

Beyond Worst-case: A Probabilistic Analysis of Affine Policies in Dynamic Optimization

Omar El Housni * Vineet Goyal *

June 20, 2017

Abstract

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, the worst-case approximation bound for affine policies is $O(\sqrt{m})$ that is also tight (see Bertsimas and Goyal [8]), whereas observed empirical performance is near-optimal. In this paper, we aim to address this stark-contrast between the worst-case and the empirical performance of affine policies. In particular, we show that affine policies give a good approximation for the two-stage adjustable robust optimization problem with high probability on random instances where the constraint coefficients are generated i.i.d. from a large class of distributions; thereby, providing a theoretical justification of the observed empirical performance. On the other hand, we also present a distribution such that the performance bound for affine policies on instances generated according to that distribution is $\Omega(\sqrt{m})$ with high probability; however, the constraint coefficients are not i.i.d.. This demonstrates that the empirical performance of affine policies can depend on the generative model for instances.

1 Introduction

In most real word problems, parameters are uncertain at the optimization phase and decisions need to be made in the face of uncertainty. 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 [13]. We refer the reader to Kall and Wallace [19], Prekopa [20], Shapiro [21], Shapiro et al. [22] 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 [23] and has been extensively studied in recent past. We refer the reader to Ben-Tal and Nemirovski [3, 4, 5], El Ghaoui and Lebret [14], Bertsimas and Sim [10, 11], Goldfarb and Iyengar [17], Bertsimas et al. [6] and Ben-Tal et al. [1] for a detailed discussion of robust optimization. However, in both these paradigms, computing an optimal dynamic solution is intractable in general due to the “curse of dimensionality”.

This intractability of computing the optimal adjustable solution necessitates considering approximate solution policies such as static and affine policies where the decision in any period t is

*Industrial Engineering and Operations Research, Columbia University, Email: {oe2148,vg2277}@columbia.edu.

restricted to a particular 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. Our goal in this paper is to address this stark contrast between the worst-case performance bounds and near-optimal empirical performance of affine policies.

In particular, we consider the following two-stage adjustable robust linear optimization problems with uncertain demand requirements:

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

where $\mathbf{A} \in \mathbb{R}_+^{m \times n}$, $\mathbf{c} \in \mathbb{R}_+^n$, $\mathbf{d} \in \mathbb{R}_+^n$, $\mathbf{B} \in \mathbb{R}_+^{m \times n}$. The right-hand-side \mathbf{h} belongs to a compact convex uncertainty set $\mathcal{U} \subseteq \mathbb{R}_+^m$. The goal in this problem is to select the first-stage decision \mathbf{x} , and the second-stage recourse decision, $\mathbf{y}(\mathbf{h})$, as a function of the uncertain right hand side realization, \mathbf{h} such that the worst-case cost over all realizations of $\mathbf{h} \in \mathcal{U}$ is minimized. We assume without loss of generality that $\mathbf{c} = \mathbf{e}$ and $\mathbf{d} = \bar{d} \cdot \mathbf{e}$ (by appropriately scaling \mathbf{A} and \mathbf{B}). Here, \bar{d} can be interpreted as the inflation factor for costs in the second-stage.

This model captures many important applications including set cover, facility location and network design problems under uncertain demand. Here the right hand side, \mathbf{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. [16] show that $\Pi_{\text{AR}}(\mathcal{U})$ (1.1) is hard to approximate within any factor that is better than $\Omega(\log n)$.

Both static and affine policy approximations have been studied in the literature for (1.1). In a static solution, we compute a single optimal solution (\mathbf{x}, \mathbf{y}) that is feasible for all realizations of the uncertain right hand side. Bertsimas et al. [9] 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 is close to being centrally symmetric. However, the performance of static solutions can be arbitrarily large for a general convex uncertainty set with the worst case performance being $\Omega(m)$. El Housni and Goyal [15] consider piecewise static policies for two-stage adjustable robust problem with uncertain constraint coefficients. These are a generalization of static policies where we divide the uncertainty set into several pieces and specify a static solution for each piece. However, they 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.

An affine policy restricts the second-stage decisions, $\mathbf{y}(\mathbf{h})$ to being an affine function of the uncertain right-hand-side \mathbf{h} , i.e., $\mathbf{y}(\mathbf{h}) = \mathbf{P}\mathbf{h} + \mathbf{q}$ for some $\mathbf{P} \in \mathbb{R}^{n \times m}$ and $\mathbf{q} \in \mathbb{R}^n$ are decision variables. Affine policies in this context were introduced in Ben-Tal et al. [2] and can be formulated as:

$$\begin{aligned}
z_{\text{Aff}}(\mathbf{c}, \mathbf{d}, \mathbf{A}, \mathbf{B}, \mathcal{U}) &= \min_{\mathbf{x}} \mathbf{c}^T \mathbf{x} + \max_{\mathbf{h} \in \mathcal{U}} \min_{\mathbf{P}, \mathbf{q}} \mathbf{d}^T (\mathbf{P}\mathbf{h} + \mathbf{q}) \\
&\mathbf{A}\mathbf{x} + \mathbf{B}(\mathbf{P}\mathbf{h} + \mathbf{q}) \geq \mathbf{h} \quad \forall \mathbf{h} \in \mathcal{U} \\
&\mathbf{P}\mathbf{h} + \mathbf{q} \geq \mathbf{0} \quad \forall \mathbf{h} \in \mathcal{U} \\
&\mathbf{x} \in \mathbb{R}_+^n
\end{aligned} \tag{1.2}$$

An optimal affine policy can be computed efficiently for a large class of problems. Bertsimas and Goyal [8] show that affine policies give a $O(\sqrt{m})$ -approximation to the optimal dynamic solution for (1.1). Furthermore, they show that the approximation bound $O(\sqrt{m})$ is tight. However, the observed empirical performance for affine policies is near-optimal for a large set of synthetic instances of (1.1).

1.1 Our Contributions

Our goal in this paper is to address this stark contrast by providing a theoretical analysis of the performance of affine policies on synthetic instances of the problem generated from a probabilistic model. In particular, we consider random instances of the two-stage adjustable problem (1.1) where the entries of the constraint matrix \mathbf{B} are random from a given distribution and analyze the performance of affine policies for a large class of distributions. Our main contributions are summarized below.

Independent and Identically distributed Constraint Coefficients. We consider random instances of the two-stage adjustable problem where the entries of \mathbf{B} are generated i.i.d. according to a given distribution and show that an affine policy gives a good approximation for a large class of distributions including distributions with bounded support and unbounded distributions with Gaussian and sub-gaussian tails.

In particular, for distributions with bounded support in $[0, b]$ and expectation μ , we show that for sufficiently large values of m and n , affine policy gives a b/μ -approximation to the adjustable problem (1.1). More specifically, with probability at least $(1 - 1/m)$, we have that

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

where $\epsilon = b/\mu\sqrt{\log m/n}$ (Theorem 2.1). Therefore, if the distribution is *symmetric*, affine policy gives a 2-approximation for the adjustable problem (1.1). For instance, for the case of uniform distribution, or Bernoulli distribution with parameter $p = 1/2$, affine gives a nearly 2-approximation for (1.1).

While the above bound 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. More specifically, we show that if B_{ij} are i.i.d. according to an unbounded distribution with a sub-gaussian tail, then for sufficiently large values of m and n , with probability at least $(1 - 1/m)$,

$$z_{\text{Aff}}(\mathbf{c}, \mathbf{d}, \mathbf{A}, \mathbf{B}, \mathcal{U}) \leq O(\sqrt{\log mn}) \cdot z_{\text{AR}}(\mathbf{c}, \mathbf{d}, \mathbf{A}, \mathbf{B}, \mathcal{U}).$$

Here we assume that the parameters of the distributions are constants independent of the problem dimension. We prove the case of *folded normal* distribution in Theorem 2.6.

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 which is tight. For the random instances where B_{ij} are i.i.d. according to above distributions, the performance 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 [7] 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 \mathbf{B} . Using the probabilistic structure of \mathbf{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 is computed using a MIP.

Worst-case distribution for Affine policies. While for a large class of commonly used distributions, affine policies give a good approximation with high probability for random i.i.d. instances according to the given distribution, we present a distribution where the performance of affine policies is $\Omega(\sqrt{m})$ with high probability for instances generated from this distribution. Note that this matches the worst-case deterministic bound for affine policies. We would like to remark that in the worst-case distribution, the coefficients B_{ij} are not identically distributed. Our analysis 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.

2 Random instances with i.i.d. coefficients

In this section, we theoretically characterize the performance of affine policies for random instances of (1.1) for a large class of generative distributions including both bounded and unbounded support distributions. In particular, we consider the two-stage problem where constraint coefficients \mathbf{A} and \mathbf{B} are i.i.d. according to a given distribution. We consider a polyhedral uncertainty set \mathcal{U} given as

$$\mathcal{U} = \{\mathbf{h} \in \mathbb{R}_+^m \mid \mathbf{R}\mathbf{h} \leq \mathbf{r}\} \quad (2.1)$$

where $\mathbf{R} \in \mathbb{R}_+^{L \times m}$ and $\mathbf{r} \in \mathbb{R}_+^L$. This is a fairly general class of uncertainty sets that includes many commonly used sets such as hypercube and *budget uncertainty* sets.

Our analysis of the performance of affine policies does not depend on the structure of first stage constraint matrix \mathbf{A} or cost \mathbf{c} . The second-stage cost, as already mentioned, is wlog of the form $\mathbf{d} = \bar{d}\mathbf{e}$. Therefore, we restrict our attention only to the distribution of coefficients of the second stage matrix \mathbf{B} . We will use the notation $\tilde{\mathbf{B}}$ to emphasize that \mathbf{B} is random. For simplicity, we refer to $z_{\text{AR}}(\mathbf{c}, \mathbf{d}, \mathbf{A}, \mathbf{B}, \mathcal{U})$ as $z_{\text{AR}}(\mathbf{B})$ and to $z_{\text{Aff}}(\mathbf{c}, \mathbf{d}, \mathbf{A}, \mathbf{B}, \mathcal{U})$ as $z_{\text{Aff}}(\mathbf{B})$.

2.1 Distributions with bounded support

We first consider the case when \tilde{B}_{ij} are i.i.d. according to a bounded distribution with support in $[0, b]$ for some constant b independent of the dimension of the problem. We show a performance bound of affine policies as compared to the optimal dynamic solution. The bound depends only on the distribution of $\tilde{\mathbf{B}}$ and holds for any polyhedral uncertainty set \mathcal{U} . In particular, we have the following theorem.

Theorem 2.1. *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 $\mathbb{E}[\tilde{B}_{ij}] = \mu \forall i \in [m] \forall j \in [n]$. For n and m sufficiently large, we have with probability at least $1 - \frac{1}{m}$,*

$$z_{\text{AR}}(\tilde{\mathbf{B}}) \leq z_{\text{Aff}}(\tilde{\mathbf{B}}) \leq \frac{b}{\mu(1 - \epsilon)} \cdot z_{\text{AR}}(\tilde{\mathbf{B}})$$

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

The above theorem shows 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 Theorem 2.1 we have with probability at least $1 - 1/m$,

$$z_{\text{AR}}(\tilde{\mathbf{B}}) \leq z_{\text{Aff}}(\tilde{\mathbf{B}}) \leq \frac{2}{1 - \epsilon} \cdot z_{\text{AR}}(\tilde{\mathbf{B}})$$

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 Theorem 2.1 we have with probability at least $1 - \frac{1}{m}$,

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

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. We discuss the intuition and the proof of Theorem 2.1 in the following subsections.

2.1.1 Preliminaries

In order to prove Theorem 2.1, we need to introduce certain preliminary results. We first introduce the following formulation for the adjustable problem (1.1) based on ideas in Bertsimas and de Ruiter [7].

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

where the set \mathcal{W} is defined as

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

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

Lemma 2.2. Let $z_{\text{AR}}(\mathbf{B})$ be as defined in (1.1) and $z_{\text{d-AR}}(\mathbf{B})$ as defined in (2.2). Then,

$$z_{\text{AR}}(\mathbf{B}) = z_{\text{d-AR}}(\mathbf{B}).$$

The proof follows from [7]. For completeness, we present it in Appendix A. Reformulation (2.2) can be interpreted as a new two-stage adjustable problem over *dualized* uncertainty set \mathcal{W} and decision $\boldsymbol{\lambda}(\mathbf{w})$. Following [7], we refer to (2.2) as the *dualized* formulation and to (1.1) as the *primal* formulation. Bertsimas and de Ruiter [7] 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 [7].

Lemma 2.3. (Theorem 2 in Bertsimas and de Ruiter [7]) Let $z_{\text{d-Aff}}(\mathbf{B})$ be the objective value when $\boldsymbol{\lambda}(\mathbf{w})$ is restricted to be affine function of \mathbf{w} and $z_{\text{Aff}}(\mathbf{B})$ as defined in (1.2). Then,

$$z_{\text{d-Aff}}(\mathbf{B}) = z_{\text{Aff}}(\mathbf{B}).$$

Bertsimas and Goyal [8] 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 [8]. In particular, we have the following lemma.

Lemma 2.4. Suppose the set \mathcal{W} is a simplex, i.e. a convex combination of $m+1$ affinely independent points, then affine policy is optimal for the adjustable problem (2.2), i.e. $z_{\text{d-Aff}}(\mathbf{B}) = z_{\text{d-AR}}(\mathbf{B})$.

The proof proceeds along similar lines as in 2.4. For completeness, we provide it in Appendix A. 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 2.5. 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,

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

Furthermore,

$$z_{\text{AR}}(\mathbf{B}) \leq z_{\text{Aff}}(\mathbf{B}) \leq \kappa \cdot z_{\text{AR}}(\mathbf{B}).$$

The proof of Lemma 2.5 is presented in Appendix A.

2.1.2 Proof of Theorem 2.1

We consider instances of problem (1.1) where \tilde{B}_{ij} are i.i.d. according to a bounded distribution with support in $[0, b]$ and $\mathbb{E}[\tilde{B}_{ij}] = \mu$ for all $i \in [m], j \in [n]$. Denote the dualized uncertainty set $\tilde{\mathcal{W}} = \{\mathbf{w} \in \mathbb{R}_+^m \mid \tilde{\mathbf{B}}^T \mathbf{w} \leq \bar{d} \cdot \mathbf{e}\}$. 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,

$$\mathcal{S} = \left\{ \mathbf{w} \in \mathbb{R}_+^m \mid \sum_{i=1}^m w_i \leq \frac{\bar{d}}{b} \right\}. \quad (2.4)$$

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 $\mathbf{w} \in \mathcal{S}$. For any any $i = 1, \dots, n$

$$\sum_{j=1}^m \tilde{B}_{ji} w_j \leq b \sum_{j=1}^m w_j \leq \bar{d}.$$

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

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

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

Using Hoeffding's inequality [18] (see Appendix B) with $\tau = b\sqrt{\frac{\log m}{n}}$, we have

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

and a union bound over $j = 1, \dots, m$ gives us

$$\mathbb{P} \left(\frac{\sum_{i=1}^n \tilde{B}_{ji}}{n} \geq \mu - \tau \quad \forall j = 1, \dots, m \right) \geq \left(1 - \frac{1}{m^2} \right)^m \geq 1 - \frac{1}{m}.$$

where the last inequality follows from Bernoulli's inequality. Therefore, with probability at least $1 - \frac{1}{m}$, we have

$$\sum_{j=1}^m w_j \leq \sum_{j=1}^m \frac{1}{\mu - \tau} \left(\frac{\sum_{i=1}^n \tilde{B}_{ji}}{n} \right) \cdot w_j \leq \frac{\bar{d}}{(\mu - \tau)} = \frac{b}{\mu(1 - \epsilon)} \cdot \frac{\bar{d}}{b}$$

where the second inequality follows from (2.5). Note that for m sufficiently large, we have $\mu - \tau > 0$. Then, $\mathbf{w} \in \frac{b}{\mu(1 - \epsilon)} \cdot \mathcal{S}$ for any $\mathbf{w} \in \tilde{\mathcal{W}}$ and consequently $\mathcal{S} \subseteq \tilde{\mathcal{W}} \subseteq \frac{b}{\mu(1 - \epsilon)} \cdot \mathcal{S}$ with probability at least $1 - 1/m$. Finally, we apply the result of Lemma 2.5 to conclude. \square

2.2 Unbounded distributions

While the approximation bound in Theorem 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 even for the case of distributions with unbounded support and sub-gaussian tails. In this section, we consider the special case where \tilde{B}_{ij} are i.i.d. according to absolute value of a standard Gaussian, also called the *folded normal* distribution, and show a logarithmic approximation bound for affine policies. In particular, we have the following theorem.

Theorem 2.6. *Consider the two-stage adjustable problem (1.1) where $\forall i \in [n], j \in [m], \tilde{B}_{ij} = |\tilde{G}_{ij}|$ and \tilde{G}_{ij} are i.i.d. according to a standard Gaussian distribution. For n and m sufficiently large, we have with probability at least $1 - \frac{1}{m}$,*

$$z_{\text{AR}}(\tilde{\mathbf{B}}) \leq z_{\text{Aff}}(\tilde{\mathbf{B}}) \leq \kappa \cdot z_{\text{AR}}(\tilde{\mathbf{B}})$$

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

Proof. Denote $\tilde{\mathcal{W}} = \{\mathbf{w} \in \mathbb{R}_+^m \mid \tilde{\mathbf{B}}^T \mathbf{w} \leq \bar{d} \cdot \mathbf{e}\}$ and $\mathcal{S} = \{\mathbf{w} \in \mathbb{R}_+^m \mid \sum_{i=1}^m w_i \leq \bar{d}\}$. Our goal is to sandwich $\tilde{\mathcal{W}}$ between two simplicies and use Lemma 2.5. Using the following tail inequality for Gaussian random variables $\tilde{G} \sim \mathcal{N}(\mu, \sigma^2)$, $\mathbb{P}(|\tilde{G} - \mu| \geq t) \leq 2e^{-\frac{t^2}{2\sigma^2}}$, we have

$$\begin{aligned} \mathbb{P}(\tilde{B}_{ij} \leq \sqrt{6 \log(mn)}) &= 1 - \mathbb{P}\left(|\tilde{G}_{ij}| \geq \sqrt{6 \log(mn)}\right) \\ &\geq 1 - 2 \exp\left(\frac{-6 \log(mn)}{2}\right) = 1 - \frac{2}{(mn)^3} \geq 1 - \frac{1}{(mn)^2} \end{aligned}$$

Therefore by taking a union bound,

$$\mathbb{P}\left(\tilde{B}_{ij} \leq \sqrt{6 \log(mn)} \quad \forall i \in [n], \forall j \in [m]\right) \geq \left(1 - \frac{1}{(mn)^2}\right)^{mn} \geq 1 - \frac{1}{mn}$$

where the last inequality follows from Bernoulli's inequality. Therefore for any $w \in \mathcal{S}$, we have with probability at least $1 - \frac{1}{mn}$,

$$\sum_{j=1}^m \tilde{B}_{ji} w_j \leq \sqrt{6 \log(mn)} \sum_{j=1}^m w_j \leq \sqrt{6 \log(mn)} \cdot \bar{d} \quad \forall i \in [n]$$

Hence, with probability at least $1 - \frac{1}{mn}$ we have, $\mathcal{S} \subseteq \sqrt{6 \log(mn)} \cdot \tilde{\mathcal{W}}$.

Now, we want to find a simplex that includes $\tilde{\mathcal{W}}$. We follow a similar approach to the proof of Theorem 2.1. Consider any $\mathbf{w} \in \tilde{\mathcal{W}}$. We have similarly to equation (2.5)

$$\sum_{j=1}^m \left(\frac{\sum_{i=1}^n \tilde{B}_{ji}}{n}\right) \cdot w_j \leq \bar{d}. \quad (2.6)$$

We have the following concentration inequality for non-negative random variables (see Theorem 7 in [12]),

$$\mathbb{P}\left(\frac{\sum_{i=1}^n \tilde{B}_{ji}}{n} \geq \mu - \tau\right) \geq 1 - \exp\left(\frac{-n\tau^2}{2\mathbb{E}(\tilde{B}_{11}^2)}\right) = 1 - \exp\left(\frac{-n\tau^2}{2}\right) = 1 - \frac{1}{m^2}$$

where $\tau = 2\sqrt{\frac{\log m}{n}}$ and $\mu = \mathbb{E}[\tilde{B}_{ji}] = \sqrt{\frac{2}{\pi}}$ is the expectation of a folded standard normal distribution. Then, union bound over $j = 1, \dots, m$ gives us

$$\mathbb{P}\left(\frac{\sum_{i=1}^n \tilde{B}_{ji}}{n} \geq \mu - \tau \quad \forall j = 1, \dots, m\right) \geq \left(1 - \frac{1}{m^2}\right)^m \geq 1 - \frac{1}{m}.$$

where the last inequality follows from Bernoulli's inequality. Therefore, combining this result with inequality (2.6), we have with probability at least $1 - \frac{1}{m}$, $\tilde{\mathcal{W}} \subseteq \frac{1}{\mu - \tau} \mathcal{S}$. Denote, $\mathcal{S}' = \frac{1}{\sqrt{6 \log(mn)}} \mathcal{S}$.

Then, we have with probability at least $1 - \frac{1}{m}$, $\mathcal{S}' \subseteq \tilde{\mathcal{W}} \subseteq \kappa \cdot \mathcal{S}'$ where

$$\kappa = \frac{\sqrt{6 \log(mn)}}{\sqrt{\frac{2}{\pi}} - 2\sqrt{\frac{\log m}{n}}} = O\left(\sqrt{\log m + \log n}\right),$$

for sufficiently large values of m and n . We finally use Lemma 2.5 to conclude. \square

We can extend the analysis and show a similar bound for the class of distributions with sub-gaussian tails. The bound of $O(\sqrt{\log m + \log n})$ depends on the dimension of the problem unlike the case of uniform bounded distribution. But, it is significantly better than the worst-case of $O(\sqrt{m})$ [8] for general instances. Furthermore, this bound holds for all uncertainty sets with high probability. 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 affine policies are near optimal.

3 Family of worst-case distribution: perturbation of i.i.d. coefficients

For any m sufficiently large, the authors in [8] 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 [8] 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 measure zero set. In fact, we exhibit a distribution and an uncertainty set such that a random instance 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 not identically distributed. The instance can be given as follows.

$$\begin{aligned} n &= m, \quad \mathbf{A} = \mathbf{0}, \quad \mathbf{c} = \mathbf{0}, \quad \mathbf{d} = \mathbf{e} \\ \mathcal{U} &= \text{conv}(\mathbf{0}, \mathbf{e}_1, \dots, \mathbf{e}_m, \boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_m) \quad \text{where } \boldsymbol{\nu}_i = \frac{1}{\sqrt{m}}(\mathbf{e} - \mathbf{e}_i) \quad \forall i \in [m]. \\ \tilde{B}_{ij} &= \begin{cases} 1 & \text{if } i = j \\ \frac{1}{\sqrt{m}} \cdot \tilde{u}_{ij} & \text{if } i \neq j \end{cases} \quad \text{where for all } i \neq j, \tilde{u}_{ij} \text{ are i.i.d. uniform}[0, 1]. \end{aligned} \tag{3.1}$$

Theorem 3.1. *For the instance defined in (3.1), we have with probability at least $1 - 1/m$,*

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

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 [8]. We would like to note that both uncertainty set and distribution of coefficients in our instance (3.1) 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.

To prove Theorem 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 3.2. *Consider the two-stage adjustable problem (1.1) where: $n = m, \mathbf{c} = \mathbf{0}, \mathbf{d} = \mathbf{e}, \mathbf{A} = \mathbf{0}$,*

$$B_{ij} = \begin{cases} 1 & \text{if } i = j \\ \frac{1}{\sqrt{m}} & \text{if } i \neq j \end{cases} \tag{3.2}$$

and the uncertainty set is defined as

$$\mathcal{U} = \text{conv}(\mathbf{0}, \mathbf{e}_1, \dots, \mathbf{e}_m, \boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_m) \tag{3.3}$$

where $\boldsymbol{\nu}_i = \frac{1}{\sqrt{m}}(\mathbf{e} - \mathbf{e}_i)$ for $i = 1, \dots, m$. Then, $z_{\text{Aff}}(\mathbf{B}) = \Omega(\sqrt{m}) \cdot z_{\text{AR}}(\mathbf{B})$.

Proof. First, let us prove that $z_{\text{AR}}(\mathbf{B}) \leq 1$. 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 $i = 1, \dots, m$.

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

We have $\mathbf{B}\mathbf{y}(\mathbf{0}) = \mathbf{0}$ and for $i \in [m]$

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

and

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

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

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

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

$$\mathbf{P} = \begin{pmatrix} \theta & \mu & \dots & \mu \\ \mu & \theta & \dots & \mu \\ \vdots & \vdots & \ddots & \vdots \\ \mu & \mu & \dots & \theta \end{pmatrix} \quad (3.4)$$

and $\mathbf{q} = \lambda\mathbf{e}$.

We have $\mathbf{y}(\mathbf{0}) = \lambda\mathbf{e} \geq \mathbf{0}$ hence

$$\lambda \geq 0. \quad (3.5)$$

We know that

$$z_{\text{Aff}}(\mathbf{B}) \geq \mathbf{d}^T \mathbf{y}(\mathbf{0}) = \lambda m. \quad (3.6)$$

Case 1: If $\lambda \geq \frac{1}{6\sqrt{m}}$, then from (3.6) we have $z_{\text{Aff}}(\mathbf{B}) \geq \frac{\sqrt{m}}{6}$.

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

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

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

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

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

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,

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

In particular we have ,

$$\begin{aligned} z_{\text{Aff}}(\mathbf{B}) &\geq \mathbf{d}^T \mathbf{y}(\boldsymbol{\nu}_1) = \frac{1}{\sqrt{m}} ((m-1)\theta + (m-1)^2\mu) + \lambda m \\ &\geq \frac{m-1}{\sqrt{m}} \left(\frac{1}{3} + (m-1)\mu \right). \end{aligned} \quad (3.7)$$

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 (3.7)

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

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

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

i.e.

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

and from (3.7)

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

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

Proof of Theorem 3.1

Proof. Denote $\mathcal{W} = \{\mathbf{w} \in \mathbb{R}_+^m \mid \mathbf{B}^T \mathbf{w} \leq \bar{d}\mathbf{e}\}$ and $\tilde{\mathcal{W}} = \{\mathbf{w} \in \mathbb{R}_+^m \mid \tilde{\mathbf{B}}^T \mathbf{w} \leq \bar{d}\mathbf{e}\}$ where \mathbf{B} is defined in (3.2) and $\tilde{\mathbf{B}}$ is defined in (3.1). Since for all i, j in $\{1, \dots, m\}$ we have $\tilde{B}_{ij} \leq B_{ij}$. Hence, for any $\mathbf{w} \in \mathcal{W}$, we have $\tilde{\mathbf{B}}^T \mathbf{w} \leq \mathbf{B}^T \mathbf{w} \leq \bar{d}\mathbf{e}$. Therefore $\mathbf{w} \in \tilde{\mathcal{W}}$ and consequently $\mathcal{W} \subseteq \tilde{\mathcal{W}}$.

Now, suppose $\mathbf{w} \in \tilde{\mathcal{W}}$, we have for all $i = 1, \dots, m$

$$w_i + \frac{1}{\sqrt{m}} \sum_{\substack{j=1 \\ j \neq i}}^m \tilde{u}_{ji} w_j \leq \bar{d}. \quad (3.8)$$

By taking the sum over i , dividing by m and rearranging, we get

$$\sum_{i=1}^m w_i \left(\frac{1}{m} + \frac{1}{m\sqrt{m}} \sum_{\substack{j=1 \\ j \neq i}}^m \tilde{u}_{ij} \right) \leq \bar{d}. \quad (3.9)$$

Here, similarly to the proof of Lemma 2.1 we apply Hoeffding's inequality [18](see appendix B), with $\tau = \sqrt{\frac{\log m}{m-1}}$

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

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

$$\mathbb{P} \left(\frac{\sum_{i=1}^n \tilde{u}_{ij}}{m-1} \geq \frac{1}{2} - \epsilon \quad \forall j = 1, \dots, m \right) \geq \left(1 - \frac{1}{m^2} \right)^m \geq 1 - \frac{1}{m}. \quad (3.10)$$

where the last inequality follows from Bernoulli's inequality. Therefore, we conclude from (3.9) and (3.10), 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 (3.8) that for all i we have $w_i \leq \bar{d}$. Hence with probability at least $1 - \frac{1}{m}$, we have for all $i = 1, \dots, m$

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

Therefore, $\mathbf{w} \in 5 \cdot \mathcal{W}$ for any \mathbf{w} in \mathcal{W} and consequently we have with probability at least $1 - \frac{1}{m}$, $\tilde{\mathcal{W}} \subseteq 5 \cdot \mathcal{W}$. All together we have proved with probability at least $1 - \frac{1}{m}$ $\mathcal{W} \subseteq \tilde{\mathcal{W}} \subseteq 5 \cdot \mathcal{W}$. This implies with probability at least $1 - \frac{1}{m}$, that $z_{\text{d-Aff}}(\tilde{\mathbf{B}}) \geq z_{\text{d-Aff}}(\mathbf{B})$ and $z_{\text{d-AR}}(\mathbf{B}) \geq \frac{z_{\text{d-AR}}(\tilde{\mathbf{B}})}{5}$. We know from Lemma 2.3 and Lemma 2.2 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 $z_{\text{Aff}}(\tilde{\mathbf{B}}) \geq z_{\text{Aff}}(\mathbf{B})$ and $z_{\text{AR}}(\mathbf{B}) \geq \frac{z_{\text{AR}}(\tilde{\mathbf{B}})}{5}$.

Moreover, we know from Lemma 3.2 that $z_{\text{Aff}}(\mathbf{B}) \geq \Omega(\sqrt{m}) \cdot z_{\text{AR}}(\mathbf{B})$. Therefore, $z_{\text{Aff}}(\tilde{\mathbf{B}}) \geq \Omega(\sqrt{m})z_{\text{AR}}(\tilde{\mathbf{B}})$ with probability at least $1 - \frac{1}{m}$. □

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 $\tilde{\mathbf{B}}$ are i.i.d. uniform $[0, 1]$, and *ii*) Coefficients of $\tilde{\mathbf{B}}$ are absolute value of i.i.d. standard Gaussian. We consider the following *budget of uncertainty* set.

$$\mathcal{U} = \left\{ \mathbf{h} \in [0, 1]^m \mid \sum_{i=1}^m h_i \leq \sqrt{m} \right\}. \quad (4.1)$$

Note that the set (4.1) 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 $O(\log n)$ [16]. We consider $n = m$, $\mathbf{d} = \mathbf{e}$. Also, we consider $\mathbf{c} = \mathbf{0}$, $\mathbf{A} = \mathbf{0}$. 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 \mathbf{c} and matrix \mathbf{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 $r = z_{\text{Aff}}(\tilde{\mathbf{B}})/z_{\text{AR}}(\tilde{\mathbf{B}})$ in Table 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 $T_{\text{AR}}(s)$ and the running time of affine policy $T_{\text{Aff}}(s)$. 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

$$\begin{aligned} z_{\text{Aff}}(\mathbf{B}) &= \min_{\mathbf{x}} \mathbf{c}^T \mathbf{x} + z \\ z &\geq \mathbf{d}^T (\mathbf{P}\mathbf{h} + \mathbf{q}) \quad \forall \mathbf{h} \in \mathcal{U} \\ \mathbf{A}\mathbf{x} + \mathbf{B}(\mathbf{P}\mathbf{h} + \mathbf{q}) &\geq \mathbf{h} \quad \forall \mathbf{h} \in \mathcal{U} \\ \mathbf{P}\mathbf{h} + \mathbf{q} &\geq \mathbf{0} \quad \forall \mathbf{h} \in \mathcal{U} \\ \mathbf{x} &\in \mathbb{R}_+^n \end{aligned}$$

Note that this formulation has infinitely many constraints but we can write a compact LP formulation using standard techniques from duality. For example, the first constraint is equivalent to

$$z - \mathbf{d}^T \mathbf{q} \geq \max \{ \mathbf{d}^T \mathbf{P}\mathbf{h} \mid \mathbf{R}\mathbf{h} \leq \mathbf{r}, \mathbf{h} \geq \mathbf{0} \}.$$

By taking the dual of the maximization problem, the constraint becomes

$$z - \mathbf{d}^T \mathbf{q} \geq \min \{ \mathbf{r}^T \mathbf{v} \mid \mathbf{R}^T \mathbf{v} \geq \mathbf{P}^T \mathbf{d}, \mathbf{v} \geq \mathbf{0} \}.$$

We can then drop the min and introduce \mathbf{v} as a variable, hence we obtain the following linear constraints

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

We can apply the same techniques for the other constraints. The complete LP formulation and its proof of correctness is presented in Appendix C.

MIP Formulation for the adjustable problem (1.1). For the adjustable problem (1.1), we show that the separation problem (4.2) can be formulated as a mixed integer program. The separation problem can be formulated as follows: Given $\hat{\mathbf{x}}$ and \hat{z} decide whether

$$\max \{ (\mathbf{h} - \mathbf{A}\hat{\mathbf{x}})^T \mathbf{w} \mid \mathbf{w} \in \mathcal{W}, \mathbf{h} \in \mathcal{U} \} > \hat{z} \quad (4.2)$$

The correctness of formulation (4.2) follows from equation (A.1) in the proof of Lemma 2.2 in Appendix A. The constraints in (4.2) are linear but the objective function contains a bilinear term, $\mathbf{h}^T \mathbf{w}$. We linearize this using a standard *digitized reformulation*. In particular, we consider finite bit representations of continuous variables, h_i and 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 C.

For general $\mathbf{A} \neq 0$, 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 $\mathbf{A} = 0, \mathbf{c} = 0$ 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 2.1 and Theorem 2.6. For instance, Theorem 2.1 implies that affine policy is a 2-approximation with high probability for random instances from a uniform distribution. 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 2.6 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}	$T_{\text{AR}}(s)$	$T_{\text{Aff}}(s)$
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}	$T_{\text{AR}}(s)$	$T_{\text{Aff}}(s)$
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 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 $z_{\text{Aff}}(\tilde{\mathbf{B}})/z_{\text{AR}}(\tilde{\mathbf{B}})$ and present the average and max ratios. Here, $T_{\text{AR}}(s)$ denotes the running time for the adjustable policy and $T_{\text{Aff}}(s)$ 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.

References

- [1] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*. Princeton University press, 2009.
- [2] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.
- [3] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.
- [4] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–14, 1999.
- [5] A. Ben-Tal and A. Nemirovski. Robust optimization—methodology and applications. *Mathematical Programming*, 92(3):453–480, 2002.
- [6] D. Bertsimas, D. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM review*, 53(3):464–501, 2011.
- [7] D. Bertsimas and F. J. de Ruiter. Duality in two-stage adaptive linear optimization: Faster computation and stronger bounds. *INFORMS Journal on Computing*, 28(3):500–511, 2016.
- [8] D. Bertsimas and V. Goyal. On the Power and Limitations of Affine Policies in Two-Stage Adaptive Optimization. *Mathematical Programming*, 134(2):491–531, 2012.
- [9] 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*, 36(1):24–54, 2011.
- [10] D. Bertsimas and M. Sim. Robust Discrete Optimization and Network Flows. *Mathematical Programming Series B*, 98:49–71, 2003.
- [11] D. Bertsimas and M. Sim. The Price of Robustness. *Operations Research*, 52(2):35–53, 2004.
- [12] F. Chung and L. Lu. Concentration inequalities and martingale inequalities: a survey. *Internet Mathematics*, 3(1):79–127, 2006.
- [13] G. Dantzig. Linear programming under uncertainty. *Management Science*, 1:197–206, 1955.
- [14] L. El Ghaoui and H. Lebret. Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications*, 18:1035–1064, 1997.
- [15] O. El Housni and V. Goyal. Piecewise static policies for two-stage adjustable robust linear optimization. *Mathematical Programming*, pages 1–17, 2017.
- [16] U. Feige, K. Jain, M. Mahdian, and V. Mirrokni. Robust combinatorial optimization with exponential scenarios. *Lecture Notes in Computer Science*, 4513:439–453, 2007.
- [17] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.
- [18] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.
- [19] P. Kall and S. Wallace. *Stochastic programming*. Wiley New York, 1994.
- [20] A. Prékopa. *Stochastic programming*. Kluwer Academic Publishers, Dordrecht, Boston, 1995.
- [21] A. Shapiro. Stochastic programming approach to optimization under uncertainty. *Mathematical Programming, Series B*, 112(1):183–220, 2008.
- [22] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming: modeling and theory*. Society for Industrial and Applied Mathematics, 2009.
- [23] A. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research*, 21(5):1154–1157, 1973.

A Proofs of preliminaries

Proof of Lemma 2.2

Proof. We have

$$\begin{aligned}
z_{\text{AR}}(\mathbf{B}) &= \min_{\mathbf{x} \geq \mathbf{0}} \mathbf{c}^T \mathbf{x} + \max_{\mathbf{h} \in \mathcal{U}} \min_{\substack{\mathbf{B}\mathbf{y} \geq \mathbf{h} - \mathbf{A}\mathbf{x} \\ \mathbf{y} \geq \mathbf{0}}} \mathbf{d}^T \mathbf{y} \\
&= \min_{\mathbf{x} \geq \mathbf{0}} \mathbf{c}^T \mathbf{x} + \max_{\mathbf{h} \in \mathcal{U}} \max_{\substack{\mathbf{B}^T \mathbf{w} \leq \mathbf{d} \\ \mathbf{w} \geq \mathbf{0}}} (\mathbf{h} - \mathbf{A}\mathbf{x})^T \mathbf{w} \\
&= \min_{\mathbf{x} \geq \mathbf{0}} \mathbf{c}^T \mathbf{x} + \max_{\mathbf{w} \in \mathcal{W}} -(\mathbf{A}\mathbf{x})^T \mathbf{w} + \max_{\substack{\mathbf{R}\mathbf{h} \leq \mathbf{r} \\ \mathbf{h} \geq \mathbf{0}}} \mathbf{h}^T \mathbf{w} \\
&= \min_{\mathbf{x} \geq \mathbf{0}} \mathbf{c}^T \mathbf{x} + \max_{\mathbf{w} \in \mathcal{W}} -(\mathbf{A}\mathbf{x})^T \mathbf{w} + \min_{\substack{\mathbf{R}^T \boldsymbol{\lambda} \geq \mathbf{w} \\ \boldsymbol{\lambda} \geq \mathbf{0}}} \mathbf{r}^T \boldsymbol{\lambda} \\
&= z_{\text{d-AR}}(\mathbf{B}).
\end{aligned} \tag{A.1}$$

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. \square

Proof of Lemma 2.4

Proof. We restate the same proof in [8] in our setting. First, since the adjustable problem is a relaxation of the affine problem then $z_{\text{d-Aff}}(\mathbf{B}) \leq z_{\text{d-AR}}(\mathbf{B})$.

Now let's prove the other inequality. Consider $\mathcal{W} = \{\mathbf{w} \in \mathbb{R}_+^m \mid \mathbf{B}^T \mathbf{w} \leq \mathbf{d}\}$ which is a simplex. Note that $\mathbf{0}$ is always an extreme point of the simplex \mathcal{W} and denote $\mathbf{w}^1, \mathbf{w}^2, \dots, \mathbf{w}^m$ the remaining m points. In particular, we have for any $\mathbf{w} \in \mathcal{W}$

$$\mathbf{w} = \sum_{j=1}^m \alpha_j \mathbf{w}^j = \mathbf{Q}\boldsymbol{\alpha}$$

where $\sum_{j=1}^m \alpha_j \leq 1$ and $\mathbf{Q} = [\mathbf{w}^1 | \mathbf{w}^2 | \dots | \mathbf{w}^m]$. Note that \mathbf{Q} is invertible since $\mathbf{w}^1, \mathbf{w}^2, \dots, \mathbf{w}^m$ are linearly independent. Hence, $\boldsymbol{\alpha} = \mathbf{Q}^{-1} \mathbf{w}$. Denote $\mathbf{x}^*, \boldsymbol{\lambda}^*(\mathbf{w})$ for $\mathbf{w} \in \mathcal{W}$ the optimal adjustable solution of the adjustable problem (2.2). We define the following affine solution $\mathbf{x} = \mathbf{x}^*$ and for $\mathbf{w} \in \mathcal{W}$, $\boldsymbol{\lambda}(\mathbf{w}) = \mathbf{P}\mathbf{Q}^{-1} \mathbf{w}$ where $\mathbf{P} = [\boldsymbol{\lambda}^*(\mathbf{w}^1) | \boldsymbol{\lambda}^*(\mathbf{w}^2) | \dots | \boldsymbol{\lambda}^*(\mathbf{w}^m)]$. In particular, we have

$$\boldsymbol{\lambda}(\mathbf{w}) = \sum_{j=1}^m \alpha_j \boldsymbol{\lambda}^*(\mathbf{w}^j).$$

Let us first check the feasibility of the solution

$$\mathbf{R}^T \boldsymbol{\lambda}(\mathbf{w}) = \sum_{j=1}^m \alpha_j \mathbf{R}^T \boldsymbol{\lambda}^*(\mathbf{w}^j) \geq \sum_{j=1}^m \alpha_j \mathbf{w}^j = \mathbf{w}$$

where the inequality follows from the feasibility of the adjustable solution. Therefore,

$$\begin{aligned}
z_{\text{d-Aff}}(\mathbf{B}) &\leq \mathbf{c}^T \mathbf{x} + \max_{\mathbf{w} \in \mathcal{W}} (-\mathbf{A}\mathbf{x})^T \mathbf{w} + \mathbf{r}^T \boldsymbol{\lambda}(\mathbf{w}) \\
&= \mathbf{c}^T \mathbf{x}^* + \max_{\boldsymbol{\alpha}} (-\mathbf{A}\mathbf{x}^*)^T \mathbf{w} + \sum_{j=1}^m \alpha_j \mathbf{r}^T \boldsymbol{\lambda}^*(\mathbf{w}^j) \\
&= \mathbf{c}^T \mathbf{x}^* + \max_{\boldsymbol{\alpha}} \sum_{j=1}^m \alpha_j ((-\mathbf{A}\mathbf{x}^*)^T \mathbf{w}^j + \mathbf{r}^T \boldsymbol{\lambda}^*(\mathbf{w}^j)) \\
&\leq \mathbf{c}^T \mathbf{x}^* + \max_{\mathbf{w} \in \mathcal{W}} ((-\mathbf{A}\mathbf{x}^*)^T \mathbf{w} + \mathbf{r}^T \boldsymbol{\lambda}^*(\mathbf{w})) \max_{\boldsymbol{\alpha}} \sum_{j=1}^m \alpha_j = z_{\text{d-AR}}(\mathbf{B})
\end{aligned}$$

where the last inequality holds because $\sum_{j=1}^m \alpha_j \leq 1$. We conclude that $z_{\text{d-Aff}}(\mathbf{B}) = z_{\text{d-AR}}(\mathbf{B})$. \square

Proof of Lemma 2.5

Proof. First the inequality $z_{\text{d-AR}}(\mathbf{B}) \leq z_{\text{d-Aff}}(\mathbf{B})$ 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,

$$z_{\text{d-Aff}}(\mathbf{B}) \leq \kappa \cdot z_{\text{d-Aff}}(\mathbf{B}, \mathcal{S})$$

where we denote $z_{\text{d-Aff}}(\mathbf{B}, \mathcal{S})$ the dualized affine problem over \mathcal{S} (it's the same problem as $z_{\text{d-Aff}}(\mathbf{B})$ where we only replace \mathcal{W} by \mathcal{S}). Since \mathcal{S} is a simplex, from Lemma 2.4, we have $z_{\text{d-Aff}}(\mathbf{B}, \mathcal{S}) = z_{\text{d-AR}}(\mathbf{B}, \mathcal{S})$. Moreover, $z_{\text{d-AR}}(\mathbf{B}, \mathcal{S}) \leq z_{\text{d-AR}}(\mathbf{B})$ because $\mathcal{S} \subseteq \mathcal{W}$. We conclude that

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

Furthermore, since $z_{\text{d-AR}}(\mathbf{B}) = z_{\text{AR}}(\mathbf{B})$ from Lemma 2.2 and $z_{\text{d-Aff}}(\mathbf{B}) = z_{\text{Aff}}(\mathbf{B})$ from Lemma 2.3, then

$$z_{\text{AR}}(\mathbf{B}) \leq z_{\text{Aff}}(\mathbf{B}) \leq \kappa \cdot z_{\text{AR}}(\mathbf{B}).$$

\square

B Hoeffding's inequality

Hoeffding's inequality[18]. Let Z_1, \dots, 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,

$$\mathbb{P}(Z - \mathbb{E}(Z) \leq -\tau) \leq \exp\left(\frac{-2n\tau^2}{(b-a)^2}\right).$$

C LP and MIP formulations for the empirical section

LP formulation for the affine problem. The affine problem (1.2) can be formulated as the following LP

$$\begin{aligned}
z_{\text{Aff}}(\mathbf{B}) &= \min \mathbf{c}^T \mathbf{x} + z \\
z - \mathbf{d}^T \mathbf{q} &\geq \mathbf{r}^T \mathbf{v} \\
\mathbf{R}^T \mathbf{v} &\geq \mathbf{P}^T \mathbf{d} \\
\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{q} &\geq \mathbf{V}^T \mathbf{r} \\
\mathbf{R}^T \mathbf{V} &\geq \mathbf{I}_m - \mathbf{B}\mathbf{P} \\
\mathbf{q} &\geq \mathbf{U}^T \mathbf{r} \\
\mathbf{U}^T \mathbf{R} + \mathbf{P} &\geq \mathbf{0} \\
\mathbf{x} \in \mathbb{R}_+^n, \mathbf{v} \in \mathbb{R}_+^L, \mathbf{U} \in \mathbb{R}_+^{L \times n}, \mathbf{V} \in \mathbb{R}_+^{L \times m}
\end{aligned} \tag{C.1}$$

Proof. The affine problem (1.2) can be reformulated as follows

$$\begin{aligned}
z_{\text{Aff}}(\mathbf{B}) &= \min_{\mathbf{x}} \mathbf{c}^T \mathbf{x} + z \\
z &\geq \mathbf{d}^T (\mathbf{P}\mathbf{h} + \mathbf{q}) \quad \forall \mathbf{h} \in \mathcal{U} \\
\mathbf{A}\mathbf{x} + \mathbf{B}(\mathbf{P}\mathbf{h} + \mathbf{q}) &\geq \mathbf{h} \quad \forall \mathbf{h} \in \mathcal{U} \\
\mathbf{P}\mathbf{h} + \mathbf{q} &\geq \mathbf{0} \quad \forall \mathbf{h} \in \mathcal{U} \\
\mathbf{x} &\in \mathbb{R}_+^n
\end{aligned}$$

We use standard duality techniques to derive formulation (C.1). The first constraint is equivalent to

$$z - \mathbf{d}^T \mathbf{q} \geq \max_{\substack{\mathbf{R}\mathbf{h} \leq \mathbf{r} \\ \mathbf{h} \geq \mathbf{0}}} \mathbf{d}^T \mathbf{P}\mathbf{h}.$$

By taking the dual of the maximization problem, the constraint is equivalent

$$z - \mathbf{d}^T \mathbf{q} \geq \min_{\substack{\mathbf{R}^T \mathbf{v} \geq \mathbf{P}^T \mathbf{d} \\ \mathbf{v} \geq \mathbf{0}}} \mathbf{r}^T \mathbf{v}$$

We can then drop the min and introduce \mathbf{v} as a variable, hence we obtain the following linear constraints

$$\begin{aligned}
z - \mathbf{d}^T \mathbf{q} &\geq \mathbf{r}^T \mathbf{v} \\
\mathbf{R}^T \mathbf{v} &\geq \mathbf{P}^T \mathbf{d} \\
\mathbf{v} &\in \mathbb{R}_+^L
\end{aligned}$$

We use the same technique for the second sets of constraints, i.e.

$$\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{q} \geq \max_{\substack{\mathbf{R}\mathbf{h} \leq \mathbf{r} \\ \mathbf{h} \geq \mathbf{0}}} \mathbf{h}(\mathbf{I}_m - \mathbf{B}\mathbf{P})$$

By taking the dual of the maximization problem for each row and dropping the min we get the following compact formulation of these constraints

$$\begin{aligned}
\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{q} &\geq \mathbf{V}^T \mathbf{r} \\
\mathbf{R}^T \mathbf{V} &\geq \mathbf{I}_m - \mathbf{B}\mathbf{P} \\
\mathbf{V} &\in \mathbb{R}_+^{L \times m}
\end{aligned}$$

Similarly, the last constraint

$$\mathbf{q} \geq \max_{\substack{\mathbf{R}\mathbf{h} \leq \mathbf{r} \\ \mathbf{h} \geq \mathbf{0}}} -\mathbf{P}\mathbf{h}$$

is equivalent to

$$\begin{aligned}
\mathbf{q} &\geq \mathbf{U}^T \mathbf{r} \\
\mathbf{U}^T \mathbf{R} + \mathbf{P} &\geq \mathbf{0} \\
\mathbf{U} &\in \mathbb{R}_+^{L \times n}.
\end{aligned}$$

□

MIP formulation for the separation adjustable problem.

The separation problem (4.2) can be formulated as the following MIP

$$\begin{aligned}
\max \quad & \sum_{i=1}^m \sum_{j=-\Delta_{\mathcal{W}}}^s \sum_{k=-\Delta_{\mathcal{U}}}^s \frac{1}{2^{j+k}} \cdot \gamma_{ijk} - (\mathbf{A}\hat{\mathbf{x}})^T \mathbf{w} \\
\mathbf{w} = \quad & \sum_{i=1}^m \sum_{j=-\Delta_{\mathcal{W}}}^s \frac{\beta_{ij}}{2^j} \cdot \mathbf{e}_i \\
\mathbf{h} = \quad & \sum_{i=1}^m \sum_{k=-\Delta_{\mathcal{U}}}^s \frac{\alpha_{ik}}{2^k} \cdot \mathbf{e}_i \\
\gamma_{ijk} \leq \quad & \beta_{ij} & \forall i \in [m], j \in [-\Delta_{\mathcal{U}}, s], k \in [-\Delta_{\mathcal{W}}, s] \\
\gamma_{ijk} \leq \quad & \alpha_{ik} & \forall i \in [m], j \in [-\Delta_{\mathcal{U}}, s], k \in [-\Delta_{\mathcal{W}}, s] \\
\gamma_{ijk} + 1 \geq \quad & \alpha_{ik} + \beta_{ij} & \forall i \in [m], j \in [-\Delta_{\mathcal{U}}, s], k \in [-\Delta_{\mathcal{W}}, s] \\
\alpha_{ik}, \beta_{ij}, \gamma_{ijk} \in \quad & \{0, 1\} & \forall i \in [m], j \in [-\Delta_{\mathcal{U}}, s], k \in [-\Delta_{\mathcal{W}}, s] \\
\mathbf{R}\mathbf{h} \leq \quad & \mathbf{r} \\
\mathbf{B}^T \mathbf{w} \leq \quad & \mathbf{d}
\end{aligned} \tag{C.2}$$

where $s = \lceil \log_2(\frac{m}{\epsilon}) \rceil$, $\Delta_{\mathcal{W}}$ 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 (4.2) is equivalent to solving the following problem for given $\hat{\mathbf{x}}$

$$\max_{\substack{\mathbf{h} \in \mathcal{U} \\ \mathbf{w} \in \mathcal{W}}} \mathbf{h}^T \mathbf{w} - (\mathbf{A}\hat{\mathbf{x}})^T \mathbf{w}$$

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 $\mathbf{h}^T \mathbf{w}$ which is a bilinear function and write it in terms of linear constraints and binary variables. Let us write $\mathbf{h} = \sum_{i=1}^m h_i \mathbf{e}_i$. For all $i \in [m]$ we digitize the component h_i as follows

$$h_i = \sum_{k=-\Delta_{\mathcal{U}}}^s \frac{\alpha_{ik}}{2^k}$$

where $s = \lceil \log_2(\frac{m}{\epsilon}) \rceil$, $\Delta_{\mathcal{U}}$ 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

$$\mathbf{h} = \sum_{i=1}^m \sum_{k=-\Delta_{\mathcal{U}}}^s \frac{\alpha_{ik}}{2^k} \cdot \mathbf{e}_i$$

Similarly, we have

$$\mathbf{w} = \sum_{i=1}^m \sum_{j=-\Delta_{\mathcal{W}}}^s \frac{\beta_{ij}}{2^j} \cdot \mathbf{e}_i$$

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

$$\begin{aligned}
\gamma_{ijk} &\leq \beta_{ij} \\
\gamma_{ijk} &\leq \alpha_{ik} \\
\gamma_{ijk} + 1 &\geq \alpha_{ik} + \beta_{ij}
\end{aligned}$$

which leads to formulation (C.2). □