A Geometric Characterization of the Power of Finite Adaptability in Multistage Stochastic and Adaptive Optimization

A Geometric Characterization of the Power of Finite Adaptability in Multi-stage Stochastic and Adaptive Optimization

Dimitris Bertsimas

Sloan School of Management and Operations Research Center, Massachusetts Institute of Technology, Cambridge MA 02139
email: dbertsim@mit.edu `http://web.mit.edu/dbertsim/www/`

Vineet Goyal

Dept of Industrial Engineering and Operations Research, Columbia University, New York NY 10027
email: vgoyal@ieor.columbia.edu `http://www.mit.edu/~goyalv/`

Xu Andy Sun

Operations Research Center, Massachusetts Institute of Technology, Cambridge MA 02139
email: sunx@mit.edu

In this paper, we show a significant role that geometric properties of uncertainty sets, such as symmetry, play in determining the power of robust and finitely adaptable solutions in multi-stage stochastic and adaptive optimization problems. We consider a fairly general class of multi-stage mixed integer stochastic and adaptive optimization problems and propose a good approximate solution policy with performance guarantees that depend on the geometric properties of the uncertainty sets. In particular, we show that a class of *finitely adaptable solutions* is a good approximation for both the multi-stage stochastic as well as the adaptive optimization problem. A finitely adaptable solution generalizes the notion of a static robust solution and specifies a small set of solutions for each stage and the solution policy implements the best solution from the given set depending on the realization of the uncertain parameters in past stages. Therefore, it is a tractable approximation to a fully-adaptable solution for the multi-stage problems. To the best of our knowledge, these are the first approximation results for the multi-stage problem in such generality. Moreover, the results and the proof techniques are quite general and also extend to include important constraints such as integrality and linear conic constraints.

Key words: robust optimization ; multi-stage stochastic optimization ; adaptive optimization; finite adaptability

MSC2000 Subject Classification: Primary: 90C15 , 90C47

OR/MS subject classification: Primary: Robust Optimization/ Stochastic Optimization

1. Introduction.

In most real world problems, several parameters are uncertain at the optimization phase and decisions are required to be made in the face of these uncertainties. Deterministic optimization is often not useful for such problems as solutions obtained through such an approach might be sensitive to even small perturbations in the problem parameters. Stochastic optimization was introduced by Dantzig [15] and Beale [1], and since then has been extensively studied in the literature. A stochastic optimization approach assumes a probability distribution over the uncertain parameters and seeks to optimize the expected value of the objective function typically. We refer the reader to several textbooks including Infanger [22], Kall and Wallace [24], Prékopa [26], Shapiro [28], Shapiro et al. [29] and the references therein for a comprehensive review of stochastic optimization.

While the stochastic optimization approach has its merits and there has been reasonable progress in the field, there are two significant drawbacks of the approach.

(i) In a typical application, only historical data is available rather than probability distributions. Modeling uncertainty using a probability distribution is a choice we make, and it is not a primitive quantity fundamentally linked to the application.

(ii) More importantly, the stochastic optimization approach is by and large computationally intractable. Shapiro and Nemirovski [31] give hardness results for two-stage and multi-stage stochastic optimization problems where they show the multi-stage stochastic optimization is computationally intractable even if approximate solutions are desired. Dyer and Stougie [16] show that a multi-stage stochastic optimization problem where the distribution of uncertain parameters in any stage also depends on the decisions in past stages is PSPACE-hard.

To solve a two-stage stochastic optimization problem, Shapiro and Nemirovski [31] show that a sampling based algorithm provides approximate solutions given that a sufficiently large number of scenarios are sampled from the assumed distribution and the problem has a *relatively complete recourse*. A two-stage stochastic optimization problem is said to have a relatively complete recourse if for every first stage decision there is a feasible second stage recourse solution almost everywhere, i.e., with probability one (see the book by Shapiro et al. [29]). Shmoys and Swamy [32, 33] and Gupta et al. [20, 21] consider the two-stage and multi-stage stochastic set covering problem under certain restrictions on the objective coefficients, and propose sampling based algorithms that use a small number of scenario samples to construct a good first-stage decision. However, the stochastic set covering problem admits a complete recourse, i.e., for any first-stage decision there is a feasible recourse decision in each scenario, and the sampling based algorithms only work for problems with complete or relatively complete recourse.

More recently, the robust optimization approach has been considered to address optimization under uncertainty and has been studied extensively (see Ben-Tal and Nemirovski [6, 7, 8], El Ghaoui and Lebret [17], Goldfarb and Iyengar [19], Bertsimas and Sim [13], Bertsimas and Sim [14]). In a robust optimization approach, the uncertain parameters are assumed to belong to some uncertainty set and the goal is to construct a single (static) solution that is feasible for all possible realizations of the uncertain parameters from the set and optimizes the worst-case objective. We point the reader to the survey by Bertsimas et al. [9] and the book by Ben-Tal et al. [4] and the references therein for an extensive review of the literature in robust optimization. This approach is significantly more tractable as compared to a stochastic optimization approach and the robust problem is equivalent to the corresponding deterministic problem in computational complexity for a large class of problems and uncertainty sets [9]. However, the robust optimization approach has the following drawbacks. Since it optimizes over the worst-case realization of the uncertain parameters, it may produce highly conservative solutions that may not perform well in the expected case. Moreover, the robust approach computes a single (static) solution even for a multi-stage problem with several stages of decision-making as opposed a fully-adaptable solution where decisions in each stage depend on the actual realizations of the parameters in past stages. This may further add to the conservativeness.

Another approach is to consider solutions that are fully-adaptable in each stage and depend on the realizations of the parameters in past stages and optimize over the worst case. Such solution approaches have been considered in the literature and referred to as adjustable robust policies (see Ben-Tal et al. [5] and the book by Ben-Tal et al. [4] for a detailed discussion of these policies). Unfortunately, the adjustable robust problem is computationally intractable and Ben-Tal et al. [5] introduce an affinely adjustable robust solution approach to approximate the adjustable robust problem. Affine solution approaches (or just affine policies) were introduced in the context of stochastic programming (see Gartska and Wets [18] and Rockafeller and Wets [27]) and have been extensively studied in control theory (see the survey by Bemporad and Morari [3]). Affine policies are useful due to their computational tractability and strong empirical performance. Recently Bertsimas et al. [12] show that affine policies are optimal for a single dimension multi-stage problem with box constraints and box uncertainty sets. Bertsimas and Goyal [11] consider affine policies for the two-stage adaptive (or adjustable robust) problem and give a tight bound of $O(\sqrt{dim(\mathcal{U})})$ on the performance of affine policies with respect to a fully-adaptable solution where $dim(\mathcal{U})$ denotes the dimension of the uncertainty set.

In this paper, we consider a class of solutions called *finitely adaptable* solutions that were introduced by Bertsimas and Caramanis [10]. In this class of solutions, the decision-maker apriori computes a small number of solutions instead of just a single (static) solution such that for every possible realization of the uncertain parameters, at least one of them is feasible and in each stage, the decision-maker implements the best solution from the given set of solutions. Therefore, a finitely adaptable solution policy is a generalization of the static robust solution and is a middle ground between the static solution policy and the fully-adaptable policy. It can be thought of as a special case of a piecewise-affine policy where the each piece is a static solution instead of an affine solution. As compared to a fully-adaptable solution, which prescribes a solution for all possible scenarios (possibly an uncountable set), a finitely adaptable solution has only a small finite number of solutions. Therefore, the decision space is much sparser and it is significantly more tractable than the fully-adaptable solution. Furthermore, for each possible scenario, at least one of the finitely adaptable solutions is feasible. This makes it different from sampling based approaches where a small number of scenarios are sampled and an optimal decision is computed for only the sampled scenarios, while the rest of the scenarios are ignored. We believe that finitely adaptable solutions are consistent with how decisions are made in most real world problems. Unlike dynamic programming that prescribes an optimal decision for each (possibly uncountable) future state of the world, we make decisions for only a few states of the world.

We aim to analyze the performance of static robust and finitely adaptable solution policies for the two-stage and multi-stage stochastic optimization problems, respectively. We show that the performance of these solution approaches as compared to the optimal fully-adaptable stochastic solution depends on fundamental geometric properties of the uncertainty set including symmetry for a fairly general class of models. Bertsimas and Goyal [11] analyze the performance of static robust solution in two-stage stochastic problems for perfectly symmetric uncertainty sets such as ellipsoids, and norm-balls. We consider a generalized notion of symmetry of a convex set introduced in Minkowski [25], where the symmetry of a convex set is a number between 0 and 1. The symmetry of a set being equal to one implies that it is perfectly symmetric (such as an ellipsoid). We show that the performance of the static robust and the finitely adaptable solutions for the corresponding two-stage and multi-stage stochastic optimization problems depends on the symmetry of the uncertainty set. This is a two-fold generalization of the results in [11]. We extend the results in [11] of performance of a static robust solution in two-stage stochastic optimization problems, for general convex uncertainty sets using a generalized notion of symmetry. Furthermore, we also generalize the static robust solution policy to a finitely adaptable solution policy for the multi-stage stochastic optimization problem and give a similar bound on its performance that is related to the symmetry of the uncertainty sets. The results are quite general and extend to important cases such as integrality constraints on decision variables and linear conic inequalities. To the best of our knowledge, there were no approximation bounds for the multi-stage problem in such generality.

2. Models and preliminaries.

In this section, we first setup the models for two-stage and multi-stage stochastic, robust, and adaptive optimization problems considered in this paper. We also discuss the solution concept of finitely adaptability for multi-stage problems. Then we introduce the geometric quantity of symmetry of a convex compact set and some properties that will be used in the paper. Lastly, we define a quantity called translation factor of a convex set.

2.1 Two-stage optimization models.

Throughout the paper, we denote vectors and matrices using boldface letters. For instance, $\boldsymbol{x}$ denotes a vector, while α denotes a scalar. A two-stage stochastic optimization problem, Π^2_{Stoch}, is defined as,

$$\begin{aligned} z_{\text{Stoch}} := \min_{\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{b})} \quad & \boldsymbol{c}^T \boldsymbol{x} + \mathbb{E}_\mu[\boldsymbol{d}^T \boldsymbol{y}(\boldsymbol{b})] \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{b}) \geq \boldsymbol{b}, \ \mu\text{-a.e. } \boldsymbol{b} \in \mathcal{U} \subseteq \mathbb{R}^m_+, \\ & \boldsymbol{x} \in \mathbb{R}^{p_1} \times \mathbb{R}^{n_1 - p_1}_+, \\ & \boldsymbol{y}(\boldsymbol{b}) \in \mathbb{R}^{p_2} \times \mathbb{R}^{n_2 - p_2}_+, \ \forall \boldsymbol{b} \in \mathcal{U}. \end{aligned} \tag{2.1}$$

Here, the first-stage decision variable is denoted as $\boldsymbol{x}$; the second-stage decision variable is $y(\boldsymbol{b})$ for $\boldsymbol{b} \in \mathcal{U}$, where $\boldsymbol{b}$ is the uncertain right-hand side with the uncertainty set denoted as $\mathcal{U}$. The optimization in (2.1) for the second stage decisions $y(\cdot)$ is performed over the space of piecewise affine functions, since there are finitely many bases of the system of linear inequalities $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{b}) \geq \boldsymbol{b}$. Note that some of the decision variables in both the first and second stage are free, i.e., not constrained to be non-negative. A probability measure μ is defined on $\mathcal{U}$. Both $\boldsymbol{A}$ and $\boldsymbol{B}$ are certain, and there is no restriction on the coefficients in $\boldsymbol{A}, \boldsymbol{B}$ or on the objective coefficients $\boldsymbol{c}, \boldsymbol{d}$. The linear constraints $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{b}) \geq \boldsymbol{b}$ hold for μ almost everywhere on $\mathcal{U}$, i.e., the set of $\boldsymbol{b} \in \mathcal{U}$ for which the linear constraints are not satisfied has measure zero. We use the notation μ-a.e. $\boldsymbol{b} \in \mathcal{U}$ to denote this.

The key assumption in the above problem is that the uncertainty set $\mathcal{U}$ is contained in the nonnegative orthant. In addition, we make the following two technical assumptions.

(i) z_{Stoch} is finite. This assumption implies that $z_{\text{Stoch}} \geq 0$. Since if $z_{\text{Stoch}} < 0$, then it must be unbounded from below as we can scale the solution by an arbitrary positive factor and still get a feasible solution.

(ii) For any first-stage solution $\boldsymbol{x}$, and a second-stage solution $\boldsymbol{y}(\cdot)$ that is feasible for μ almost everywhere on $\mathcal{U}$, $\mathbb{E}_\mu[\boldsymbol{y}(\boldsymbol{b})]$ exists.

These technical conditions are assumed to hold for all the multi-stage stochastic models considered in the paper as well. We would like to note that since $\boldsymbol{A}$ is not necessarily equal to $\boldsymbol{B}$, our model does not admit relatively complete recourse.

A two-stage adaptive optimization problem, Π^2_{Adapt}, is given as,

$$\begin{aligned} z_{\text{Adapt}} := \min_{\boldsymbol{x},\boldsymbol{y}(\boldsymbol{b})} \quad & \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{b}\in\mathcal{U}} \boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{b}) \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y}(\boldsymbol{b}) \geq \boldsymbol{b},\ \forall \boldsymbol{b} \in \mathcal{U} \subseteq \mathbb{R}^m_+, \\ & \boldsymbol{x} \in \mathbb{R}^{p_1} \times \mathbb{R}^{n_1-p_1}_+, \\ & \boldsymbol{y}(\boldsymbol{b}) \in \mathbb{R}^{p_2} \times \mathbb{R}^{n_2-p_2}_+,\ \forall \boldsymbol{b} \in \mathcal{U}. \end{aligned} \tag{2.2}$$

Note that the above problem is also referred to as an adjustable robust problem in the literature (see Ben-Tal et al. [5] and the book by Ben-Tal et al. [4]).

The corresponding static robust optimization problem, Π_{Rob}, is defined as,

$$\begin{aligned} z_{\text{Rob}} := \min_{\boldsymbol{x},\boldsymbol{y}} \quad & \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{d}^T\boldsymbol{y} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{y} \geq \boldsymbol{b},\ \forall \boldsymbol{b} \in \mathcal{U} \subseteq \mathbb{R}^m_+, \\ & \boldsymbol{x} \in \mathbb{R}^{p_1} \times \mathbb{R}^{n_1-p_1}_+, \\ & \boldsymbol{y} \in \mathbb{R}^{p_2} \times \mathbb{R}^{n_2-p_2}_+. \end{aligned} \tag{2.3}$$

Note that any feasible solution to the robust problem (2.3) is also feasible for the adaptive problem (2.2), and thus $z_{\text{Adapt}} \leq z_{\text{Rob}}$. Moreover, any feasible solution to the adaptive problem (2.2) is also feasible for the stochastic problem (2.1) and in addition, we have

$$\boldsymbol{c}^T\boldsymbol{x} + \mathbb{E}_\mu[\boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{b})] \leq \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{b}\in\mathcal{U}} \boldsymbol{d}^T\boldsymbol{y}(\boldsymbol{b}),$$

leading to $z_{\text{Stoch}} \leq z_{\text{Adapt}}$. Hence, we have $z_{\text{Stoch}} \leq z_{\text{Adapt}} \leq z_{\text{Rob}}$.

We would like to note that when the left hand side of the constraints, i.e., $\boldsymbol{A}$ and $\boldsymbol{B}$ are uncertain, even a two-stage, two-dimensional problem can have an unbounded gap between the static robust solution and the stochastic solution.

2.2 Multi-stage optimization models. For multi-stage problems, we consider a fairly general model where the evolution of the multi-stage uncertainty is given by a directed acyclic network $G = (\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ is the set of nodes corresponding to different uncertainty sets, and $\mathcal{A}$ is the set of arcs that describe the evolution of uncertainty from Stage k to $(k+1)$ for all $k = 2, \ldots, K-1$. In each Stage $(k+1)$ for $k = 1, \ldots, K-1$, the uncertain parameters $\boldsymbol{u}_k$ belong to one of the N_k uncertainty sets,

$\mathcal{U}_1^k, \ldots, \mathcal{U}_{N_k}^k \subset \mathbb{R}_+^m$. We also assume that the probability distribution of $\boldsymbol{u}_k$ conditioned on the fact that $\boldsymbol{u}_k \in \mathcal{U}_j^k$ for all $k = 2, \ldots, K$, $j = 1, \ldots, N_k$ is known. In our notation, the multi-stage uncertainty network starts from Stage 2 (and not Stage 1) and we refer to the uncertainty sets and uncertain parameters in Stage $(k+1)$ using index k. Therefore, in a K-stage problem, the index $k \in \{1, \ldots, K-1\}$.

In Stage 2, there is a single node $\mathcal{U}^1$, which we refer to as the *root node*. Therefore, $N_1 = 1$. For any $k = 1, \ldots, K-1$, suppose the uncertain parameters $\boldsymbol{u}_k$ in Stage $(k+1)$, belongs to $\mathcal{U}_j^k$ for some $j = 1, \ldots, N_k$. Then for any edge from $\mathcal{U}_j^k$ to $\mathcal{U}_{j'}^{k+1}$ in the directed network, $\boldsymbol{u}_{k+1} \in \mathcal{U}_{j'}^{k+1}$ with probability $p_{j,j'}^k$, which is an observable event in Stage $(k+2)$. In other words, in Stage $(k+2)$, we can observe which state transition happened in stage $(k+1)$. Therefore, at every stage, we know the realizations of the uncertain parameters in past stages as well as the path of the uncertainty evolution in G and the decisions in each stage depend on both of these. Note that since we also observe the path of the uncertainty evolution (i.e. what edges in the directed network were realized in each stage), the uncertainty sets in a given stage need not be disjoint. This model of uncertainty is a generalization of the scenario tree model often used in stochastic programming (see Shapiro et al. [30]) where the multi-stage uncertainty is described by a tree. If the directed acyclic network G in our model is a tree and each uncertainty set is a singleton, our model reduces to a scenario tree model described in [30].

The evolution of the multi-stage uncertainty is illustrated in Figure 1. We would like to note that this is a very general model of multi-stage uncertainty. For instance, consider a multi-period inventory management problem, where the demand is uncertain. In each stage, we observe the realized demand and also a signal from the market about the next period demand such as the weekly retail sales index. In our model, the observed market signals correspond to the path in the multi-stage uncertainty network and the observed demand is the actual realization of the uncertain parameters. As another example, consider a multi-period asset management problem with uncertain asset returns. In each period, we observe the asset returns and also a market signal such as S&P 500 or NASDAQ indices. Again, the market signals correspond to the path in the multi-stage uncertainty network in our model and observed asset returns are the realization of uncertain parameters in the model.

Let $\mathcal{P}$ denote the set of directed paths in G from the root node, $\mathcal{U}_1^1$, to node $\mathcal{U}_j^{K-1}$, for all $j = 1, \ldots, N_{K-1}$. We denote any path $\boldsymbol{P} \in \mathcal{P}$ as an ordered sequence, $(j_1, \ldots, j_{K-1})$ of the indices of the uncertainty sets in each stage that occur on $\boldsymbol{P}$. Let $\boldsymbol{\Omega}(\boldsymbol{P})$ denote the set of possible realizations of the

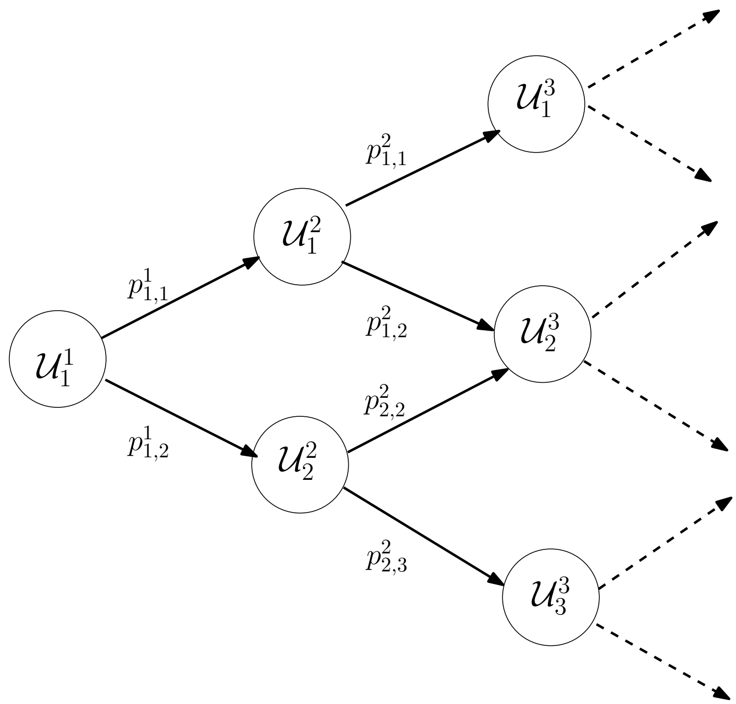

Figure 1: Illustration of the evolution of uncertainty in a multi-stage problem.

multi-stage uncertainty from uncertainty sets in $\boldsymbol{P}$. For any $\boldsymbol{P} = (j_1, \ldots, j_{K-1}) \in \mathcal{P}$,

$$\boldsymbol{\Omega}(\boldsymbol{P}) = \left\{\boldsymbol{\omega} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{K-1}) \;\middle|\; \boldsymbol{b}_k \in \mathcal{U}_{j_k}^k,\ \forall k = 1, \ldots, K-1\right\}. \tag{2.4}$$

For any $\boldsymbol{P} = (j_1, \ldots, j_{K-1}) \in \mathcal{P}$, and $k = 1, \ldots, K-1$, let $\boldsymbol{P}[k]$ denote the index sequence of Path $\boldsymbol{P}$ from Stage 2 to Stage $(k+1)$, i.e., $\boldsymbol{P}[k] = (j_1, \ldots, j_k)$. Let $\mathcal{P}[k] = \{\boldsymbol{P}[k] \mid \boldsymbol{P} \in \mathcal{P}\}$. Also, for any $\boldsymbol{\omega} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{K-1}) \in \boldsymbol{\Omega}(\boldsymbol{P})$, let $\boldsymbol{\omega}[k]$ denote the subsequence of first k elements of $\boldsymbol{\omega}$, i.e., $\boldsymbol{\omega}[k] = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k)$.

We define a probability measure for the multi-stage uncertainty as follows. For any $k = 1, \ldots, K-1$, $j = 1, \ldots, N_k$, let μ_j^k be a probability measure defined on $\mathcal{U}_j^k$ that is independent of the probability measures over other uncertainty sets in the network. Therefore, for any $\boldsymbol{P} = (j_1, \ldots, j_{K-1}) \in \mathcal{P}$, the probability measure $\mu_{\boldsymbol{P}}$ over the set $\Omega(\boldsymbol{P})$ is the product measure of $\mu_{j_k}^k$, $k = 1, \ldots, K-1$. Moreover, the probability that the uncertain parameters realize from path $\boldsymbol{P}$ is given by $\prod_{k=1}^{K-2} p_{j_k, j_{k+1}}^k$. This defines a probability measure on the set of realizations of the multi-stage uncertainty, $\bigcup_{\boldsymbol{P} \in \mathcal{P}} \Omega(\boldsymbol{P})$.

We can now formulate the K-stage stochastic optimization problem Π_{Stoch}^K, where the right-hand side of the constraints is uncertain, as follows.

$$\begin{aligned} z_{\text{Stoch}}^K = \min\ & \boldsymbol{c}^T\boldsymbol{x} + \sum_{k=1}^{K-1} \mathbb{E}_{\boldsymbol{P}\in\mathcal{P}}\left[\mathbb{E}_{\mu_{\boldsymbol{P}}}\left[\boldsymbol{d}_k^T\boldsymbol{y}_k(\boldsymbol{\omega}[k],\boldsymbol{P}[k])\right]\right] \\ \text{s.t. } & \forall \boldsymbol{P}\in\mathcal{P},\ \mu_{\boldsymbol{P}}\text{-a.e. } \boldsymbol{\omega}=(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_{K-1})\in\boldsymbol{\Omega}(\boldsymbol{P}) \\ & \boldsymbol{A}\boldsymbol{x} + \sum_{k=1}^{K-1}\boldsymbol{B}_k\boldsymbol{y}_k(\boldsymbol{\omega}[k],\boldsymbol{P}[k]) \geq \sum_{k=1}^{K-1}\boldsymbol{b}_k, \\ & \boldsymbol{x}\in\mathbb{R}^{p_1}\times\mathbb{R}_+^{n_1-p_1}, \\ & \boldsymbol{y}_k(\boldsymbol{\omega}[k],\boldsymbol{P}[k])\in\mathbb{R}^{p_k}\times\mathbb{R}_+^{n_k-p_k},\ \forall k=1,\ldots,K-1, \end{aligned} \tag{2.5}$$

There is no restriction on the coefficients in the constraint matrices $\boldsymbol{A},\boldsymbol{B}_k$ or on the objective coefficients $\boldsymbol{c},\boldsymbol{d}_k, k=1,\ldots,K-1$. However, we require that each uncertainty set $\mathcal{U}_j^k\subseteq\mathbb{R}_+^m$ for all $j=1,\ldots,N_k$, $k=1,\ldots,K-1$. As mentioned before, we assume that z_{Stoch}^K is finite and the expectation of every feasible multi-stage solution exists.

We also formulate the K-stage adaptive optimization problem Π_{Adapt}^K as follows.

$$\begin{aligned} z_{\text{Adapt}}^K = \min\ & \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{P}\in\mathcal{P},\boldsymbol{\omega}\in\Omega(\boldsymbol{P})}\ \min_{\boldsymbol{y}_k(\boldsymbol{\omega}[k],\boldsymbol{P}[k]),k=1,\ldots,K-1}\ \sum_{k=1}^{K-1}\boldsymbol{d}_k^T\boldsymbol{y}_k(\boldsymbol{\omega}[k],\boldsymbol{P}[k]) \\ \text{s.t. } & \forall \boldsymbol{P}\in\mathcal{P},\ \forall\boldsymbol{\omega}=(\boldsymbol{b}_1,\ldots,\boldsymbol{b}_{K-1})\in\boldsymbol{\Omega}(\boldsymbol{P}) \\ & \boldsymbol{A}\boldsymbol{x} + \sum_{k=1}^{K-1}\boldsymbol{B}_k\boldsymbol{y}_k(\boldsymbol{\omega}[k],\boldsymbol{P}[k]) \geq \sum_{k=1}^{K-1}\boldsymbol{b}_k, \\ & \boldsymbol{x}\in\mathbb{R}^{p_1}\times\mathbb{R}_+^{n_1-p_1}, \\ & \boldsymbol{y}_k(\boldsymbol{\omega}[k],\boldsymbol{P}[k])\in\mathbb{R}^{p_k}\times\mathbb{R}_+^{n_k-p_k},\ \forall k=1,\ldots,K-1. \end{aligned} \tag{2.6}$$

We also formulate the K-stage stochastic optimization problem $\Pi_{\text{Stoch}(\boldsymbol{b},\boldsymbol{d})}^K$, where both the right-hand side and the objective coefficients are uncertain, as follows. The subscript Stoch$(\boldsymbol{b},\boldsymbol{d})$ in the subscript of the problem name denotes that both the right hand side $\boldsymbol{b}$ and the objective coefficient $\boldsymbol{d}$ are uncertain. Here, $\boldsymbol{\omega}$ denotes the sequence of uncertain right-hand side and objective coefficients realizations and for any $k=1,\ldots,K-1$, $\boldsymbol{\omega}[k]$ denotes the subsequence of first k elements. Also, for any $\boldsymbol{P}\in\mathcal{P}$, the measure $\mu_{\boldsymbol{P}}$ is the product measure of the measures on the uncertainty sets in Path $\boldsymbol{P}$.

$$\begin{aligned} z_{\text{Stoch}(\boldsymbol{b},\boldsymbol{d})}^K = \min\quad & \boldsymbol{c}^T\boldsymbol{x} + \sum_{k=1}^{K-1} \mathbb{E}_{\boldsymbol{P}\in\mathcal{P}}\left[\mathbb{E}_{\mu_{\boldsymbol{P}}}\left[\boldsymbol{d}_k^T\boldsymbol{y}_k(\boldsymbol{\omega}[k],\boldsymbol{P}[k])\right]\right] \\ \text{s.t. } & \forall \boldsymbol{P}\in\mathcal{P},\mu_{\boldsymbol{P}}\text{-a.e. } \boldsymbol{\omega}=\big((\boldsymbol{b}_1,\boldsymbol{d}_1),\ldots,(\boldsymbol{b}_{K-1},\boldsymbol{d}_{K-1})\big)\in\boldsymbol{\Omega}(\boldsymbol{P}) \\ & \boldsymbol{A}\boldsymbol{x} + \sum_{k=1}^{K-1}\boldsymbol{B}_k\boldsymbol{y}_k(\boldsymbol{\omega}[k],\boldsymbol{P}[k]) \geq \sum_{k=1}^{K-1}\boldsymbol{b}_k, \\ & \boldsymbol{x}\in\mathbb{R}^{p_1}\times\mathbb{R}_+^{n_1-p_1}, \\ & \boldsymbol{y}_k(\boldsymbol{\omega}[k],\boldsymbol{P}[k])\in\mathbb{R}^{p_k}\times\mathbb{R}_+^{n_k-p_k},\ \forall k=1,\ldots,K-1. \end{aligned} \tag{2.7}$$

Also, we can formulate the K-stage adaptive problem $\Pi^K_{\text{Adapt}(\boldsymbol{b},\boldsymbol{d})}$,

$$\begin{aligned} z^K_{\text{Adapt}(\boldsymbol{b},\boldsymbol{d})} = \min \quad & \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{P}\in\mathcal{P},\boldsymbol{\omega}\in\Omega(\boldsymbol{P})} \ \min_{\boldsymbol{y}_k(\boldsymbol{\omega}[k],\boldsymbol{P}[k]),k=1,\ldots,K-1} \sum_{k=1}^{K-1} \boldsymbol{d}_k^T \boldsymbol{y}_k(\boldsymbol{\omega}[k],\boldsymbol{P}[k]) \\ \text{s.t.} \ & \forall \boldsymbol{P}\in\mathcal{P},\ \forall \boldsymbol{\omega} = \big((\boldsymbol{b}_1,\boldsymbol{d}_1),\ldots,(\boldsymbol{b}_{K-1},\boldsymbol{d}_{K-1})\big)\in\boldsymbol{\Omega} \\ & \boldsymbol{A}\boldsymbol{x} + \sum_{k=1}^{K} \boldsymbol{B}_k \boldsymbol{y}_k(\boldsymbol{\omega}[k],\boldsymbol{P}[k]) \geq \sum_{k=1}^{K} \boldsymbol{b}_k, \\ & \boldsymbol{x}\in\mathbb{R}^{p_1}\times\mathbb{R}^{n_1-p_1}_+, \\ & \boldsymbol{y}_k(\boldsymbol{\omega}[k],\boldsymbol{P}[k]) \in \mathbb{R}^{p_k}\times\mathbb{R}^{n_k-p_k}_+,\ \forall k=1,\ldots,K-1. \end{aligned} \tag{2.8}$$

In Section 7, we consider an extension to the case where the constraints are general linear conic inequalities and the right hand side uncertainty set is a convex and compact subset of the underlying cone. Furthermore, in Section 8, we also consider extensions of the above two-stage and multi-stage models where some decision variables in each stage are integer.

2.3 Examples.

In this section, we show two classical problems that can be formulated in our framework to illustrate the applicability of our models.

Multi-period inventory management problem. We show that we can model the classical multi-period inventory management problem as a special case of (2.5). In a classical single-item inventory management problem, the goal in each period is to decide on the quantity of the item to order under an uncertain future period demand. In each period, each unit of excess inventory incurs a holding cost and each unit of backlogged demand incurs a per-unit penalty cost and the goal is to make ordering decisions such that the sum of expected holding and backorder-penalty cost is minimized.

As an example, we model a 3-stage problem in our framework. Let $\boldsymbol{\Omega}$ denote the set of demand scenarios with a probability measure μ, x denote the initial inventory, $y_k(b_1,\ldots,b_k)$ denote the backlog and $z_k(b_1,\ldots,b_k)$ denote the order quantity in Stage $(k+1)$ when the first k-period demand is $(b_1,\ldots,b_k)$. Let h_k denote the per unit holding cost and p_k denote the per-unit backlog penalty in Stage $(k+1)$. We model the 3-stage inventory management problem as follows where the decision variables are x, $y_1(b_1), y_2(b_1,b_2)$ and $z_1(b_1)$ for all $(b_1,b_2)\in\mathcal{U}$.

$$\begin{aligned} \min \quad & \mathbb{E}_{b_1}\Big[h_1\big(x+y_1(b_1)-b_1\big)+p_1y_1(b_1)+\mathbb{E}_{b_2|b_1}\big[h_2\big(x+z_1(b_1)+y_2(b_1,b_2)-(b_1+b_2)\big)+p_2y_2(b_1,b_2)\big]\Big] \\ \text{s.t.} \quad & x+y_1(b_1)\geq b_1,\ \mu\text{-a.e. } (b_1,b_2)\in\Omega \\ & x+z_1(b_1)+y_2(b_1,b_2)\geq b_1+b_2,\ \mu\text{-a.e. } (b_1,b_2)\in\Omega \\ & x, y_1(b_1), y_2(b_1,b_2), z_1(b_1)\geq 0. \end{aligned}$$

The above formulation can be generalized to a multi-period problem in a straightforward manner. We can also generalize to multi-item variants of the problem. However, we would like to note that we can not model a capacity constraint on the order quantity, since we require that the constraints are of "greater than or equal to" form with nonnegative right-hand sides. Nevertheless, the formulation is fairly general even with this restriction.

Capacity planning under demand uncertainty. In many important applications, we encounter the problem of planning for capacity to serve an uncertain future demand. For instance, in a call center, staffing capacity decisions need to be made well in advance of the realization of the uncertain demand. In electricity grid operations, generator unit-commitment decisions are required to be made at least a day ahead of the realization of the uncertain electricity demand because of the large startup time of the generators. In a facility location problem with uncertain future demand, the facilities need to be opened well in advance of the realized demand. Therefore, the capacity planning problem under demand uncertainty is important and widely applicable.

We show that this can be modeled in our framework using the example of the facility location problem. For illustration, we use a 2-stage problem. Let $\mathcal{F}$ denote the set of facilities and $\mathcal{D}$ denote the set of demand points. For each facility $i \in \mathcal{F}$, let x_i be an integer decision variable that denotes the capacity for facility i. For each point $j \in \mathcal{D}$, let b_j denote the uncertain future demand and let $y_{ij}(\boldsymbol{b})$ denote the amount of demand of j assigned to facility i when the demand vector is $\boldsymbol{b}$. Also, let d_{ij} denote the cost of assigning a unit demand from point j to facility i and let c_i denote the per-unit capacity cost of facility i. Therefore, we can formulate the problem as follows. Let $\mathcal{U}$ be the uncertainty set for the demand and let μ be a probability measure defined on $\mathcal{U}$.

$$
\begin{aligned}
\min \quad & \sum_{i \in \mathcal{F}} c_i x_i + \mathbb{E}_\mu \left[\sum_{j \in \mathcal{D}} d_{ij} y_{ij}(\boldsymbol{b}) \right] \\
\text{s.t.} \quad & \sum_{i \in \mathcal{F}} y_{ij}(\boldsymbol{b}) \geq b_j, \ \mu\text{-a.e. } \boldsymbol{b} \in \mathcal{U}, \ \forall j \in \mathcal{D} \\
& x_i - \sum_{j \in \mathcal{D}} y_{ij}(\boldsymbol{b}) \geq 0, \ \mu\text{-a.e. } \boldsymbol{b} \in \mathcal{U}, \ \forall i \in \mathcal{F} \\
& x_i \in \mathbb{Z}_+, \ \forall i \in \mathcal{F} \\
& y_{ij}(\boldsymbol{b}) \in \mathbb{R}_+, \ \forall i \in \mathcal{F}, j \in \mathcal{D}, \mu\text{-a.e. } \boldsymbol{b} \in \mathcal{U}.
\end{aligned}
$$

2.4 Finitely adaptable solutions.

We consider a *finitely adaptable* class of solutions for the multi-stage stochastic and adaptive optimization problems described above. This class of solutions was introduced by Bertsimas and Caramanis [10] where the decision-maker computes a small set of solutions in each stage apriori.

A static robust solution policy specifies a single solution that is feasible for all possible realizations of the uncertain parameters. On the other extreme, a fully-adaptable solution policy specifies a solution for each possible realization of the uncertain parameters in past stages. Typically, the set of possible realizations of the uncertain parameters is uncountable which implies that an optimal fully-adaptable solution policy is a function from an uncountable set of scenarios to optimal decisions for each scenario and often suffers from the "curse of dimensionality". A finitely adaptable solution is a tractable approach that bridges the gap between a static robust solution and a fully-adaptable solution. In a general finitely adaptable solution policy, instead of computing an optimal decision for each scenario, we partition the scenarios into a small number of sets and compute a single solution for each possible set. The partitioning of the set of scenarios is problem specific and is chosen by the decision-maker. A finitely adaptable solution policy is a special case of a piecewise affine solution where the solution in each piece is a static solution.

For the multi-stage stochastic and adaptive problems (2.5) and (2.6), we consider the partition of the set of scenarios based on the realized path in the multi-stage uncertainty network. In particular, in each Stage $(k+1)$, the decision $\boldsymbol{y}_k$ depends on the path of the uncertainty realization until Stage $(k+1)$. Therefore, there are $|\mathcal{P}[k]|$ different solutions for each Stage $(k+1)$, $k = 1, \ldots, K-1$; one corresponding to each directed path from the root node to a node in Stage $(k+1)$ in the multi-stage uncertainty network. For any realization of uncertain parameters and the path in the uncertainty network, the solution policy implements the solution corresponding to the realized path. Figure 2 illustrates the number of solutions in the finitely adaptable solution for each stage and each uncertainty set in a multi-stage uncertainty network. In the example in Figure 2, there are following four directed paths from Stage 2 to Stage K $(K = 4)$: $\boldsymbol{P}_1 = (1,1,1), \boldsymbol{P}_2 = (1,1,2), \boldsymbol{P}_3 = (1,2,2), \boldsymbol{P}_4 = (1,2,3)$. We know that $\boldsymbol{P}_j[1] = (1)$ for all $j = 1, \ldots, 4$. Also, $\boldsymbol{P}_1[2] = \boldsymbol{P}_2[2] = (1,1)$ and $\boldsymbol{P}_3[2] = \boldsymbol{P}_4[2] = (1,2)$. Therefore, in a finitely adaptable solution, we have the following decision variables apart from $\boldsymbol{x}$.

$$\begin{aligned}
\text{Stage 2} &: \quad \boldsymbol{y}_1(1) \\
\text{Stage 3} &: \quad \boldsymbol{y}_2(1,1), \boldsymbol{y}_2(1,2) \\
\text{Stage 4} &: \quad \boldsymbol{y}_3(1,1,1), \boldsymbol{y}_3(1,1,2), \boldsymbol{y}_3(1,2,2), \boldsymbol{y}_3(1,2,3).
\end{aligned}$$

2.5 The symmetry of a convex set. Given a nonempty compact convex set $\mathcal{U} \subset \mathbb{R}^m$ and a point $\boldsymbol{u} \in \mathcal{U}$, we define the symmetry of $\boldsymbol{u}$ with respect to $\mathcal{U}$ as follows.

$$\mathbf{sym}(\boldsymbol{u}, \mathcal{U}) := \max\{\alpha \geq 0 : \boldsymbol{u} + \alpha(\boldsymbol{u} - \boldsymbol{u}') \in \mathcal{U}, \ \forall \boldsymbol{u}' \in \mathcal{U}\}. \tag{2.9}$$

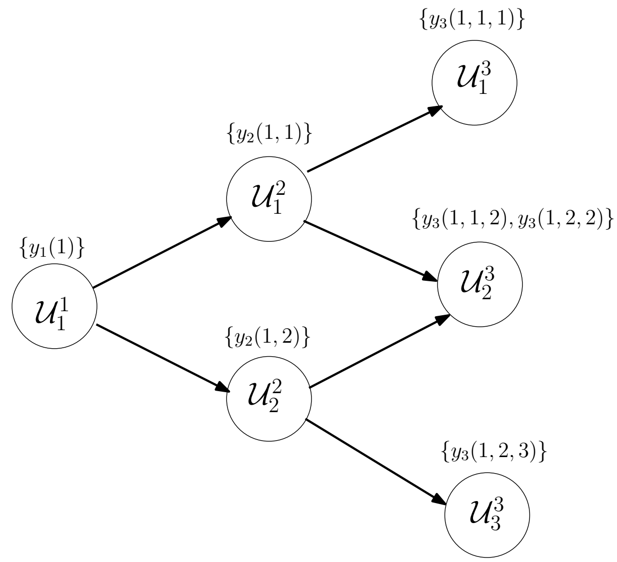

Figure 2: A Finitely Adaptable Solution for a 4-stage problem. For each uncertainty set, we specify the set of corresponding solutions in the finitely adaptable solution policy.

In order to develop a geometric intuition on $\mathbf{sym}(\boldsymbol{u}, \mathcal{U})$, we first show the following result. For this discussion, assume $\mathcal{U}$ is full-dimensional.

Lemma 2.1 *Let $\mathcal{L}$ be the set of lines in $\mathbb{R}^m$ passing through $\boldsymbol{u}$. For any line $\boldsymbol{l} \in \mathcal{L}$, let $\boldsymbol{u}_{\boldsymbol{l}}'$ and $\boldsymbol{u}_{\boldsymbol{l}}''$ be the points of intersection of $\boldsymbol{l}$ with the boundary of $\mathcal{U}$, denoted $\delta(\mathcal{U})$ (these exist as $\mathcal{U}$ is full-dimensional and compact). Then,*

a) $\mathbf{sym}(\boldsymbol{u}, \mathcal{U}) \leq \min\left(\dfrac{\|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}''\|}{\|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}'\|}, \dfrac{\|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}'\|}{\|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}''\|}\right)$,

b) $\mathbf{sym}(\boldsymbol{u}, \mathcal{U}) = \min\limits_{\boldsymbol{l} \in \mathcal{L}} \min\left(\dfrac{\|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}''\|}{\|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}'\|}, \dfrac{\|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}'\|}{\|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}''\|}\right)$.

Proof. Let $\mathbf{sym}(\boldsymbol{u}, \mathcal{U}) = \alpha_1$. Consider any $\boldsymbol{l} \in \mathcal{L}$ and consider $\boldsymbol{u}_{\boldsymbol{l}}' \in \delta(\mathcal{U})$ as illustrated in Figure 3. Let $\boldsymbol{u}_{\boldsymbol{l},r}' = \boldsymbol{u} + \alpha_1(\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}')$. From (2.9), we know that $\boldsymbol{u}_{\boldsymbol{l},r}' \in \mathcal{U}$. Note that $\boldsymbol{u}_{\boldsymbol{l},r}'$ is a *scaled reflection* of $\boldsymbol{u}_{\boldsymbol{l}}'$ about $\boldsymbol{u}$ by a factor α_1. Furthermore, $\boldsymbol{u}_{\boldsymbol{l},r}'$ lies on $\boldsymbol{l}$. Therefore, $\boldsymbol{u}_{\boldsymbol{l},r}' \in \mathcal{U}$ implies that

$$\|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l},r}'\| \leq \|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}''\| \Rightarrow \alpha_1 \cdot \|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}'\| \leq \|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}''\|,$$

which in turn implies that

$$\mathbf{sym}(\boldsymbol{u}, \mathcal{U}) = \alpha_1 \leq \frac{\|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}''\|}{\|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}'\|}.$$

Using a similar argument starting with $\boldsymbol{u}_{\boldsymbol{l}}''$ instead of $\boldsymbol{u}_{\boldsymbol{l}}'$, we obtain that

$$\mathbf{sym}(\boldsymbol{u}, \mathcal{U}) \leq \frac{\|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}'\|}{\|\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}''\|}.$$

To prove the second part, let

$$\alpha_2 = \min_{\boldsymbol{l}\in\mathcal{L}} \min\left(\frac{\|\boldsymbol{u}-\boldsymbol{u}_{\boldsymbol{l}}''\|}{\|\boldsymbol{u}-\boldsymbol{u}_{\boldsymbol{l}}'\|}, \frac{\|\boldsymbol{u}-\boldsymbol{u}_{\boldsymbol{l}}'\|}{\|\boldsymbol{u}-\boldsymbol{u}_{\boldsymbol{l}}''\|}\right).$$

From the above argument, we know that $\mathbf{sym}(\boldsymbol{u},\mathcal{U}) \leq \alpha_2$. Consider any $\boldsymbol{u}' \in \mathcal{U}$. We show that $(\boldsymbol{u} + \alpha_2(\boldsymbol{u} - \boldsymbol{u}')) \in \mathcal{U}$.

Let $\boldsymbol{l}$ denote the line joining $\boldsymbol{u}$ and $\boldsymbol{u}'$. Clearly, $\boldsymbol{l} \in \mathcal{L}$. Without loss of generality, suppose $\boldsymbol{u}'$ belongs to the line segment between $\boldsymbol{u}$ and $\boldsymbol{u}_{\boldsymbol{l}}'$. Therefore, $\boldsymbol{u} - \boldsymbol{u}' = \gamma(\boldsymbol{u} - \boldsymbol{u}_{\boldsymbol{l}}')$ for some $0 \leq \gamma \leq 1$ and

$$\|\boldsymbol{u}-\boldsymbol{u}'\| \leq \|\boldsymbol{u}-\boldsymbol{u}_{\boldsymbol{l}}'\|. \tag{2.10}$$

We know that

$$\alpha_2 \leq \frac{\|\boldsymbol{u}-\boldsymbol{u}_{\boldsymbol{l}}''\|}{\|\boldsymbol{u}-\boldsymbol{u}_{\boldsymbol{l}}'\|}$$

$$\Rightarrow \alpha_2\|\boldsymbol{u}-\boldsymbol{u}_{\boldsymbol{l}}'\| \leq \|\boldsymbol{u}-\boldsymbol{u}_{\boldsymbol{l}}''\|$$

$$\Rightarrow \alpha_2\|\boldsymbol{u}-\boldsymbol{u}'\| \leq \|\boldsymbol{u}-\boldsymbol{u}_{\boldsymbol{l}}''\|,$$

where the last inequality follows from (2.10). Therefore, $\boldsymbol{u} + \alpha_2\|\boldsymbol{u} - \boldsymbol{u}'\|$ belongs to the line segment between $\boldsymbol{u}$ and $\boldsymbol{u}_{\boldsymbol{l}}''$ and thus, belongs to $\mathcal{U}$. Therefore, $\mathbf{sym}(\boldsymbol{u},\mathcal{U}) \geq \alpha_2$. □

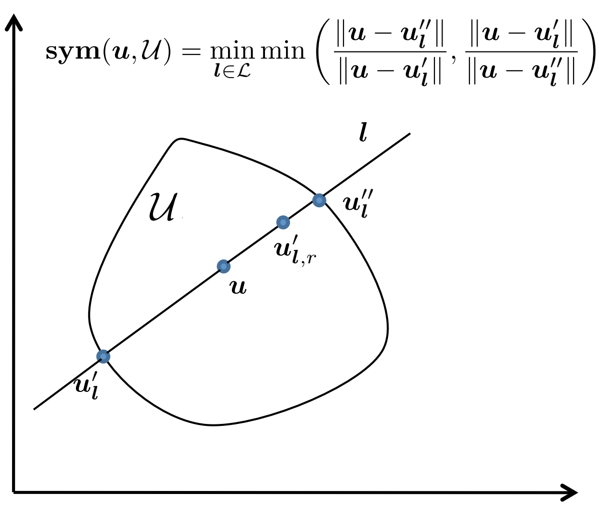

Figure 3: Geometric illustration of $\mathbf{sym}(\boldsymbol{u},\mathcal{U})$.

The symmetry of set $\mathcal{U}$ is defined as

$$\mathbf{sym}(\mathcal{U}) := \max\{\mathbf{sym}(\boldsymbol{u},\mathcal{U}) \mid \boldsymbol{u} \in \mathcal{U}\}. \tag{2.11}$$

An optimizer $\boldsymbol{u}_0$ of (2.11) is called a point of symmetry of $\mathcal{U}$. This definition of symmetry can be traced back to Minkowski [25] and is the first and the most widely used symmetry measure. Refer to Belloni and Freund [2] for a broad investigation of the properties of the symmetry measure defined in (2.9) and (2.11).

Note that the above definition generalizes the notion of perfect symmetry considered by Bertsimas and Goyal [11]. In [11], the authors define that a set $\mathcal{U}$ is symmetric if there exists $\boldsymbol{u}_0 \in \mathcal{U}$ such that, for any $\boldsymbol{z} \in \mathbb{R}^m$, $(\boldsymbol{u}_0 + \boldsymbol{z}) \in \mathcal{U} \Leftrightarrow (\boldsymbol{u}_0 - \boldsymbol{z}) \in \mathcal{U}$. Equivalently, $\boldsymbol{u} \in \mathcal{U} \Leftrightarrow (2\boldsymbol{u}_0 - \boldsymbol{u}) \in \mathcal{U}$. According to the definition in (2.11), $\mathbf{sym}(\mathcal{U}) = 1$ for such a set. Figure 4 illustrates symmetries of several interesting convex sets.

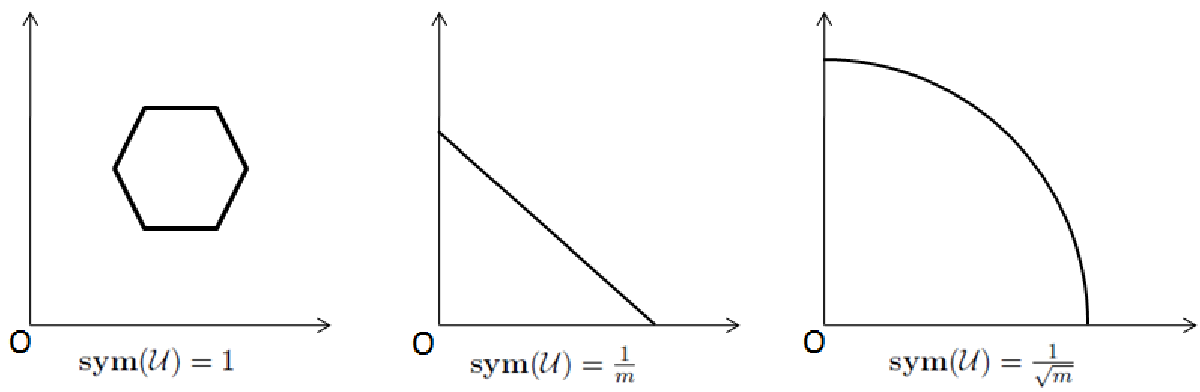

Figure 4: The figure on the left is a symmetric polytope with symmetry 1. The middle figure illustrates a standard simplex in $\mathbb{R}^m$ with symmetry $1/m$. The right figure shows the intersection of a Euclidean ball with $\mathbb{R}^m_+$, which has symmetry $1/\sqrt{m}$.

Lemma 2.2 (Belloni and Freund [2]) *For any nonempty convex compact set $\mathcal{U} \subseteq \mathbb{R}^m$, the symmetry of $\mathcal{U}$ satisfies,*

$$\frac{1}{m} \leq \mathbf{sym}(\mathcal{U}) \leq 1.$$

The symmetry of a convex set is at most 1, which is achieved for a perfectly symmetric set; and at least $1/m$, which is achieved by a standard simplex defined as $\Delta = \{\boldsymbol{x} \in \mathbb{R}^m_+ \mid \sum_{i=1}^m x_i \leq 1\}$. The lower bound follows from Löwner-John Theorem [23] (see Belloni and Freund [2]). The following lemma is used later in the paper.

Lemma 2.3 *Let $\mathcal{U} \subset \mathbb{R}^m_+$ be a convex and compact set such that $\boldsymbol{u}_0$ is the point of symmetry of $\mathcal{U}$. Then,*

$$\left(1 + \frac{1}{\mathbf{sym}(\mathcal{U})}\right) \cdot \boldsymbol{u}_0 \geq \boldsymbol{u}, \ \forall \boldsymbol{u} \in \mathcal{U}.$$

Proof. From the definition of symmetry in (2.11), we have that for any $\boldsymbol{u} \in \mathcal{U}$,

$$\boldsymbol{u}_0 + \mathbf{sym}(\mathcal{U})(\boldsymbol{u}_0 - \boldsymbol{u}) = (\mathbf{sym}(\mathcal{U}) + 1)\boldsymbol{u}_0 - \mathbf{sym}(\mathcal{U})\boldsymbol{u} \in \mathcal{U},$$

which implies that $(\mathbf{sym}(\mathcal{U}) + 1)\boldsymbol{u}_0 - \mathbf{sym}(\mathcal{U})\boldsymbol{u} \geq \mathbf{0}$ since $\mathcal{U} \subset \mathbb{R}^m_+$. □

2.6 The translation factor $\rho(u, \mathcal{U})$. For a convex compact set $\mathcal{U} \subset \mathbb{R}^m_+$, we define a *translation factor* $\rho(\boldsymbol{u}, \mathcal{U})$, the *translation factor* of $\boldsymbol{u} \in \mathcal{U}$ with respect to $\mathcal{U}$, as follows.

$$\rho(\boldsymbol{u}, \mathcal{U}) = \min\{\alpha \in \mathbb{R}_+ \mid \mathcal{U} - (1-\alpha) \cdot \boldsymbol{u} \subset \mathbb{R}^m_+\}.$$

In other words, $\mathcal{U}' := \mathcal{U} - (1-\rho)\boldsymbol{u}$ is the maximum possible translation of $\mathcal{U}$ in the direction $-\boldsymbol{u}$ such that $\mathcal{U}' \subset \mathbb{R}^m_+$. Figure 5 gives a geometric picture. Note that for $\alpha = 1$, $\mathcal{U} - (1-\alpha) \cdot \boldsymbol{u} = \mathcal{U} \subset \mathbb{R}^m_+$. Therefore, $0 < \rho \leq 1$. And ρ approaches 0, when the set $\mathcal{U}$ moves away from the origin. If there exists $\boldsymbol{u} \in \mathcal{U}$ such that $\boldsymbol{u}$ is at the boundary of $\mathbb{R}^m_+$, then $\rho = 1$. We denote

$$\rho(\mathcal{U}) := \rho(\boldsymbol{u}_0, \mathcal{U}),$$

where $\boldsymbol{u}_0$ is the symmetry point of the set $\mathcal{U}$. The following lemma is used later in the paper.

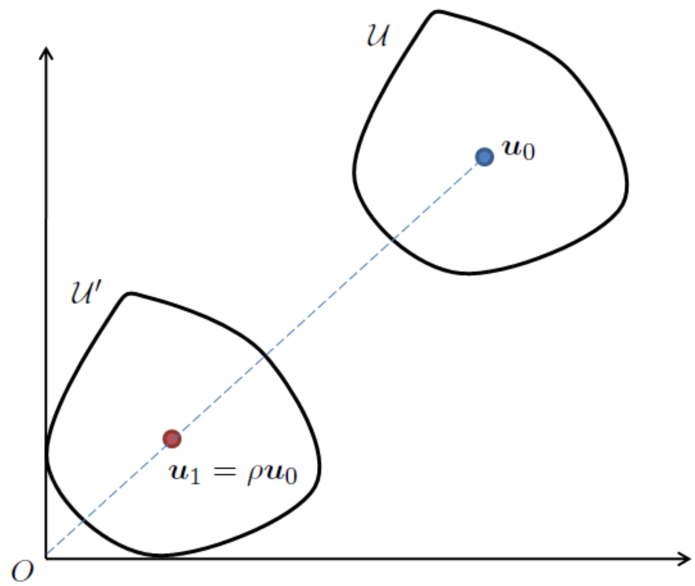

Figure 5: Geometry of the translation factor.

Lemma 2.4 *Let $\mathcal{U} \subset \mathbb{R}^m_+$ be a convex and compact set such that $\boldsymbol{u}_0$ is the point of symmetry of $\mathcal{U}$. Let $s = \mathbf{sym}(\mathcal{U}) = \mathbf{sym}(\boldsymbol{u}_0, \mathcal{U})$ and $\rho = \rho(\mathcal{U}) = \rho(\boldsymbol{u}_0, \mathcal{U})$. Then,*

$$\left(1 + \frac{\rho}{s}\right) \cdot \boldsymbol{u}_0 \geq \boldsymbol{u}, \ \forall \boldsymbol{u} \in \mathcal{U}.$$

Proof. Let $\mathcal{U}' = \mathcal{U} - (1-\rho)\boldsymbol{u}_0$. Let $\boldsymbol{u}_1 := \boldsymbol{u}_0 - (1-\rho)\boldsymbol{u}_0$. Also let $\boldsymbol{z} := \boldsymbol{u}_0 - \boldsymbol{u}_1 = (1-\rho)\boldsymbol{u}_0$. Figure 5 gives a geometric picture. Note that $\mathbf{sym}(\mathcal{U}') = \mathbf{sym}(\mathcal{U}) = s$. From Lemma 2.3, we know that

$$\left(1 + \frac{1}{s}\right)\boldsymbol{u}_1 \geq \boldsymbol{u}', \quad \forall \boldsymbol{u}' \in \mathcal{U}'.$$

Adding $\boldsymbol{z}$ on both sides, we have,

$$\begin{aligned}
&\left(1+\frac{1}{s}\right)\boldsymbol{u}_1+\boldsymbol{z}\geq\boldsymbol{u}, \quad \forall\boldsymbol{u}\in\mathcal{U},\\
\Rightarrow&\left(1+\frac{1}{s}\right)\rho\boldsymbol{u}_0+(1-\rho)\boldsymbol{u}_0\geq\boldsymbol{u}, \quad \forall\boldsymbol{u}\in\mathcal{U},\\
\Rightarrow&\left(1+\frac{\rho}{s}\right)\boldsymbol{u}_0\geq\boldsymbol{u}, \quad \forall\boldsymbol{u}\in\mathcal{U}.
\end{aligned}$$

□

3. Our contributions.

The contributions of this paper are two-fold. We present two significant generalizations of the model and results in Bertsimas and Goyal [11], where the authors characterize the performance of static robust solutions for two-stage stochastic and adaptive optimization problems under the assumption that the uncertainty sets are perfectly symmetric.

Firstly, we generalize the two-stage results to general uncertainty sets. We show that the performance of a static robust solution for two-stage stochastic and adaptive optimization problems depends on the general notion of symmetry (2.11) of the uncertainty set. The bounds are independent of the constraint matrices and any other problem data. Our bounds are also tight for all possible values of symmetry and reduce to the results in [11] for perfectly symmetric sets.

Secondly, we consider the multi-stage extensions of the two-stage models in [11] as described above and show that a class of finitely adaptable solutions which is a generalization of the static robust solution, is a good approximation for both the stochastic and the adaptive problem. The proof techniques are very general and easily extend to the case where some of the decision variables are integer constrained and the case where the constraints are linear conic inequalities for a general convex cone. To the best of our knowledge, these are the first performance bounds for the multi-stage problem in such generality.

Our main contributions are summarized below.

Stochastic optimization. For the two-stage stochastic optimization problem under right-hand side uncertainty, we show the following bound under a fairly general condition on the probability measure,

$$z_{\text{Rob}}\leq\left(1+\frac{\rho}{s}\right)\cdot z_{\text{Stoch}},$$

where $s=\mathbf{sym}(\mathcal{U})$ and $\rho=\rho(\mathcal{U})$ are the symmetry and the translation factor of the uncertainty set, respectively. Note that the above bound compares the cost of an optimal static solution with the expected cost of an optimal fully-adaptable two-stage stochastic solution and shows that the static cost is at most $(1+\rho/s)$ times the optimal stochastic cost. The performance of the static robust solution for the two-stage stochastic problem can possibly be even better. For the two-stage problem, we only implement the first-stage part of the static robust solution. For the second-stage, we compute an optimal solution after the uncertain parameters (right hand side in our case) is realized. Therefore, the expected second-stage cost for this solution policy is at most the second-stage cost of the static robust solution and the total expected cost is at most $z_{\text{Rob}} \leq (1+\rho/s) \cdot z_{\text{Stoch}}$. Since z_{Rob} is an upper bound on the expected cost of the solution obtained from an optimal static solution, the bound obtained by comparing z_{Rob} to the optimal stochastic cost is in fact a conservative bound.

For multi-stage problems, we show that a K-stage stochastic optimization problem, Π^K_{Stoch}, can be well approximated efficiently by a finitely adaptable solution. In particular, there is a finitely adaptable solution with at most $|\mathcal{P}|$ solutions that is a $(1+\rho/s)$-approximation of the original problem where s is the minimum symmetry over all sets in the uncertainty network and ρ is the maximum translation factor of any uncertainty set. Note that when all sets are perfectly symmetric, i.e., $s=1$, the finitely adaptable solution is a 2-approximation, generalizing the result of [11] to multi-stage. We also show that the bound of $(1+\rho/s)$ is tight.

Note that since $0<\rho\leq 1$ and $s \geq 1/m$,

$$\left(1+\frac{\rho}{s}\right) \leq (m+1),$$

which shows that, for a fixed dimension, the performance bound for robust and finitely adaptable solutions is *independent* of the particular data of an instance. This is surprising since when the left-hand side has uncertainty, i.e., $\boldsymbol{A}, \boldsymbol{B}$ are uncertain, even a two-stage two-dimensional problem can have an unbounded gap between the static robust solution and the stochastic solution as mentioned earlier. This also indicates that our assumptions on the model, namely, the right-hand side uncertainty (and/or cost uncertainty) and the uncertainty set contained in the nonnegative orthant, are tight. If these assumptions are relaxed, the performance gap becomes unbounded even for fixed dimensional uncertainty.

For the case when both cost and right-hand side are uncertain in $\Pi^K_{\text{Stoch}(\boldsymbol{b},\boldsymbol{d})}$, the performance of any finitely adaptable solution can be arbitrarily worse as compared to the optimal fully-adaptable stochastic solution. This result follows along the lines of arbitrary bad performance of a static robust solution in two-stage stochastic problems when both cost and right-hand sides are uncertain as shown in Bertsimas and Goyal [11].

Adaptive optimization. We show that for a multi-stage adaptive optimization problem, Π^K_{Adapt}, where only the right-hand side of the constraints is uncertain, the cost of a finitely adaptable solution is at most $(1+\rho/s)$ times the optimal cost of a fully-adaptable multi-stage solution, where s is the minimum symmetry of all sets in the uncertainty network and ρ is the maximum translation factor of the point of symmetry over all the uncertainty sets. This bound follows from the bound for the performance of a finitely adaptable solution for the multi-stage stochastic problem. Furthermore, if the uncertainty comes from hypercube sets, then a finitely adaptable solution with at most $|\mathcal{P}|$ solutions at each node of the uncertainty network is an optimal solution for the adaptive problem.

For the case when both cost and right-hand side are uncertain in $\Pi^K_{\text{Adapt}(\boldsymbol{b},\boldsymbol{d})}$, we show that the worst-case cost a finitely adaptable solution with at most $|\mathcal{P}|$ different solutions at each node of the uncertainty network is at most $(1+\rho/s)^2$ times the cost of an optimal fully-adaptable solution.

Extensions. We consider an extension of the above multi-stage models to the case where the constraints are linear conic inequalities and the uncertainty set belongs to the underlying cone. We also consider the case where some of the decision variables are constrained to be integers. Our proof techniques are quite general and the results extend to both these cases.

For the case of linear conic inequalities and the uncertainty set contained in the underlying convex cone, we show that a finitely adaptable (static robust) solution is a $(1+\rho/s)$-approximation for the multi-stage (two-stage respectively) stochastic and adaptive problem with right hand side uncertainty. The result also holds for the adaptive problem with both the right hand side and objective coefficient uncertainty and the performance bound for the finitely adaptable (static robust) solution is $(1+\rho/s)^2$ for the multi-stage (two-stage respectively) problem.

We also consider the case where some of the decision variables are integer constrained. For the multi-stage (two-stage) stochastic problem with right hand side uncertainty, if some of the first-stage decision variables are integer constrained, a finitely adaptable (static robust respectively) solution is a $\lceil(1+\rho/s)\rceil$ approximation with respect to an optimal fully-adaptable solution. For the multi-stage adaptive problem, we can handle integer constrained decision variables in all stages unlike the stochastic problem where we can handle integrality constraints only on the first stage decision variables. We show that for the multi-stage (two-stage) adaptive problem with right hand side uncertainty and integrality constraints on some of the decision variables in each stage, a finitely adaptable (static robust respectively) solution is a $\lceil(1+\rho/s)\rceil$-approximation. For the multi-stage (two-stage) adaptive problem with both right hand side and objective coefficient uncertainty and integrality constraints on variables, a finitely adaptable (static robust respectively) solution is a $\lceil(1+\rho/s)\rceil\cdot(1+\rho/s)$-approximation.

Outline. The rest of the paper is organized as follows. In Section 4, we present the performance bound of a static-robust solution that depends on the symmetry of the uncertainty set for the two-stage stochastic optimization problem. We also show that the bound is tight and present explicit bounds for several interesting and commonly used uncertainty sets. In Section 5, we present the finitely adaptable solution policy for the multi-stage stochastic optimization problem and discuss its performance bounds. In Section 6, we discuss the results for multi-stage adaptive optimization problems. In Sections 7 and 8, we present extensions of our results for the models with linear conic constraints and integrality constraints respectively.

4. Two-stage stochastic optimization problem.

In this section, we consider the two-stage stochastic optimization problem (2.1) and show that the performance of a static robust solution depends on the symmetry of the uncertainty set.

THEOREM 4.1 *Consider the two-stage stochastic optimization problem in* (2.1)*. Let* μ *be the probability measure on the uncertainty set* $\mathcal{U} \subset \mathbb{R}^m_+$*,* $\boldsymbol{b}_0$ *be the point of symmetry of* $\mathcal{U}$*, and* $\rho = \rho(\boldsymbol{b}_0, \mathcal{U})$ *be the translation factor of* $\boldsymbol{b}_0$ *with respect to* $\mathcal{U}$*. Denote* $s = \mathbf{sym}(\mathcal{U})$*. Assume the probability measure* μ *satisfies,*

$$\mathbb{E}_\mu[\boldsymbol{b}] \geq \boldsymbol{b}_0. \tag{4.1}$$

Then,

$$z_{\text{Rob}} \leq \left(1 + \frac{\rho}{s}\right) \cdot z_{\text{Stoch}}. \tag{4.2}$$

PROOF. From Lemma 2.4, we know that

$$\left(1 + \frac{\rho}{s}\right)\boldsymbol{b}_0 \geq \boldsymbol{b}, \quad \forall \boldsymbol{b} \in \mathcal{U}. \tag{4.3}$$

For brevity, let $\tau := (1+\rho/s)$. Suppose $(\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{b}) : \boldsymbol{b} \in \mathcal{U})$ is an optimal fully-adaptable stochastic solution. We first show that the solution, $(\tau\boldsymbol{x}, \tau\mathbb{E}[\boldsymbol{y}(\boldsymbol{b})])$, is a feasible solution to the robust problem (2.3). By the feasibility of $(\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{b})$ μ-a.e. $\boldsymbol{b} \in \mathcal{U})$, we have

$$\boldsymbol{A}(\tau\boldsymbol{x}) + \boldsymbol{B}(\tau\boldsymbol{y}(\boldsymbol{b})) \quad \geq \quad \tau\boldsymbol{b}, \; \mu - a.e.\boldsymbol{b} \in \mathcal{U}.$$

Taking expectation on both sides, we have

$$\boldsymbol{A}(\tau\boldsymbol{x}) + \boldsymbol{B}\big(\tau\mathbb{E}[\boldsymbol{y}(\boldsymbol{b})]\big) \geq \tau\mathbb{E}[\boldsymbol{b}] \geq \tau\boldsymbol{b}_0 \geq \boldsymbol{b}, \; \forall \boldsymbol{b} \in \mathcal{U},$$

where the second inequality follows from (4.1) and the last inequality follows from (4.3). Therefore, $(\tau\boldsymbol{x}, \tau\mathbb{E}[\boldsymbol{y}(\boldsymbol{b})])$ is a feasible solution for the robust problem (2.3) and,

$$z_{\text{Rob}} \leq \tau \cdot (\boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{d}^T\mathbb{E}[\boldsymbol{y}(\boldsymbol{b})]). \tag{4.4}$$

Also, by definition, $z_{\text{Stoch}} = \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{d}^T\mathbb{E}_{\boldsymbol{b}}[\boldsymbol{y}(\boldsymbol{b})]$, which implies that $z_{\text{Stoch}} \leq \tau \cdot z_{\text{Rob}}$ □

For brevity, we refer to (4.2) as the symmetry bound. The following comments are in order.

(i) The symmetry bound (4.2) is *independent* of the problem data $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{c}, \boldsymbol{d}$, depending only on the geometric properties of the uncertainty set $\mathcal{U}$, namely the symmetry and the translation factor of $\mathcal{U}$.

(ii) Theorem 4.1 can also be stated in a more general way, removing Assumption (4.1) on the probability measure, and use the symmetry of the expectation point. In particular, the bound (4.2) holds for $s = \mathbf{sym}(\mathbb{E}_\mu[\boldsymbol{b}], \mathcal{U})$ and $\rho = \rho(\mathbb{E}_\mu[\boldsymbol{b}], \mathcal{U})$.

However, we would like to note that (4.1) is a mild assumption, especially for symmetric uncertainty sets. Any symmetric probability measure on a symmetric uncertainty set satisfies this assumption. It also emphasizes the role that the symmetry of the uncertainty set plays in the bound.

(iii) As already mentioned in Section 3, a small relaxation from the assumptions of our model would cause unbounded performance gap. In particular, if the assumption, $\mathcal{U} \subset \mathbb{R}^m_+$, is relaxed, or the constraint coefficients are uncertain, the gap between z_{Rob} and z_{Stoch} cannot be bounded even in small dimensional problems. The following examples illustrate this fact.

(a) $\mathcal{U} \not\subseteq \mathbb{R}^m_+$: Consider the instance where $m = 1$ and $c = 0, d = 1, A = 0, B = 1$, the uncertainty set $\mathcal{U} = [-1, 1]$, and a uniform distribution on $\mathcal{U}$. The optimal stochastic solution has cost $z_{\text{Stoch}}(b) = 0$, while $z_{\text{Rob}}(b) = 2$. Thus, the gap is unbounded.

(b) $\boldsymbol{A}, \boldsymbol{B}$ uncertain: Consider the following instance taken from Ben-Tal et al. [4],

$$\begin{aligned} \min \quad & x \\ \text{s.t.} \quad & \begin{bmatrix} -\frac{1}{2}b \\ 1 \end{bmatrix} x + \begin{bmatrix} 1 \\ b-2 \end{bmatrix} y(b) \geq \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \forall b \in [0, r], \\ & x, y(b) \geq 0, \forall b \in [0, r], \end{aligned}$$

where $0 < r < 1$. From the constraints, we have that $x \geq 2/(1-r)$. Therefore, the optimal cost of a static solution is at least $2/(1-r)$. However, the optimal stochastic cost is at most 4. For details, refer to [4]. When r approaches 1, the gap tends to infinity.

4.1 Tightness of the bound.

In this section, we show that the bound given in Theorem 4.1 is tight. In particular, we show that for any given symmetry and translation factor, there exist a family of instances of the two-stage stochastic optimization problem such that the bound in Theorem 4.1 holds with equality.

THEOREM 4.2 *Given any symmetry $1/m \le s \le 1$ and translation factor, $0 < \rho \le 1$, there exist a family of instances such that $z_{\text{Rob}} = (1 + \rho/s) \cdot z_{\text{Stoch}}$.*

PROOF. For $p \ge 1$, let

$$\mathbf{B}_p^+ = \{\boldsymbol{b} \in \mathbb{R}_+^m \mid \|\boldsymbol{b}\|_p \le 1\}.$$

In Appendix A.2, we show that

$$\mathbf{sym}(\mathbf{B}_p^+) = \left(\frac{1}{m}\right)^{\frac{1}{p}},$$

and the symmetry point is,

$$\boldsymbol{b}_0(\mathbf{B}_p^+) = \frac{1}{m^{1/p}+1}\boldsymbol{e},$$

where $\boldsymbol{b}_0(\mathcal{U})$ denotes the symmetry point for any set $\mathcal{U}$. Also, let $(\mathbf{B}_p^+)' := \mathbf{B}_p^+ + r\boldsymbol{e}$, for some $r \ge 0$. Then, given any symmetry $s \ge 1/m$ and translation factor, $\rho \le 1$, we can find a $p \ge 1$ and $r \ge 0$ such that

$$s = \left(\frac{1}{m}\right)^{\frac{1}{p}}, \quad \rho = \frac{1}{(m^{1/p}+1)r+1}.$$

Now consider a problem instance where $\boldsymbol{A} = \mathbf{0}, \boldsymbol{B} = \boldsymbol{I}, \boldsymbol{c} = \mathbf{0}, \boldsymbol{d} = \boldsymbol{e}$, and a probability measure whose expectation is at the symmetry point of $(\mathbf{B}_p^+)'$. The optimal static robust solution is $\boldsymbol{y} = (r+1)\boldsymbol{e}$ and the optimal stochastic solution is $\boldsymbol{y}(\boldsymbol{b}) = \boldsymbol{b}$ for all $\boldsymbol{b} \in (\mathbf{B}_p^+)'$. Therefore,

$$z_{\text{Stoch}} = \left(r + \frac{1}{m^{1/p}+1}\right)m, \quad z_{\text{Rob}} = (r+1)m,$$

and,

$$\frac{z_{\text{Rob}}}{z_{\text{Stoch}}} = \frac{r+1}{\left(r + \frac{1}{m^{1/p}+1}\right)} = 1 + \frac{\rho}{s},$$

which shows the bound in Theorem 4.1 is tight. □

4.2 An alternative bound.

In this section, we present another performance bound on the robust solution for the two-stage stochastic problems, and compare it with the symmetry bound (4.2).

For an uncertainty set $\mathcal{U}$, let $\boldsymbol{b}^h$ as $b_j^h := \max_{\boldsymbol{b} \in \mathcal{U}} b_j$. Also, suppose the probability measure μ satisfies (4.1), i.e., $\mathbb{E}_\mu[\boldsymbol{b}] \ge \boldsymbol{b}_0$, where $\boldsymbol{b}_0$ is the symmetry point of $\mathcal{U}$. Let

$$\theta_s^* := \min\{\theta : \ \theta \cdot \boldsymbol{b}_0 \ge \boldsymbol{b}^h\}. \tag{4.5}$$

Using an argument similar to the proof of Theorem 4.1, we can show that the performance gap is at most θ_s^*, where the subscript s stands for stochasticity, i.e.,

$$z_{\text{Rob}} \leq \theta_s^* \cdot z_{\text{Stoch}}. \tag{4.6}$$

Let $s = \textbf{sym}(\boldsymbol{b}_0, \mathcal{U})$ and $\rho = \rho(\boldsymbol{b}_0, \mathcal{U})$. From Lemma (2.4), we also know that

$$\left(1 + \frac{\rho}{s}\right) \cdot \boldsymbol{b}_0 \geq \boldsymbol{b}, \; \forall \boldsymbol{b} \in \mathcal{U}.$$

Therefore, $\theta_s^* \leq (1 + \rho/s)$. So, θ_s^* is upper bounded by the symmetry bound obtained in Theorem 4.1. For brevity, we refer to (4.5) as a scaling bound. A geometric picture is given in Figure 6. As shown in the figure, $\overline{\boldsymbol{u}}$ is obtained from scaling $\mathbb{E}_\mu[\boldsymbol{b}]$ by a factor greater than one such that $\overline{\boldsymbol{u}}$ dominates all the points in $\mathcal{U}$, and θ_s^* is the smallest such factor.

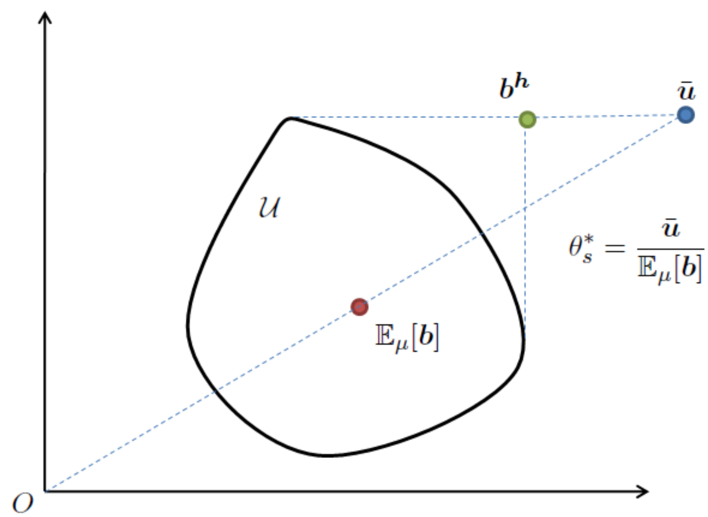

Figure 6: A geometric perspective on the stochasticity gap.

In the following section, we show that the scaling bound can be strictly tighter than the symmetry bound. On the other hand, the symmetry bound relates the performance of a static robust solution to key geometric properties of the uncertainty set. Furthermore, both bounds are equal for several interesting classes of uncertainty sets as discussed in the following section. In Theorem 4.2, we show that the symmetry bound is tight for any $\rho \leq 1$, $s \geq 1/m$ for a family of uncertainty sets.

The symmetry bound reveals several key qualitative properties of the performance gap that are difficult to see from the scaling bound. For example, for any symmetric uncertainty set, without the need to compute $\boldsymbol{b}^h$ and θ_s^*, we have a general bound of two on the stochasticity gap. Since symmetry of any convex set is at least $1/m$, the bound for any convex uncertainty set is at most $(m + 1)$. Both these bounds are not obvious from the scaling bound. Most importantly, in practice the symmetry bound gives an informative guideline in designing uncertainty sets where the robust solution has guaranteed performance.

4.3 Examples: Stochasticity gap for specific uncertainty sets. In this subsection, we give examples of specific uncertainty sets and characterize their symmetry. Both bounds on the stochasticity gap, the symmetry bound (4.2) as well the scaling bound (4.5), are presented in Table 1 for several interesting uncertainty sets. In most of the cases, the scaling bound is equal to the symmetry bound. The proofs of the symmetry computation of various uncertainty sets are deferred to the Appendix.

No.	Uncertainty set	Symmetry	Stochasticity gap
1	$\{\boldsymbol{b} : \\|\boldsymbol{b}\\|_p \leq 1, \boldsymbol{b} \geq \boldsymbol{0}\}$	$\frac{1}{m^{1/p}}$	$1 + m^{\frac{1}{p}}$ (a,b)
2	$\{\boldsymbol{b} : \\|\boldsymbol{b} - \bar{\boldsymbol{b}}\\|_p \leq 1, b_1 \geq \bar{b}_1\} \subset \mathbb{R}^m_+$ $1 \leq p < \infty$	$\frac{1}{2^{1/p}}$	$1 + \max_{1 \leq i \leq m} \bar{b}_i^{-1}$ (b)
3	$\{\boldsymbol{b} : \\|\boldsymbol{b} - \bar{\boldsymbol{b}}\\|_\infty \leq 1, b_1 \geq \bar{b}_1\} \subset \mathbb{R}^m_+$	1	$1 + \max_{1 \leq i \leq m} \bar{b}_i^{-1}$ (a,b)
4	$\{\boldsymbol{b} : \\|\boldsymbol{b} - \bar{\boldsymbol{b}}\\|_p \leq 1\} \subset \mathbb{R}^m_+$	1	$1 + \max_{1 \leq i \leq m} \bar{b}_i^{-1}$ (a,b)
5	$\{\boldsymbol{b} : \\|\boldsymbol{E}(\boldsymbol{b} - \bar{\boldsymbol{b}})\\|_2 \leq 1\} \subset \mathbb{R}^m_+$	1	$1 + \max_{1 \leq i \leq m} \frac{E_{ii}^{-1}}{\bar{b}_i}$ (a,b)
6	$\{\boldsymbol{b} : \\|\boldsymbol{b}\\|_{p_1} \leq 1, \\|\boldsymbol{b}\\|_{p_2} \leq r, \boldsymbol{b} \geq \boldsymbol{0}\}$	$\frac{1}{rm^{1/p_1}}$	$1 + rm^{\frac{1}{p_1}}$ (a,b)
7	Budgeted uncertainty set Δ_k $(1 \leq k \leq m)$	$\frac{k}{m}$	$1 + \frac{m}{k}$ (a,b)
8	Demand uncertainty set $\mathbf{DU}(\mu, \Gamma)$ $(\mu \geq \Gamma)$	1	$1 + \frac{\Gamma}{\mu}$ (a,b)
9	Demand uncertainty set $\mathbf{DU}(\mu, \Gamma)$ $(\frac{1}{\sqrt{m}}\Gamma < \mu < \Gamma)$	$\frac{\sqrt{m}\mu + \Gamma}{(1 + \sqrt{m})\Gamma}$	$1 + \frac{(1 + \sqrt{m})\Gamma}{\sqrt{m}\mu + \Gamma}$ (a,b)
10	Demand uncertainty set $\mathbf{DU}(\mu, \Gamma)$ $(0 \leq \mu \leq \frac{1}{\sqrt{m}}\Gamma)$	$\frac{\sqrt{m}\mu + \Gamma}{\sqrt{m}(\mu + \Gamma)}$	$1 + \frac{\sqrt{m}(\mu + \Gamma)}{\sqrt{m}\mu + \Gamma}$ (a,b)
11	$\{\boldsymbol{b} : L \leq \boldsymbol{e}^T\boldsymbol{b} \leq U, \boldsymbol{b} \geq \boldsymbol{0}\}$	$\frac{1}{m - \frac{L}{U}}$	$(m + 1) - \frac{L}{U}$ (a,b)
12	$\mathbf{conv}(\Delta_1, \{\boldsymbol{e}\})$	$\frac{1}{m - 1}$	m (a,b)
13	$\mathbf{conv}(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k)$	$\frac{1}{m} \leq s \leq 1$	$\max_{1 \leq i \leq m} \frac{b_i^h}{\bar{b}_i}$ (b)
14	$\{\boldsymbol{B} \in \mathbb{S}^m_+ \mid \boldsymbol{I} \bullet \boldsymbol{B} \leq 1\}$	$\frac{1}{m}$	$1 + m$ (a)

Table 1: Symmetry and corresponding stochasticity gap for various uncertainty sets. Note that footnote (a) or (b) means the tight bound is from the symmetry bound (4.2) or the scaling bound (4.6).

An L_p-ball intersected with the nonnegative orthant. We define,

$$\mathbf{B}_p^+ := \{\boldsymbol{b} \in \mathbb{R}_+^m \mid \|\boldsymbol{b}\|_p \leq 1\},$$

for $p \geq 1$. In Appendix A.2 we show the following.

$$\mathbf{sym}(\mathbf{B}_p^+) = \left(\frac{1}{m}\right)^{\frac{1}{p}}, \quad \boldsymbol{b}_0(\mathbf{B}_p^+) = \frac{1}{m^{\frac{1}{p}}+1}\boldsymbol{e}. \tag{4.7}$$

Therefore, if $\mathcal{U} = \mathbf{B}_p^+$ and the probability measure satisfies condition (4.1), then $z_{\text{Rob}} \leq (1+m^{\frac{1}{p}})z_{\text{Stoch}}$. The bound is tight as shown in Theorem 4.2. There are several interesting cases for $\mathbf{B}_p^+$ uncertainty sets. In particular,

(i) For $p = 1$, $\mathbf{B}_p^+$ is the standard simplex centered at the origin. The symmetry is $1/m$ and stochasticity gap is $m+1$.

(ii) For $p = 2$, $\mathbf{B}_p^+$ is the Euclidean ball in the nonnegative orthant. Its symmetry is $1/\sqrt{m}$, and the stochasticity gap is $1+\sqrt{m}$.

(iii) For $p = \infty$, $\mathbf{B}_p^+$ is a hypercube centered at $\boldsymbol{e}/2$ and touches the origin. The symmetry is 1 and the stochasticity gap is 2.

Ellipsoidal uncertainty set. An ellipsoidal uncertainty set is defined as,

$$\mathcal{U} := \{\boldsymbol{b} \in \mathbb{R}_+^m \mid \|\boldsymbol{E}(\boldsymbol{b} - \overline{\boldsymbol{b}})\|_2 \leq 1\}, \tag{4.8}$$

where $\boldsymbol{E}$ is an invertible matrix. Since $\mathcal{U}$ is symmetric, the stochasticity gap is bounded by 2 if $\mathcal{U} \subseteq \mathbb{R}_+^m$ and the probability measure satisfies (4.1). In Appendix A.3, we show that the bound can be improved to the following.

$$z_{\text{Rob}} \leq \left(1 + \max_{1\leq i\leq m} \frac{E_{ii}^{-1}}{\overline{b}_i}\right) z_{\text{Stoch}} \leq 2 z_{\text{Stoch}}.$$

Intersection of two L_p-balls. Consider the following uncertainty set.

$$\mathcal{U} := \{\boldsymbol{b} \in \mathbb{R}_+^m \mid \|\boldsymbol{b}\|_{p_1} \leq 1, \|\boldsymbol{b}\|_{p_2} \leq r\}, \quad \text{for } 0 < r < 1 \text{ and } 1 \leq p_1 < p_2. \tag{4.9}$$

Assume the following condition holds,

$$\frac{r}{\|\boldsymbol{e}\|_{p_2}} > \frac{1}{\|\boldsymbol{e}\|_{p_1}} \Leftrightarrow r m^{\frac{1}{p_1}} > m^{\frac{1}{p_2}}, \tag{4.10}$$

which guarantees that the intersection of the two unit norm balls is non-trivial. In Appendix A.4, we show that

$$\mathbf{sym}(\mathcal{U}) = \frac{1}{r m^{\frac{1}{p_1}}}, \quad \boldsymbol{b}_0(\mathcal{U}) = \frac{r}{r m^{\frac{1}{p_1}}+1}\boldsymbol{e}. \tag{4.11}$$

The budgeted uncertainty set. The budgeted uncertainty set is defined as,

$$\Delta_k := \left\{ \boldsymbol{b} \in [0,1]^m \;\middle|\; \sum_{i=1}^m b_i \leq k \right\}, \quad \text{for } 1 \leq k \leq m. \tag{4.12}$$

In Appendix A.5, we show that

$$\mathbf{sym}(\Delta_k) = \frac{k}{m}, \quad \boldsymbol{b}_0(\Delta_k) = \frac{k}{m+k}\boldsymbol{e}. \tag{4.13}$$

Demand uncertainty set. We define the following set,

$$\mathbf{DU} := \left\{ \boldsymbol{b} \in \mathbb{R}^m_+ \;\middle|\; \left| \frac{\sum_{i \in S} \boldsymbol{b}_i - |S|\mu}{\sqrt{|S|}} \right| \leq \Gamma, \; \forall S \subseteq N := \{1, \ldots, m\} \right\}. \tag{4.14}$$

Such a set can model the demand uncertainty where $\boldsymbol{b}$ is the demand of m products; μ and Γ are the center and the span of the uncertain range.

The set **DU** has different symmetry properties, depending on the relation between μ and Γ. If $\mu \geq \Gamma$, the set **DU** is in fact symmetric. Intuitively, **DU** is the intersection of an L_∞ ball centered at $\mu\boldsymbol{e}$ with $(2^m - 2m)$ halfspaces that are symmetric with respect to $\mu\boldsymbol{e}$. If $\mu < \Gamma$, **DU** is not symmetric any more — part of it is cut off by the nonnegative orthant. In Appendix A.6, we present a proof of the following proposition, which summarizes the symmetry property of **DU** for all the cases.

Proposition 4.1 *Assume the uncertainty set is* **DU**,

(*i*) *If* $\mu \geq \Gamma$, *then,*

$$\mathbf{sym}(\mathbf{DU}) = 1, \quad \boldsymbol{b}_0(\mathbf{DU}) = \mu\boldsymbol{e}. \tag{4.15}$$

(*ii*) *If* $\frac{1}{\sqrt{m}}\Gamma < \mu < \Gamma$, *then,*

$$\mathbf{sym}(\mathbf{DU}) = \frac{\sqrt{m}\mu + \Gamma}{(1+\sqrt{m})\Gamma}, \quad \boldsymbol{b}_0(\mathbf{DU}) = \frac{(\sqrt{m}\mu + \Gamma)(\mu + \Gamma)}{\sqrt{m}\mu + (2+\sqrt{m})\Gamma}\boldsymbol{e}. \tag{4.16}$$

(*iii*) *If* $0 \leq \mu \leq \frac{1}{\sqrt{m}}\Gamma$, *then,*

$$\mathbf{sym}(\mathbf{DU}) = \frac{\sqrt{m}\mu + \Gamma}{\sqrt{m}(\mu + \Gamma)}, \quad \boldsymbol{b}_0(\mathbf{DU}) = \frac{(\sqrt{m}\mu + \Gamma)(\mu + \Gamma)}{2\sqrt{m}\mu + (1+\sqrt{m})\Gamma}\boldsymbol{e}. \tag{4.17}$$

5. Multi-stage stochastic problem under RHS uncertainty.

In this section, we consider the multi-stage stochastic optimization problem, Π^K_{Stoch}, under right hand side uncertainty where the multi-stage uncertainty is described by a directed network as discussed earlier. We show that a finitely adaptable class of solutions is a good approximation for the fully-adaptable multi-stage problem. Furthermore, the performance ratio of the finitely adaptable solution depends on the geometric properties of the uncertainty sets in the multi-stage uncertainty network. The number of solutions at each Stage $(k+1)$ for $k=1,\ldots,K-1$ depends on the number of directed paths in the uncertainty network from the root node to nodes in Stage $(k+1)$. Therefore, if $\mathcal{P}$ is the set of all directed paths from Stage 2 to Stage K, the total number of solutions in any stage in the finitely adaptable solution policy is bounded by $|\mathcal{P}|$.

THEOREM 5.1 *Let $s=\min_{k,j}\mathbf{sym}(\boldsymbol{u}_{k,j},\mathcal{U}_j^k)$ and $\rho=\max_{k,j}\rho(\boldsymbol{u}_{k,j},\mathcal{U}_j^k)$. Suppose $\mathbb{E}_{\boldsymbol{b}}[\boldsymbol{b}\mid\boldsymbol{b}\in\mathcal{U}_j^k]\geq\boldsymbol{u}_{k,j}$ where $\boldsymbol{u}_{k,j}$ is the point of symmetry of $\mathcal{U}_j^k\subseteq\mathbb{R}_+^m$ for all $j=1,\ldots,N_k$, $k=1,\ldots,K-1$. Then there is a finitely adaptable solution policy that can be computed efficiently and has at most $|\mathcal{P}|$ solutions in each stage, where $\mathcal{P}$ is the set of directed paths from the root node in Stage 2 to nodes in Stage K in the multi-stage uncertainty network, such that the expected cost is at most $(1+\rho/s)$ times the optimal cost of Π^K_{Stoch}.*

5.1 Algorithm. We first describe an algorithm to construct a finitely adaptable solution. In each Stage $(k+1)$ for $k=1,\ldots,K-1$, the set of finitely adaptable solutions contains a unique solution corresponding to each directed path from the root node in Stage 2 to a node in Stage $(k+1)$. Therefore, we consider $|\mathcal{P}[k]|$ solutions in Stage $(k+1)$ each indexed by $\boldsymbol{P}[k]$ for all $\boldsymbol{P}\in\mathcal{P}$. In other words, the finitely adaptable solution is specified by the first-stage solution $\boldsymbol{x}$, and for each Stage $(k+1)$ for $k=1,\ldots,K-1$, $\boldsymbol{y}_k(\boldsymbol{P}[k])$ for all $\boldsymbol{P}[k]\in\mathcal{P}[k]$. Recall that for any $\boldsymbol{P}=(j_1,\ldots,j_{K-1})\in\mathcal{P}$, the probability that the uncertain parameters realize from the path $\boldsymbol{P}$ is given by,

$$Pr(j_1,\ldots,j_{K-1})=\prod_{k=1}^{K-2}p^k_{j_k,j_{k+1}}, \tag{5.1}$$

and the measure $\mu_{\boldsymbol{P}}$ is defined as a product measure of the measures on the uncertainty sets in Path $\boldsymbol{P}$. We first show that for any $k=1,\ldots,K-1$, $j=1,\ldots,N_k$, $(1+\rho/s)\cdot\boldsymbol{u}_{k,j}$ dominates all points in $\mathcal{U}_j^k$ coordinatewise.

LEMMA 5.1 *For any $k=1,\ldots,K$, $j=1,\ldots,N_k$, for all $\boldsymbol{u}\in\mathcal{U}_j^k$,*

$$\left(1+\frac{\rho}{s}\right)\cdot\boldsymbol{u}_{k,j}\geq\boldsymbol{u}.$$

PROOF. Let $\rho_1=\rho(\boldsymbol{u}_{k,j},\mathcal{U}_j^k)$, $s_1=\mathbf{sym}(\boldsymbol{u}_{k,j},\mathcal{U}_j^k)$. From Lemma 2.4, we know that

$$\left(1+\frac{\rho_1}{s_1}\right)\cdot\boldsymbol{u}_{k,j}\geq\boldsymbol{u},\ \forall\boldsymbol{u}\in\mathcal{U}_j^k.$$

By definition in Theorem 5.1, we know that $\rho_1\leq\rho$ and $s_1\geq s$. Therefore, for all $\boldsymbol{u}\in\mathcal{U}_j^k$,

$$\left(1+\frac{\rho}{s}\right)\cdot\boldsymbol{u}_{k,j}\geq\left(1+\frac{\rho_1}{s_1}\right)\cdot\boldsymbol{u}_{k,j}\geq\boldsymbol{u}.$$

□

We consider the following multi-stage problem to compute the finitely adaptable solution.

$$
\begin{aligned}
z_{\mathcal{A}} = \min \quad & \boldsymbol{c}^T \boldsymbol{x} + \sum_{k=1}^{K-1} \mathbb{E}_{\boldsymbol{P}\in\mathcal{P}}[\boldsymbol{d}_k^T \boldsymbol{y}_k(\boldsymbol{P}[k])] \\
\text{s.t. } & \forall \boldsymbol{P} = (j_1, \ldots, j_{K-1}) \in \mathcal{P} \\
& \boldsymbol{A}\boldsymbol{x} + \sum_{k=1}^{K-1} \boldsymbol{B}_k \boldsymbol{y}_k(\boldsymbol{P}[k]) \geq \left(1 + \frac{\rho}{s}\right) \cdot \sum_{k=1}^{K-1} \boldsymbol{u}_{k,j_k} \\
& \boldsymbol{x} \in \mathbb{R}^{p_1} \times \mathbb{R}_+^{n_1 - p_1}, \\
& \boldsymbol{y}_k(\boldsymbol{P}[k]) \in \mathbb{R}^{p_k} \times \mathbb{R}_+^{n_k - p_k}, \ \forall k = 1, \ldots, K-1.
\end{aligned}
\tag{5.2}
$$

The number of constraints in (5.2) is equal to $|\mathcal{P}| \cdot m$, where m is the number of rows of $\boldsymbol{A}, \boldsymbol{B}_k$ for all $k = 1, \ldots, K-1$. Also, the number of decision variables in each Stage $(k+1)$ for $k = 1, \ldots, K-1$ is $|\mathcal{P}[k]|$, namely, $\boldsymbol{y}_k(\boldsymbol{P}[k])$ for all $\boldsymbol{P}[k] \in \mathcal{P}[k]$. Therefore, the solution is finitely adaptable as there are only a finite number of solutions in each stage. Furthermore, in some cases, the number of directed paths is small and polynomial in the size of the uncertainty network. For instance, if the uncertainty network is a path, we require exactly one solution at each stage in our finitely adaptable solution. If the network is a tree, then there is a single path to each uncertainty set from the root node in Stage 2 and therefore, the number of solutions in the finitely adaptable solution is exactly equal to the number of uncertainty sets in the uncertainty network. However, in general, the number of directed paths can be exponential in the input size. For example, for the case of a recombining directed network in Figure 1, the number of directed paths from the root node to the node j in Stage $(k+1)$, $j = 1, \ldots, N_k$ is equal to $\binom{k}{j}$, which is exponential in k, while the input size is $\mathrm{O}(k^2)$.

In the next subsection, we show that the finitely adaptable solution computed in (5.2) is a good approximation of a fully-adaptable optimal stochastic solution.

5.2 Proof of Theorem 5.1.

For brevity, let $\tau = (1 + \rho/s)$. The proof proceeds as follows. We first consider a particular finitely adaptable solution feasible to (5.2), which implies that its expected cost is at least $z_{\mathcal{A}}$. We then extend this particular finitely adaptable solution to a fully-adaptable one without changing its expected cost. Finally, we show that the expected cost of the extended solution is equal to τ times the expected cost of an optimal fully-adaptable solution.

Let $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}_k(\boldsymbol{\omega}[k], \boldsymbol{P}[k])$ denote an optimal fully-adaptable solution for Π^K_{Stoch} for all $k = 1, \ldots, K-1$, $\boldsymbol{P} \in \mathcal{P}$, $\mu_{\boldsymbol{P}}$-a.e. $\boldsymbol{\omega} \in \Omega(\boldsymbol{P})$. For the first step of the proof, we consider the following particular finitely adaptable solution for (5.2). For all $\boldsymbol{P} \in \mathcal{P}$, $k = 1, \ldots, K-1$, let

$$\overline{\boldsymbol{y}}_k(\boldsymbol{P}[k]) = \tau \cdot \mathbb{E}_{\boldsymbol{\omega} \in \Omega(\boldsymbol{P})} \left[\hat{\boldsymbol{y}}_k(\boldsymbol{\omega}[k], \boldsymbol{P}[k]) \right]. \tag{5.3}$$

Also, let $\overline{\boldsymbol{x}} = \tau \cdot \hat{\boldsymbol{x}}$. We show that the above finitely adaptable solution is feasible for (5.2). Consider any $\boldsymbol{P} = (j_1, \ldots, j_{K-1}) \in \mathcal{P}$. Now,

$$\boldsymbol{A}\hat{\boldsymbol{x}} + \sum_{k=1}^{K-1} \boldsymbol{B}_k \hat{\boldsymbol{y}}_k(\boldsymbol{\omega}[k], \boldsymbol{P}[k]) \geq \sum_{k=1}^{K-1} \boldsymbol{b}_k, \mu_{\boldsymbol{P}}\text{-a.e. } \boldsymbol{\omega} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{K-1}) \in \Omega(\boldsymbol{P}).$$

Taking conditional expectation with respect to $\mu_{\boldsymbol{P}}$ on both sides, we have that

$$\boldsymbol{A}\hat{\boldsymbol{x}} + \sum_{k=1}^{K-1} \boldsymbol{B}_k \mathbb{E}_{\mu_{\boldsymbol{P}}} \left[\hat{\boldsymbol{y}}_k\left(\boldsymbol{\omega}[k], \boldsymbol{P}[k]\right) \right] \geq \sum_{k=1}^{K-1} \mathbb{E}_{\mu_{\boldsymbol{P}}} \left[\boldsymbol{b}_k \right] \geq \sum_{k=1}^{K} \boldsymbol{u}_{k,j_k}, \tag{5.4}$$

where the last inequality follows as $\mathbb{E}_{\mu_{\boldsymbol{P}}}[\boldsymbol{b}_k] \geq \boldsymbol{u}_{k,j_k}$. Therefore, for any $\boldsymbol{P} = (j_1, \ldots, j_{K-1}) \in \mathcal{P}$,

$$\begin{aligned}
\boldsymbol{A}\overline{\boldsymbol{x}} + \sum_{k=1}^{K-1} \boldsymbol{B}_k \overline{\boldsymbol{y}}_k(\boldsymbol{P}[k]) &= \boldsymbol{A}(\tau \cdot \hat{\boldsymbol{x}}) + \sum_{k=1}^{K-1} \boldsymbol{B}_k \left(\tau \cdot \mathbb{E}_{\mu_{\boldsymbol{P}}} \left[\boldsymbol{d}_k^T \hat{\boldsymbol{y}}_k\left(\boldsymbol{\omega}[k], \boldsymbol{P}[k]\right) \right] \right) \\
&= \tau \cdot \left(\boldsymbol{A}\hat{\boldsymbol{x}} + \sum_{k=1}^{K-1} \boldsymbol{B}_k \mathbb{E}_{\mu_{\boldsymbol{P}}} \left[\boldsymbol{d}_k^T \hat{\boldsymbol{y}}_k(\boldsymbol{\omega}[k], \boldsymbol{P}[k]) \right] \right) \\
&\geq \tau \cdot \left(\sum_{k=1}^{K-1} \boldsymbol{u}_{k,j_k} \right),
\end{aligned} \tag{5.5}$$

where (5.5) follows from (5.4). Therefore, the solution $\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}}$ is a feasible solution for (5.2). Let $\overline{z}$ be the expected cost of $\overline{\boldsymbol{x}}$, $\overline{\boldsymbol{y}}$, i.e.,

$$\overline{z} = \boldsymbol{c}^T \overline{\boldsymbol{x}} + \sum_{k=1}^{K-1} \mathbb{E}_{\boldsymbol{P} \in \mathcal{P}} [\boldsymbol{d}_k^T \overline{\boldsymbol{y}}_k(\boldsymbol{P}[k])].$$

Clearly, $z_{\mathcal{A}} \leq \overline{z}$. For the second step of the proof, we extend the finitely adaptable solution $\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}}$ to a feasible solution $\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}$ for Π^K_{Stoch} as follows.

$$\begin{aligned}
\tilde{\boldsymbol{x}} &= \overline{\boldsymbol{x}} \\
\tilde{\boldsymbol{y}}_k(\boldsymbol{\omega}[k], \boldsymbol{P}[k]) &= \overline{\boldsymbol{y}}_k(\boldsymbol{P}[k]), \ \forall \boldsymbol{P} \in \mathcal{P}, \ \forall k = 1, \ldots, K-1.
\end{aligned} \tag{5.6}$$

We show that the extended solution $\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}$ is a feasible solution for Π^K_{Stoch}. Consider any $\boldsymbol{P} = (j_1, \ldots, j_{K-1}) \in \mathcal{P}$, $\boldsymbol{\omega} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{K-1}) \in \Omega(\boldsymbol{P})$. Therefore,

$$\begin{aligned}
\boldsymbol{A}\tilde{\boldsymbol{x}} + \sum_{k=1}^{K-1} \boldsymbol{B}_k \tilde{\boldsymbol{y}}_k(\boldsymbol{\omega}[k], \boldsymbol{P}[k]) &= \boldsymbol{A}\overline{\boldsymbol{x}} + \sum_{k=1}^{K-1} \boldsymbol{B}_k \overline{\boldsymbol{y}}_k(\boldsymbol{P}[k]) \\
&\geq \tau \cdot \sum_{k=1}^{K-1} \boldsymbol{u}_{k,j_k} \qquad (5.7) \\
&\geq \sum_{k=1}^{K-1} \boldsymbol{b}_k,
\end{aligned}$$

where (5.7) follows from the feasibility of the solution $\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}}$ for (5.2) and the last inequality follows from

Lemma 5.1. Let $\tilde{z}$ be the expected cost of $\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}$, i.e.,

$$\begin{aligned}
\tilde{z} &= \boldsymbol{c}^T \tilde{\boldsymbol{x}} + \sum_{k=1}^{K-1} \mathbb{E}_{\boldsymbol{P}\in\mathcal{P}} \left[\mathbb{E}_{\mu_{\boldsymbol{P}}} [\boldsymbol{d}_k^T \tilde{\boldsymbol{y}}_k(\boldsymbol{\omega}[k], \boldsymbol{P}[k])] \right] \\
&= \boldsymbol{c}^T \overline{\boldsymbol{x}} + \sum_{k=1}^{K-1} \mathbb{E}_{\boldsymbol{P}\in\mathcal{P}} \left[\mathbb{E}_{\mu_{\boldsymbol{P}}} [\boldsymbol{d}_k^T \overline{\boldsymbol{y}}_k(\boldsymbol{P}[k])] \right] \\
&= \boldsymbol{c}^T \overline{\boldsymbol{x}} + \sum_{k=1}^{K-1} \mathbb{E}_{\boldsymbol{P}\in\mathcal{P}} [\boldsymbol{d}_k^T \overline{\boldsymbol{y}}_k(\boldsymbol{P}[k])] \qquad (5.8) \\
&= \overline{z}
\end{aligned}$$

where (5.8) follows from (5.6). Now, we compare $\tilde{z}$ to the expected cost of an optimal fully-adaptable solution as follows.

$$\begin{aligned}
\tilde{z} &= \overline{z} \\
&= \boldsymbol{c}^T \overline{\boldsymbol{x}} + \sum_{k=1}^{K-1} \mathbb{E}_{\boldsymbol{P}\in\mathcal{P}} [\boldsymbol{d}_k^T \overline{\boldsymbol{y}}_k(\boldsymbol{P}[k])] \\
&= \boldsymbol{c}^T (\tau \cdot \boldsymbol{x}) + \sum_{k=1}^{K-1} \mathbb{E}_{\boldsymbol{P}\in\mathcal{P}} \left[\boldsymbol{d}_k^T \left(\tau \cdot \mathbb{E}_{\mu_{\boldsymbol{P}}} \left[\hat{\boldsymbol{y}}_k(\boldsymbol{\omega}[k], \boldsymbol{P}[k]) \right] \right) \right] \\
&= \tau \cdot z_{\text{Stoch}}^K, \qquad (5.9)
\end{aligned}$$

where (5.9) follows from the optimality of $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}$ for Π_{Stoch}^K. Combining $z_{\mathcal{A}} \leq \overline{z}$ and (5.9), we obtain that $z_{\mathcal{A}} \leq \tau \cdot z_{\text{Stoch}}^K$. □

Stochastic problem under cost and RHS uncertainty. For the multi-stage stochastic problem, $\Pi_{\text{Stoch}(\boldsymbol{b},\boldsymbol{d})}^K$ where both the objective coefficients and the right hand side are uncertain, a finitely adaptable solution performs arbitrarily worse as compared to the fully-adaptable solution. It follows from one of the results in Bertsimas and Goyal [11], where the authors show that a static-robust solution may perform arbitrarily worse as compared to an optimal fully-adaptable two-stage solution for the stochastic problem when both cost and right hand side are uncertain.

6. Multi-stage adaptive problem.

In this section, we consider multi-stage adaptive optimization problems and show that a finitely adaptable solution is a good approximation to the fully-adaptable solution. Furthermore, the performance bound of a finitely adaptable solution is related to the symmetry and translation factors of the uncertainty sets as in the case of stochastic optimization problems. The approximation guarantee for the case when only the right hand sides of the multi-stage adaptive problem are uncertain, Π_{Adapt}^K, follows directly from the approximation bounds for the stochastic problem. Surprisingly, we can also show that a finitely adaptable solution is also a good approximation for the multi-stage adaptive problem when both the right hand sides and the objective coefficients are uncertain,

$\Pi^K_{\text{Adapt}(\boldsymbol{b},\boldsymbol{d})}$, unlike the stochastic counterpart. Our results generalize the performance of a static robust solution for two-stage adaptive optimization problem under right hand side and/or objective coefficient uncertainty.

6.1 Right hand side uncertainty, $\Pi^K_{\textbf{Adapt}}$. We show that there is a finitely adaptable solution with at most $|\mathcal{P}|$ solutions for each stage and each uncertainty set, that is a good approximation of the fully-adaptable problem. Furthermore, such a finitely adaptable solution can be computed efficiently using exactly the algorithm described in Section 5.1. The performance bound of the finitely adaptable solution policy follows directly from its performance bound with respect to the stochastic problem in Theorem 5.1. Therefore, we have the following theorem.

THEOREM 6.1 *Suppose $s = \min_{k,j} \textbf{sym}(\mathcal{U}^k_j)$ and $\boldsymbol{u}_{k,j} \in \mathcal{U}^k_j \subseteq \mathbb{R}^m_+$ is the point of symmetry for all $j = 1,\ldots,N_k$, $k = 1,\ldots,K-1$. Let $\rho = \max_{k,j} \rho(\mathcal{U}^k_j)$ for all $j = 1,\ldots,N_k$, $k = 1,\ldots,K-1$. Then there is a finitely adaptable solution policy that can be computed efficiently and has at most $|\mathcal{P}|$ solutions in each stage, where $\mathcal{P}$ is the set of directed paths from the root node to any node in Stage K of the multi-stage uncertainty network, such that its worst-case cost is at most $(1+\rho/s)$ times the optimal cost of Π^K_{Adapt}.*

6.2 RHS and cost uncertainty. In this section, we consider the multi-stage adaptive optimization problem where both the right hand side and the objective coefficients are uncertain. While for the stochastic problem, the performance of a finitely adaptable solution can be arbitrarily bad with respect to an optimal stochastic solution, surprisingly, we can show that there exists a finitely adaptable solution with at most $|\mathcal{P}|$ solutions for each stage for each uncertainty set, that is a good approximation for the multi-stage problem.

THEOREM 6.2 *Suppose $s = \min_{k,j} \textbf{sym}(\mathcal{U}^k_j)$ and $\boldsymbol{u}_{k,j} \in \mathcal{U}^k_j \subseteq \mathbb{R}^{m+n_k}_+$ is the point of symmetry for all $j = 1,\ldots,N_k$, $k = 1,\ldots,K-1$. Also, let $\boldsymbol{u}^{\boldsymbol{b}}_{k,j}, \boldsymbol{u}^{\boldsymbol{d}}_{k,j}$ denote the right hand side and the objective coefficient uncertainty in $\boldsymbol{u}_{k,j}$ respectively. Let $\rho = \max_{k,j} \rho(\mathcal{U}^k_j)$ for all $j = 1,\ldots,N_k$, $k = 1,\ldots,K-1$. Then there is a finitely adaptable solution policy that can be computed efficiently and has at most $|\mathcal{P}|$ solutions where $\mathcal{P}$ is the set of directed paths from the root node to any node in Stage K of the multi-stage uncertainty network, such that its worst-case cost is at most $(1+\rho/s)^2$ times the optimal cost of $\Pi^K_{\text{Adapt}(\boldsymbol{b},\boldsymbol{d})}$.*

PROOF. From Lemma 2.4, we know that for any uncertainty set $\mathcal{U}^k_j$ in Stage $(k+1)$, $\boldsymbol{u} \le (1+\rho/s)\cdot\boldsymbol{u}_{k,j}$ for all $\boldsymbol{u} \in \mathcal{U}^k_j$. We compute a finitely adaptable solution by solving the following multi-stage problem similar to the one used to compute a finitely adaptable solution for the stochastic counterpart.

$$\begin{aligned} \min \quad & \boldsymbol{c}^T\boldsymbol{x} + \max_{\boldsymbol{P}\in\mathcal{P}} \min_{\boldsymbol{y}_k(\boldsymbol{P}[k]), k=1,\ldots,K-1} \sum_{k=1}^{K-1} \boldsymbol{d}_k^T \boldsymbol{y}_k(\boldsymbol{P}[k]) \\ \text{s.t. } & \forall \boldsymbol{P} = (j_1,\ldots,j_{K-1}) \in \mathcal{P} \\ & \boldsymbol{A}\boldsymbol{x} + \sum_{k=1}^{K-1} \boldsymbol{B}_k \boldsymbol{y}_k(\boldsymbol{P}[k]) \geq \left(1+\frac{\rho}{s}\right) \cdot \sum_{k=1}^{K-1} \boldsymbol{u}^{\boldsymbol{b}}_{k,j_k}, \\ & \boldsymbol{x} \in \mathbb{R}^{p_1} \times \mathbb{R}^{n_1-p_1}_+, \\ & \boldsymbol{y}_k(\boldsymbol{\omega}[k], \boldsymbol{P}[k]) \in \mathbb{R}^{p_k} \times \mathbb{R}^{n_k-p_k}_+, \ \forall k = 1,\ldots,K-1. \end{aligned} \tag{6.1}$$

Note that the above problem is similar to (5.2) except the objective function. Using an argument similar to the second part of the proof of Theorem 5.1 (see Section 5.2), we can show that a finitely adaptable solution of (6.1) can be extended to a feasible fully-adaptable solution for $\Pi^K_{\text{Adapt}(\boldsymbol{b},\boldsymbol{d})}$ of the same worst-case cost.

Suppose $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}_k(\boldsymbol{\omega}[k], \boldsymbol{P}[k]))$ for all $k = 1,\ldots,K-1$, $\boldsymbol{P} \in \mathcal{P}, \boldsymbol{\omega} \in \Omega(\boldsymbol{P})$ denote a optimal fully-adaptable solution for $\Pi^K_{\text{Adapt}}(\boldsymbol{b},\boldsymbol{d})$. As defined earlier, for any $\boldsymbol{P} = (j_1,\ldots,j_{K-1}) \in \mathcal{P}$, let

$$\boldsymbol{\omega}_{\boldsymbol{P}} = (\boldsymbol{u}_{1,j_1}, \ldots, \boldsymbol{u}_{K-1,j_{K-1}}), \tag{6.2}$$

Consider the following approximate solution for (6.1).

$$\tilde{\boldsymbol{x}} = \left(1+\frac{\rho}{s}\right) \cdot \hat{\boldsymbol{x}}, \tag{6.3}$$

and for any $\boldsymbol{P} \in \mathcal{P}$, for all $k = 1,\ldots,K-1$,

$$\tilde{\boldsymbol{y}}_k(\boldsymbol{P}[k]) = \left(1+\frac{\rho}{s}\right) \cdot \hat{\boldsymbol{y}}_k(\boldsymbol{\omega}_{\boldsymbol{P}}[k], \boldsymbol{P}[k]). \tag{6.4}$$

For brevity, as before, let $\tau = (1+\rho/s)$. We first need to show that the solution $\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}$ is feasible for (6.1). For any $\boldsymbol{P} = (j_1,\ldots,j_{K-1}) \in \mathcal{P}$,

$$\boldsymbol{A}\tilde{\boldsymbol{x}} + \sum_{k=1}^{K-1} \boldsymbol{B}_k \tilde{\boldsymbol{y}}_k(\boldsymbol{P}[k]) = \tau \cdot \left(\boldsymbol{A}\hat{\boldsymbol{x}} + \sum_{k=1}^{K-1} \boldsymbol{B}_k \hat{\boldsymbol{y}}_k(\boldsymbol{\omega}_{\boldsymbol{P}}[k], \boldsymbol{P}[k]) \right) \geq \tau \cdot \left(\sum_{k=1}^{K-1} \boldsymbol{u}_{k,j_k} \right),$$

where the last inequality follows from the feasibility for $\hat{\boldsymbol{y}}$ for $\boldsymbol{\omega}_{\boldsymbol{P}}$. Thus, the solution $\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}$ is feasible for (6.1).

To bound the worst-case cost of the solution, we show that for any $\boldsymbol{P} \in \mathcal{P}$, the cost of the approximate finitely adaptable solution is at most $(1+\rho/s)^2$ times the worst-case cost of the optimal fully-adaptable solution. Consider any $\boldsymbol{P} = (j_1, \ldots, j_{K-1}) \in \mathcal{P}$. Now,

$$\begin{aligned}
\boldsymbol{c}^T\tilde{\boldsymbol{x}} + \sum_{k=1}^{K-1} \boldsymbol{d}_k^T \tilde{\boldsymbol{y}}_k(\boldsymbol{P}[k]) &= \tau \cdot \left(\boldsymbol{c}^T\hat{\boldsymbol{x}} + \sum_{k=1}^{K-1} \boldsymbol{d}_k^T \hat{\boldsymbol{y}}_k(\boldsymbol{\omega_P}[k], \boldsymbol{P}[k])\right) \\
&\leq \tau \cdot \left(\boldsymbol{c}^T\hat{\boldsymbol{x}} + \sum_{k=1}^{K-1} \left(\tau \cdot (\boldsymbol{u}^{\boldsymbol{d}}_{k,j_k})^T \hat{\boldsymbol{y}}_k(\boldsymbol{\omega_P}[k], \boldsymbol{P}[k])\right)\right) \qquad (6.5) \\
&= \tau \cdot \boldsymbol{c}^T\hat{\boldsymbol{x}} + \tau^2 \cdot \left(\sum_{k=1}^{K-1} (\boldsymbol{u}^{\boldsymbol{d}}_{k,j_k})^T \hat{\boldsymbol{y}}_k(\boldsymbol{\omega_P}[k], \boldsymbol{P}[k])\right) \\
&\leq \tau^2 \cdot z^K_{\text{Adapt}(\boldsymbol{b},\boldsymbol{d})}, \qquad (6.6)
\end{aligned}$$

where (6.5) follows as $\boldsymbol{d} \leq \tau \cdot \boldsymbol{u}^{\boldsymbol{d}}_{k,j_k}$ for all $(\boldsymbol{b}, \boldsymbol{d}) \in \mathcal{U}^k_{j_k}$, $k = 1, \ldots, K-1$. Inequality (6.6) follows as $\boldsymbol{\omega_P} \in \Omega(\boldsymbol{P})$ and thus, is a feasible scenario in $\Pi^K_{\text{Adapt}(\boldsymbol{b},\boldsymbol{d})}$. □

6.3 An alternative bound. In this section, we discuss an alternative bound for the performance of the finitely adaptable solution as compared to the optimal fully-adaptable solution similar to the one we present for the stochastic problem. For the sake of simplicity, we present this alternative bound for the adaptive problem under right hand side uncertainty, Π^K_{Adapt}. The bound extends in a straightforward manner to the case of both right hand side and cost uncertainty.

The proof of Theorem 6.1 (and also Theorem 6.2) is based on the construction of a good finitely adaptable solution from an optimal fully-adaptable solution in the following manner. For each $\boldsymbol{P} \in \mathcal{P}$, we consider the scenario $\boldsymbol{\omega_P}$ where in each stage the uncertain parameter realization is the point of symmetry of the corresponding uncertainty set on path $\boldsymbol{P}$. We show that the solution for scenario $\boldsymbol{\omega_P}$ scaled by a factor $(1+\rho/s)$ is a good feasible finitely adaptable solution for all $\boldsymbol{\omega} \in \Omega(\boldsymbol{P})$. Since $\boldsymbol{\omega_P} \in \Omega(\boldsymbol{P})$, the cost for an optimal solution for this scenario is a lower bound on z^K_{Adapt} which implies a bound of $(1+\rho/s)$ for the finitely adaptable solution with respect to the optimal. Now, if for some other scenario $\boldsymbol{\omega}'(\boldsymbol{P}) \in \Omega(\boldsymbol{P})$, a smaller scaling factor than $(1+\rho/s)$ suffices to obtain a feasible finitely adaptable solution for all $\boldsymbol{\omega} \in \Omega(\boldsymbol{P})$ for all $\boldsymbol{P} \in \mathcal{P}$, this would imply a smaller bound on the performance of the finitely adaptable solution.

Following the discussion in Section 4.2, we define $\boldsymbol{b}^h(\mathcal{U})$ as follows. For all $j = 1, \ldots, m$, $b^h_j(\mathcal{U}) := \max_{\boldsymbol{b}\in\mathcal{U}} b_j$. Let

$$\theta(\mathcal{U}) = \min\{\theta \mid \exists \boldsymbol{b} \in \mathcal{U}, \theta \cdot \boldsymbol{b} \geq \boldsymbol{b}^h(\mathcal{U})\}.$$

Note that $\theta(\mathcal{U}) \leq (1+\rho/s)$ as $(1+\rho/s)\cdot \boldsymbol{b}_0 \geq \boldsymbol{b}^h(\mathcal{U})$, where $\boldsymbol{b}_0$ is the point of symmetry of $\mathcal{U}$, $\rho = \rho(\boldsymbol{b}_0, \mathcal{U})$, and $s = \mathbf{sym}(\boldsymbol{b}_0, \mathcal{U})$. Let $\boldsymbol{b}^1(\mathcal{U})$ denote the vector $\boldsymbol{b} \in \mathcal{U}$ that achieves the minimum value of θ. Let

$$\theta^*_a = \max_{k=1,\ldots,K, j=1,\ldots,N_k} \theta(\mathcal{U}^k_j).$$

For each $k = 1, \ldots, K-1$, $j = 1, \ldots, N_k$, $\theta(\mathcal{U}_j^k)$ is defined by a feasible uncertainty realization from $\mathcal{U}_j^k$. Therefore, scaling the solution corresponding to $\boldsymbol{b}^1(\mathcal{U}_j^k)$ by a factor $\theta_a^* \geq \theta(\mathcal{U}_j^k)$ produces a feasible finitely adaptable solution. Also, as we note above, $\theta_a^* \leq (1+\rho/s)$. Therefore, θ_a^* is upper bounded by the bound in Theorem 6.1. We refer to θ_a^* as a scaling bound as earlier.

We can interpret the scaling bound geometrically as follows. For any $\mathcal{U} \subseteq \mathbb{R}_+^m$, let $\theta = \theta(\mathcal{U})$, $\boldsymbol{b}^1 = \boldsymbol{b}^1(\mathcal{U})$, and $\boldsymbol{b}^h = \boldsymbol{b}^h(\mathcal{U})$. We know that

$$\theta \cdot \boldsymbol{b}^1 \geq \boldsymbol{b}^h \Rightarrow \boldsymbol{b}^1 \geq \frac{1}{\theta} \cdot \boldsymbol{b}^h.$$

Therefore, $1/\theta$ is the minimum scaling factor for $\boldsymbol{b}^h$ such that it is dominated by some point in $\mathcal{U}$ coordinate-wise. Note that $(1/\theta) \cdot \boldsymbol{b}^h$ does not necessarily belong to $\mathcal{U}$ but is always contained in,

$$\tilde{\mathcal{U}} = \{\boldsymbol{b} \in \mathbb{R}_+^m \mid \exists \boldsymbol{b}' \in \mathcal{U}, \boldsymbol{b}' \geq \boldsymbol{b}\}. \tag{6.7}$$

To see this, consider the following uncertainty set, $\mathcal{U} \subset \mathbb{R}_+^3$, where $\mathcal{U} = \mathsf{conv}\big((0,0,0),(1,0,1),(0,1,1)\big)$. We can alternatively define $\mathcal{U}$ as $\mathcal{U} = \{\boldsymbol{b} \in [0,1]^3 \mid b_3 = b_1 + b_2\}$. It is easy to observe that $\boldsymbol{b}^h = (1,1,1)$. We show that $\boldsymbol{b}^1 = (1/2, 1/2, 1) \in \mathcal{U}$ and $\theta = 2$. We know that $\theta b_j^1 \geq 1$ for all $j = 1, 2$. Therefore,

$$b_3^1 = b_1^1 + b_2^1 \geq \frac{2}{\theta}.$$

Furthermore, $b_3^1 \leq 1$ which implies that $\theta \geq 2$. For $\theta = 2$, $\boldsymbol{b}^1 = (1/2, 1/2, 1) \in \mathcal{U}$ satisfies $\theta \boldsymbol{b}^1 \geq \boldsymbol{b}^h$. Now,

$$\frac{1}{\theta} \cdot \boldsymbol{b}^h = \frac{1}{2} \cdot (1,1,1) \notin \mathcal{U}.$$

However, $\boldsymbol{b}^1 \geq (1/\theta)\boldsymbol{b}^h$, and $\boldsymbol{b}^1 \in \mathcal{U}$ which implies that $(1/\theta)\boldsymbol{b}^h \in \tilde{\mathcal{U}}$ as defined in (6.7). The geometric picture is given in Figure 7. Note that in Figure 7, $\boldsymbol{b}^1 = (1/\theta)\boldsymbol{b}^h$ but this is not true in general as illustrated in the above example.

For certain uncertainty sets, the scaling bound of θ_a^* is strictly better than the bound in Theorem 6.1.

(i) **Hypercube**. Suppose each uncertainty set in the multi-stage uncertainty network is a hypercube. Therefore, $s = 1$. Also, suppose $\rho = 1$ where ρ is as defined in Theorem 6.1. The bound in Theorem 6.1 is $(1 + \rho/s) = 2$. On the other hand, $\theta_a^* = 1$ since $\boldsymbol{b}^h(\mathcal{U}) \in \mathcal{U}$, when $\mathcal{U}$ is a hypercube. Therefore, the two bounds are in fact different. The scaling bound implies that the finitely adaptable solution is optimal, while the symmetry bound implies that it is only a 2-approximation for the multi-stage adaptive optimization problem.

(ii) **Hypersphere**. Suppose the uncertainty sets are all hyperspheres in $\mathbb{R}_+^m$, i.e., L_2-balls, with unit radius and centered at $\boldsymbol{e} = (1, 1, \ldots, 1)$. Therefore, $\rho = 1$ and also $s = 1$ which implies that the

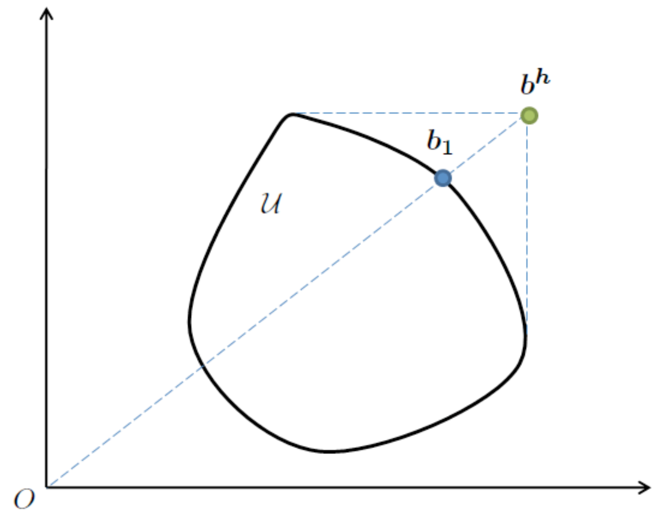

Figure 7: A geometric perspective on the adaptability gap.

bound from Theorem 6.1 is 2. On the other hand, it is easy to note that the scaling bound,

$$\theta_a^* = \frac{2}{1 + 1/\sqrt{m}}.$$

In this case, the scaling bound is not significantly better than the bound in Theorem 6.1.

7. Extension to general cones.

We consider extensions of our results to the case where the constraints are general linear conic inequalities and the uncertainty set belongs to the underlying cone. For simplicity, we discuss the two-stage case. The generalization also applies to the multi-stage problems.

7.1 Stochastic problem with linear conic constraints.

We consider the following two-stage conic stochastic optimization problem.

$$\begin{aligned} z_{\text{Stoch}}^{\mathcal{K}} := \min_{\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{b})} \quad & \boldsymbol{c}^T \boldsymbol{x} + \mathbb{E}_\mu[\boldsymbol{d}^T \boldsymbol{y}(\boldsymbol{b})] \\ \text{s.t.} \quad & \mathcal{A}\boldsymbol{x} + \mathcal{B}\boldsymbol{y}(\boldsymbol{b}) \succeq_{\mathcal{K}} \boldsymbol{b}, \ \mu\text{-a.e. } \boldsymbol{b} \in \mathcal{U}, \\ & \boldsymbol{x} \in \mathbb{R}^{p_1} \times \mathbb{R}_+^{n_1 - p_1}, \\ & \boldsymbol{y}(\boldsymbol{b}) \in \mathbb{R}^{p_2} \times \mathbb{R}_+^{n_2 - p_2}, \end{aligned} \tag{7.1}$$

where $\mathcal{K}$ is a closed pointed convex cone in a finite dimensional space, such as the nonnegative orthant $\mathbb{R}_+^m$, the second-order cone (SOC) $\{(\boldsymbol{b}, t) \in \mathbb{R}^{m+1} : \|\boldsymbol{b}\|_2 \leq t\}$, and the semidefinite cone $\mathbb{S}_+^m$. Here $\mathcal{A}, \mathcal{B}$ are mappings from $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$ to the finite dimensional space that contains $\mathcal{K}$, respectively. For example, if $\mathcal{K}$ is the nonnegative orthant or the second-order cone, both $\mathcal{A}$ and $\mathcal{B}$ are matrices of the appropriate dimension. If $\boldsymbol{\mathcal{K}}$ is the semi-definite (SDP) cone, then $\mathcal{A}, \mathcal{B}$ are linear mappings defined as,

$$\mathcal{A}\boldsymbol{x} = \sum_{i=1}^{n_1} x_i \boldsymbol{A}_i, \ \mathcal{B}\boldsymbol{y} = \sum_{i=1}^{n_2} y_i \boldsymbol{B}_i,$$

where A_i and B_i are symmetric matrices. The linear conic inequality (7.1) is equivalent to the inclusion in the cone, i.e., $\mathcal{A}\boldsymbol{x} + \mathcal{B}\boldsymbol{y}(\boldsymbol{b}) \succeq_{\boldsymbol{\mathcal{K}}} \boldsymbol{b} \ \Leftrightarrow \ \mathcal{A}\boldsymbol{x} + \mathcal{B}\boldsymbol{y}(\boldsymbol{b}) - \boldsymbol{b} \in \boldsymbol{\mathcal{K}}$.

LEMMA 7.1 *Suppose the convex, compact set $\mathcal{U} \subset \boldsymbol{\mathcal{K}}$, and $\boldsymbol{b}_0$ is the point of symmetry of $\mathcal{U}$. Then,*

$$\left(1 + \frac{1}{\mathbf{sym}(\mathcal{U})}\right) \cdot \boldsymbol{b}_0 \succeq_{\boldsymbol{\mathcal{K}}} \boldsymbol{b}, \ \forall \boldsymbol{b} \in \mathcal{U}.$$

PROOF. By the definition of symmetry and the assumption that $\mathcal{U} \subset \boldsymbol{\mathcal{K}}$, we have for any $\boldsymbol{b} \in \mathcal{U}$, $\boldsymbol{b}_0 + \mathbf{sym}(\mathcal{U})(\boldsymbol{b}_0 - \boldsymbol{b}) = (\mathbf{sym}(\mathcal{U}) + 1)\boldsymbol{b}_0 - \mathbf{sym}(\mathcal{U})\boldsymbol{b} \in \boldsymbol{\mathcal{K}}$. □

THEOREM 7.1 *Consider the two-stage stochastic optimization problem in* (7.1). *Let μ be the probability measure on the uncertainty set $\mathcal{U}$, $\boldsymbol{b}_0$ be the point of symmetry of $\mathcal{U}$, and $\rho = \rho(\boldsymbol{b}_0, \mathcal{U})$ be the translation factor of $\boldsymbol{b}_0$ with respect to $\mathcal{U}$. Denote $s = \mathbf{sym}(\mathcal{U})$. Assume the probability measure μ satisfies, $\mathbb{E}_\mu[\boldsymbol{b}] \succeq_{\boldsymbol{\mathcal{K}}} \boldsymbol{b}_0$. Then the cost of an optimal static solution is at most $(1 + \rho/s) \cdot z^{\boldsymbol{\mathcal{K}}}_{\mathrm{Stoch}}$.*

The proof of the above theorem is similar to the proof of Theorem 4.1. For the sake of completeness, we present the proof of Theorem 7.1 in Appendix B. A similar result holds for the corresponding adaptive optimization problem as well.

8. Extensions to integer variables.

We can extend our results to the case when some decision variables are integer constrained for both the stochastic as well as the adaptive optimization problems. In the case of the multi-stage stochastic optimization problem with right hand side uncertainty, we can handle integer decision variables only in the first stage. Whereas for the multi-stage adaptive problem, we can handle integer decision variables in every stage for both versions, Π^K_{Adapt} and $\Pi^K_{\mathrm{Adapt}(\boldsymbol{b},\boldsymbol{d})}$.

8.1 Multi-stage stochastic problem.

We consider the multi-stage stochastic problem, Π^K_{Stoch} as defined in (2.5) with an additional constraint that some of the first-stage decision variables $\boldsymbol{x}$ are required to be integers. Even for this case, we show that a finitely adaptable solution provides a good approximation.

THEOREM 8.1 *Consider the multi-stage stochastic problem Π^K_{Stoch}* (2.5), *with additional integer constraints on some first stage decision variables. Suppose $\mathbb{E}_{\boldsymbol{b}}[\boldsymbol{b} \mid \boldsymbol{b} \in U^k_j] = \boldsymbol{u}_{k,j}$ for all $j = 1, \ldots, N_k$, $k = 1, \ldots, K-1$. Let $s = \min_{k,j} \mathbf{sym}(\boldsymbol{u}_{k,j}, \mathcal{U}^k_j)$ and $\rho = \max_{k,j} \rho(\boldsymbol{u}_{k,j}, \mathcal{U}^k_j)$. Then there is a finitely* *adaptable solution policy that can be computed efficiently and has at most $|\mathcal{P}|$ solutions in each stage, where $\mathcal{P}$ is the set of directed paths from the root node to nodes in Stage K in the multi-stage uncertainty network, such that the expected cost is at most $\lceil(1+\rho/s)\rceil$ times the optimal cost.*

The proof of Theorem 8.1 is exactly similar to the proof of Theorem 5.1 except that we need to handle the integrality constraints in constructing a feasible finitely adaptable solution. Therefore, instead of scaling the optimal fully-adaptable solution by a factor of $(1+\rho/s)$ to construct a feasible finitely adaptable solution,we need to scale by $\lceil(1+\rho/s)\rceil$ to preserve the integrality constraints. This implies that the performance ratio of the finitely adaptable solution with respect to an optimal fully-adaptable solution is at most $\lceil(1+\rho/s)\rceil$. Note that the stochastic problem, with both right hand side and objective coefficients uncertainty, can not be well approximated by a finitely adaptable solution even without the integrality constraints.

8.2 Multi-stage adaptive optimization problem. For the multi-stage adaptive problem, we can handle integer decision variables in all stages. In particular, we have the following theorems.

THEOREM 8.2 *Consider the multi-stage adaptive problem, Π^K_{Adapt} (2.6), with additional integer constraints on some decision variables in each stage. Suppose $s = \min_{k,j} \mathbf{sym}(\mathcal{U}^k_j)$ and $\boldsymbol{u}_{k,j} \in \mathcal{U}^k_j$ is the point of symmetry for all $j = 1,\ldots,N_k$, $k = 1,\ldots,K-1$. Let $\rho = \max_{k,j} \rho(\mathcal{U}^k_j)$ for all $j = 1,\ldots,N_k$, $k = 1,\ldots,K-1$. Then there is a finitely adaptable solution policy that can be computed efficiently and has at most $|\mathcal{P}|$ solutions where $\mathcal{P}$ is the set of directed paths from the root node to any node in Stage K of the multi-stage uncertainty network, such that its worst-case cost is at most $\lceil(1+\rho/s)\rceil$ times the optimal cost.*

THEOREM 8.3 *Consider the multi-stage adaptive problem, $\Pi^K_{\text{Adapt}(\boldsymbol{b},\boldsymbol{d})}$ (2.8), with additional integer constraints on some decision variables in each stage. Suppose $s = \min_{k,j} \mathbf{sym}(\mathcal{U}^k_j)$ and $\boldsymbol{u}_{k,j} \in \mathcal{U}^k_j$ is the point of symmetry for all $j = 1,\ldots,N_k$, $k = 1,\ldots,K-1$. Also, let $\boldsymbol{u}^{\boldsymbol{b}}_{k,j}, \boldsymbol{u}^{\boldsymbol{d}}_{k,j}$ denote the right hand side and the objective coefficient uncertainty in $\boldsymbol{u}_{k,j}$ respectively. Let $\rho = \max_{k,j} \rho(\boldsymbol{u}_{k,j}, \mathcal{U}^k_j)$ for all $j = 1,\ldots,N_k$, $k = 1,\ldots,K-1$. Then there is a finitely adaptable solution policy that can be computed efficiently and has at most $|\mathcal{P}|$ solutions where $\mathcal{P}$ is the set of directed paths from the root node to any node in Stage K of the multi-stage uncertainty network, such that its worst-case cost is at most $\lceil(1+\rho/s)\rceil \cdot (1+\rho/s)$ times the optimal cost of $\Pi^K_{\text{Adapt}(\boldsymbol{b},\boldsymbol{d})}$.*

9. Conclusions. In this paper, we propose tractable solution policies for two-stage and multi-stage stochastic and adaptive optimization problems and relate the performance of the approximate solution approaches with the fundamental geometric properties of the uncertainty set. For a fairly general stochastic optimization problem, we show that the performance of a static robust solution for the two-stage problem and a finitely adaptable solution for the multi-stage problem is related to the symmetry and translation factor of the uncertainty sets. In particular, the performance bound is $(1+\rho/s)$ where ρ is the translation factor of the uncertainty sets and s is the symmetry. We also show that the bound is tight, i.e., given any symmetry and translation factor, there exists a family of instances where the uncertainty sets have the given symmetry and translation factor, and the cost of an optimal static robust solution is exactly equal to $(1 + \rho/s)$ times the optimal stochastic cost. For most commonly used uncertainty sets, the performance bound gives quite interesting results. For instance, if the sets are perfectly symmetric, i.e., $s = 1$, the bound is less than or equal to 2. Refer to Table 1 for a list of examples of several interesting uncertainty sets and corresponding bounds. In any model, the uncertainty set is the modeler's choice. Our bound offers important insights in this choice of uncertainty set as well.

While we show that the stochastic problem with only right hand side uncertainty can be well approximated, the static robust solution and the finitely adaptable solution are not a good approximation for the case where both the right hand side and the objective coefficients are uncertain. However, for the adaptive optimization problem, we show that the static robust and the finitely adaptable solution are a good approximation for the two-stage and multi-stage version respectively, even when both the right hand side and the objective coefficients are uncertain. The performance bound in this case is $(1 + \rho/s)^2$ where again ρ is the translation factor of the uncertainty sets and s is the symmetry. This bound is not as strong as the bound for the stochastic problem. We also present an alternate geometric bound for this case.

Acknowledgements. We thank the associate editor and the reviewers of the paper for several insightful comments and suggestions. The research is partially supported by NSF Grant EFRI-0735905.

References

[1] E.M.L. Beale. On Minizing A Convex Function Subject to Linear Inequalities. *Journal of the Royal Statistical Society. Series B (Methodological)*, 17(2):173–184, 1955.

[2] A. Belloni and R. M. Freund. On the symmetry function of a convex set. *Mathmematical Programming*, 111(1):57–93, 2007.

[3] A. Bemporad, F. Borrelli, and M. Morari. Min-max control of constrained uncertain discrete-time linear systems. *IEEE Transactions on Automatic Control*, 48(9):1600–1606, 2003.

[4] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust optimization*. Princeton Univ Press, 2009.

[5] A. Ben-Tal, A. Goryashko, E. Guslitzer, and A. Nemirovski. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2):351–376, 2004.

[6] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 23(4):769–805, 1998.

[7] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25(1):1–14, 1999.

[8] A. Ben-Tal and A. Nemirovski. Robust optimization–methodology and applications. *Mathematical Programming*, 92(3):453–480, 2002.

[9] D. Bertsimas, D.B. Brown, and C. Caramanis. Theory and applications of Robust Optimization. *SIAM Review (To Appear)*, 2010.

[10] D. Bertsimas and C. Caramanis. Finite adaptability in linear optimization. *IEEE Automatic Control (To Appear)*, 2010.

[11] D. Bertsimas and V. Goyal. On the power of robust solutions in two-stage stochastic and adaptive optimization problems. *Mathematics of Operations Research*, 35(2):284–305, 2010.

[12] D. Bertsimas, D.A. Iancu, and P.A. Parrilo. Optimality of Affine Policies in Multistage Robust Optimization. *Mathematics of Operations Research*, 35(2):363–394, 2010.

[13] D. Bertsimas and M. Sim. Robust Discrete Optimization and Network Flows. *Mathematical Programming Series B*, 98:49–71, 2003.

[14] D. Bertsimas and M. Sim. The Price of Robustness. *Operation Research*, 52(2):35–53, 2004.

[15] G.B. Dantzig. Linear programming under uncertainty. *Management Sci.*, 1:197–206, 1955.

[16] M. Dyer and L. Stougie. Computational complexity of stochastic programming problems. *Mathematical Programming*, 106(3):423–432, 2006.

[17] 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.

[18] S.J. Garstka and R.J.B. Wets. On decision rules in stochastic programming. *Mathematical Programming*, 7(1):117–143, 1974.

[19] D. Goldfarb and G. Iyengar. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1):1–38, 2003.

[20] A. Gupta, M. Pál, R. Ravi, and A. Sinha. Boosted sampling: approximation algorithms for stochastic optimization. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 417–426, 2004.

[21] A. Gupta, M. Pal, R. Ravi, and A. Sinha. What about Wednesday? Approximation algorithms for multistage stochastic optimization. *Lecture notes in computer science*, pages 86–98, 2005.

[22] G. Infanger. *Planning under uncertainty: solving large-scale stochastic linear programs.* Boyd & Fraser Pub Co, 1994.

[23] F. John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays, presented to R. Courant on his 60th Birthday, January 8, 1948*, 1948.

[24] P. Kall and S.W. Wallace. *Stochastic programming.* Wiley New York, 1994.

[25] H. Minkowski. Allegemeine lehzätze über konvexe polyeder. *Ges. Ahb.*, 2:103–121, 1911.

[26] A. Prékopa. *Stochastic programming.* Kluwer, 1995.

[27] RT Rockafellar and R.J.B. Wets. The optimal recourse problem in discrete time: L^1-multipliers for inequality constraints. *SIAM J. Control Optimization*, 16:16–36, 1978.

[28] A. Shapiro. Stochastic programming approach to optimization under uncertainty. *Mathematical Programming, Series B*, 112(1):183–220, 2008.

[29] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on stochastic programming: modeling and theory.* Society for Industrial and Applied Mathematics, 2009.

[30] A. Shapiro, D. Dentcheva, A. Ruszczyński, and A.P. Ruszczyński. *Lectures on stochastic programming: modeling and theory.* Society for Industrial Mathematics, 2009.

[31] A. Shapiro and A. Nemirovski. On complexity of stochastic programming problems. *Continuous Optimization: Current Trends and Applications*, V. Jeyakumar and A.M. Rubinov (Eds.):111–144, 2005.

[32] D.B. Shmoys and C. Swamy. Stochastic Optimization is (Almost) as easy as Deterministic Optimization. In *Annual Symposium on Foundations of Computer Science*, volume 45, pages 228–237, 2004.

[33] C. Swamy and D.B. Shmoys. Sampling-based approximation algorithms for multi-stage stochastic optimization. In *46th Annual IEEE Symposium on Foundations of Computer Science, 2005. FOCS 2005.*, pages 357–366, 2005.

Appendix A. Examples: Symmetry of specific sets.

Our main tool in calculating the symmetry of a convex compact set is the following proposition in Belloni and Freund [2],

PROPOSITION A.1 (BELLONI AND FREUND [2]) *Let S be a convex body, and consider the representation of S as the intersection of halfspaces: $S = \{x \in \mathbb{R}^m | a_i^T \boldsymbol{x} \le b_i,\ i \in I\}$ for some (possibly unbounded) index set I, and let $\delta_i^* := \max_{\boldsymbol{x}\in S}\{-a_i^T \boldsymbol{x}\}$ for $i \in I$. Then for all $\boldsymbol{x} \in S$,*

$$\mathbf{sym}(\boldsymbol{x}, S) = \inf_{i\in I}\left\{\frac{b_i - a_i^T \boldsymbol{x}}{\delta_i^* + a_i^T \boldsymbol{x}}\right\}.$$

A.1 General L_p half-ball for $p \ge 1$.

Define an L_p half-ball as

$$\mathbf{HB}_p := \{\boldsymbol{b} \in \mathbb{R}^m \mid \|\boldsymbol{b}\|_p \le 1, b_1 \ge 0\}.$$

The dual norm is the L_q norm with $\frac{1}{p} + \frac{1}{q} = 1$. The symmetry of L_p half-ball is summarized as follows,

PROPOSITION A.2 *The symmetry and point of symmetry of an L_p half-ball are,*

$$\mathbf{sym}(\mathbf{HB}_p) = \left(\frac{1}{2}\right)^{\frac{1}{p}}, \quad \boldsymbol{b}_0(\mathbf{HB}_p) = \frac{1}{2^{\frac{1}{p}} + 1}\boldsymbol{e_1}.$$

PROOF. The L_p half-ball can be represented by halfspaces as

$$\mathbf{HB}_p = \{\boldsymbol{b} \in \mathbb{R}^m \mid -b_1 \le 0, \boldsymbol{\pi}^T \boldsymbol{b} \le 1, \forall \|\boldsymbol{\pi}\|_q \le 1, \pi_1 \ge 0\}.$$

For each $\boldsymbol{\pi}$ in the dual unit ball (i.e., $\|\boldsymbol{\pi}\|_q = 1$), define

$$\delta^*(\boldsymbol{\pi}) := \max_{\|\boldsymbol{b}\|_p \le 1, b_1 \ge 0} -\boldsymbol{\pi}^T \boldsymbol{b},$$

whose optimum $\boldsymbol{b}^*$ satisfies $b_1^* = 0, \|\boldsymbol{b}^*\|_p = 1$. Therefore we have,

$$\delta^*(\boldsymbol{\pi}) = \max_{\|\tilde{\boldsymbol{b}}\|_q = 1} -\tilde{\boldsymbol{\pi}}^T \tilde{\boldsymbol{b}} = \|\tilde{\boldsymbol{\pi}}\|_q = (1 - \pi_1^q)^{1/q},$$

where $\boldsymbol{b} = (b_1; \tilde{\boldsymbol{b}})$ and $\boldsymbol{\pi} = (\pi_1; \tilde{\boldsymbol{\pi}})$. Also define $\delta^*(\boldsymbol{e}_1) := \max_{\boldsymbol{b}\in\mathbf{HB}_p} \boldsymbol{b}_1 = 1$. Now we can compute the symmetry of $\mathbf{HB}_p$. Due to the geometry, the point of symmetry has the form $\boldsymbol{b}_0 = \alpha \boldsymbol{e}_1$. We have,

$$\mathbf{sym}(\mathbf{HB}_p) = \max_{\alpha\in[0,1]} \min\left\{\inf_{\pi_1\in[0,1]} \frac{1 - \pi_1\alpha}{(1 - \pi_1^q)^{\frac{1}{q}} + \pi_1\alpha}, \frac{\alpha}{1 - \alpha}\right\}.$$

The maximum is achieved when the inf term is equal to the second term, because the inf term is decreasing in α and the second term is increasing in α. The inf term has optimality condition,

$$-\alpha((1 - \pi_1^q)^{\frac{1}{q}} + \pi_1\alpha) = (1 - \pi_1\alpha)\left(\frac{-\pi_1^{q-1}}{(1 - \pi_1)^{\frac{1}{p}}} + \alpha\right).$$

Therefore, we have,

$$\frac{\alpha}{\frac{-\pi_1^{q-1}}{(1-\pi_1)^{\frac{1}{p}}}-\alpha}=\frac{\alpha}{1-\alpha},$$

which implies that $\frac{-\pi_1^{q-1}}{(1-\pi_1)^{\frac{1}{p}}}=1$, i.e., $\pi_1=(\frac{1}{2})^{\frac{1}{q}}$, and $\alpha=\frac{1}{2^{\frac{1}{p}}+1}$. □

A.2 Intersection of an L_p ball with $\mathbb{R}^m_+$ for $p\geq 1$. Recall that an L_p-ball intersected with the nonnegative orthant is defined as $\mathbf{B}_p^+ := \{\boldsymbol{b}\in\mathbb{R}^m \mid \|\boldsymbol{b}\|_p\leq 1, \boldsymbol{b}\geq 0\}$. We show that the symmetry and symmetry point of $\mathbf{B}_p^+$ are as defined in (4.7). The L_p half-ball can be represented by halfspaces as

$$\mathbf{B}_p^+=\{\boldsymbol{b}\in\mathbb{R}^m \mid -\boldsymbol{b}\leq \mathbf{0}, \boldsymbol{\pi}^T\boldsymbol{b}\leq 1, \forall\|\boldsymbol{\pi}\|_q\leq 1, \boldsymbol{\pi}\geq\mathbf{0}\},$$

where $\|\cdot\|_q$ is the dual-norm with $\frac{1}{p}+\frac{1}{q}=1$. Therefore, define $\delta^*(\boldsymbol{\pi}) := \max_{\boldsymbol{b}\in\mathbf{B}_p^+} -\boldsymbol{\pi}^T\boldsymbol{b}=0$, and $\delta^*(\boldsymbol{e}_k) := \max_{\boldsymbol{b}\in\mathbf{B}_p^+} b_k=1$. Therefore, the symmetry can be computed as,

$$\mathbf{sym}(\mathbf{B}_p^+) := \max_{\alpha\in[0,(\frac{1}{m})^{\frac{1}{p}}]}\min\left\{\inf_{\|\boldsymbol{\pi}\|_q=1,\boldsymbol{\pi}\geq\mathbf{0}}\frac{1-\alpha\boldsymbol{e}^T\boldsymbol{\pi}}{\alpha\boldsymbol{e}^T\boldsymbol{\pi}}, \frac{\alpha}{1-\alpha}\right\},$$

where we use the property that the symmetry point $\boldsymbol{b}_0=\alpha\boldsymbol{e}$ for some $\alpha\in[0,1/\|\boldsymbol{e}\|_p]$. The maximum is achieved when the inf term is equal to the second term. The inf term can be computed, since $\max_{\|\boldsymbol{\pi}\|_q=1,\boldsymbol{\pi}\geq\mathbf{0}}\boldsymbol{e}^T\boldsymbol{\pi}=\|\boldsymbol{e}\|_p$. Thus, at the symmetry point, we have $\frac{1}{\alpha\|\boldsymbol{e}\|_p}-1=\frac{\alpha}{1-\alpha}$. Therefore, $\alpha=\frac{1}{\|\boldsymbol{e}\|_p+1}=\frac{1}{m^{\frac{1}{p}}+1}$, which gives the results.

A.3 Ellipsoidal uncertainty set. An ellipsoidal uncertainty set that is contained in the nonnegative orthant can be defined as,

$$\mathcal{U} := \{\boldsymbol{b} \mid \|\boldsymbol{E}(\boldsymbol{b}-\overline{\boldsymbol{b}})\|_2\leq 1\}\subset\mathbb{R}^m_+. \tag{A.1}$$

We assume the ellipsoid has full dimension, thus, $\overline{\boldsymbol{b}}>\mathbf{0}$. The symmetry of an ellipsoid is 1. But the translation factor depends on the position of the center $\overline{\boldsymbol{b}}$. The following proposition computes the translation factor.

Proposition A.3 *Assume the uncertainty set $\mathcal{U}$ is defined in (A.1). The translation factor $\rho(\overline{\boldsymbol{b}},\mathcal{U})$ is given as,*

$$\rho(\overline{\boldsymbol{b}},\mathcal{U})=\max_{1\leq i\leq m}\frac{\sqrt{E_{ii}^{-1}}}{\overline{b}_i},$$

where E_{ii}^{-1} is the i-th diagonal element of the inverse matrix E^{-1}.

Proof. From the definition, the translation factor is the smallest ρ such that

$$b_i^l := \min\{b_i \| \|\boldsymbol{E}(\boldsymbol{b}-\rho\overline{\boldsymbol{b}})\|_2\leq 1\}\geq 0, \quad \forall i=1,2,\ldots,m.$$

From the optimality conditions, we can get that $b_i^l = \rho\bar{b}_i - \sqrt{E_{ii}^{-1}}$, which gives the result. □

A.4 Intersection of two L_p balls with $\mathbb{R}_+^m$. Consider the uncertainty set $\mathcal{U}$ defined in (4.9) where $1 \le p_1 < p_2$, $0 < r < 1$, and suppose (4.10) holds. We show that the symmetry and symmetry point of $\mathcal{U}$ are as defined in (4.11).

The uncertainty set $\mathcal{U}$ can be represented by the intersection of halfspaces,

$$\mathcal{U} = \{\boldsymbol{b} \in \mathbb{R}^m \mid -\boldsymbol{b} \le \boldsymbol{0}, \boldsymbol{\pi}^T\boldsymbol{b} \le 1, \boldsymbol{\lambda}^T\boldsymbol{b} \le r, \forall \|\boldsymbol{\pi}\|_{q_1} \le 1, \boldsymbol{\pi} \ge \boldsymbol{0}, \|\boldsymbol{\lambda}\|_{q_2} \le r, \boldsymbol{\lambda} \ge \boldsymbol{0}\},$$

where $\frac{1}{p_1} + \frac{1}{q_1} = 1$, $\frac{1}{p_2} + \frac{1}{q_2} = 1$. Compute the following quantities,

$$\delta^*(\boldsymbol{\pi}) := \max_{\boldsymbol{b}\in\mathcal{U}} -\boldsymbol{\pi}^T\boldsymbol{b} = 0, \quad \forall \|\boldsymbol{\pi}\|_{q_1} \le 1, \boldsymbol{\pi} \ge \boldsymbol{0},$$

$$\delta^*(\boldsymbol{\lambda}) := \max_{\boldsymbol{b}\in\mathcal{U}} -\boldsymbol{\lambda}^T\boldsymbol{b} = 0, \quad \forall \|\boldsymbol{\lambda}\|_{q_2} \le r, \boldsymbol{\lambda} \ge \boldsymbol{0},$$

$$\delta^*(\boldsymbol{e}_k) := \max_{\boldsymbol{b}\in\mathcal{U}} b_k = r, \quad \forall k = 1, \ldots, m.$$

Since the set $\mathcal{U}$ is symmetric with respect to the direction $\boldsymbol{e}$, the symmetry point must have the form $\boldsymbol{b}_0 = \alpha\boldsymbol{e}$. The symmetry can be computed as,

$$\mathbf{sym}(\mathcal{U}) := \max_{\alpha \in [0, (\frac{1}{m})^{\frac{1}{p_1}}]} \min\left\{ \inf_{\|\boldsymbol{\pi}\|_{q_1}=1, \boldsymbol{\pi}\ge\boldsymbol{0}} \frac{1-\alpha\boldsymbol{e}^T\boldsymbol{\pi}}{\alpha\boldsymbol{e}^T\boldsymbol{\pi}}, \inf_{\|\boldsymbol{\lambda}\|_{q_2}=1, \boldsymbol{\lambda}\ge\boldsymbol{0}} \frac{1-\alpha\boldsymbol{e}^T\boldsymbol{\lambda}}{\alpha\boldsymbol{e}^T\boldsymbol{\lambda}}, \frac{\alpha}{r-\alpha} \right\}, \tag{A.2}$$

As is shown in the proof of the previous proposition, we have,

$$\inf_{\|\boldsymbol{\pi}\|_{q_1}=1, \boldsymbol{\pi}\ge\boldsymbol{0}} \frac{1-\alpha\boldsymbol{e}^T\boldsymbol{\pi}}{\alpha\boldsymbol{e}^T\boldsymbol{\pi}} = \frac{1}{\alpha\|\boldsymbol{e}\|_{p_1}} - 1,$$
$$\inf_{\|\boldsymbol{\lambda}\|_{q_2}=1, \boldsymbol{\lambda}\ge\boldsymbol{0}} \frac{1-\alpha\boldsymbol{e}^T\boldsymbol{\lambda}}{\alpha\boldsymbol{e}^T\boldsymbol{\lambda}} = \frac{r}{\alpha\|\boldsymbol{e}\|_{p_2}} - 1.$$

Using condition (4.10), we know the first term is dominated by the second term for any α, therefore, the maximum in the symmetry formula (A.2) is achieved when

$$\frac{1}{\alpha\|\boldsymbol{e}\|_{p_1}} - 1 = \frac{\alpha}{r-\alpha}.$$

Therefore, $\alpha = \dfrac{r}{r\|\boldsymbol{e}\|_{p_1} + 1} = \dfrac{r}{rm^{\frac{1}{p_1}} + 1}$, which gives the results.

A.5 Budgeted uncertainty set. An important type of uncertainty sets is the budgeted uncertainty set, Δ_k as defined in (4.12). We show that the symmetry and the symmetry point are as defined in (4.13).

First, we observe that the symmetry point $\boldsymbol{b}_0(\Delta_k)$ must be of the form $\boldsymbol{b}_0 = \alpha\boldsymbol{e}$, due to the geometry of Δ_k. Then, use Proposition A.1, for $k \ge 1$, we have,

$$\mathbf{sym}(\Delta_k) = \max_{0\le\alpha\le\frac{k}{m}} \min\left\{\frac{\alpha}{1-\alpha}, \frac{k-m\alpha}{m\alpha}\right\}.$$

By the monotonicity of $\frac{\alpha}{1-\alpha}$ and $\frac{k-m\alpha}{m\alpha}$, the maximum is achieved at $\alpha = \frac{k}{m+k}$. Thus, the symmetry point of Δ_k is $\boldsymbol{b}_0 = \frac{k}{m+k}e$ and $\mathbf{sym}(\Delta_k) = \frac{k}{m}$.

A.6 Demand uncertainty set. Proof of Proposition 4.1. For $\mu \geq \Gamma$, the hypercube centered at $\mu\boldsymbol{e}$ is completely contained in the positive orthant. For each S, define

$$\delta^*(S_+) := \max_{\boldsymbol{b}\in\mathbf{DU}} -\boldsymbol{e}_S^T\boldsymbol{b} = -|S|\mu + \sqrt{|S|}\Gamma,$$

$$\delta^*(S_-) := \max_{\boldsymbol{b}\in\mathbf{DU}} \boldsymbol{e}_S^T\boldsymbol{b} = |S|\mu + \sqrt{|S|}\Gamma.$$

The symmetry of $\boldsymbol{b}$ is given as follows,

$$\mathbf{sym}(\boldsymbol{b},\mathbf{DU}) = \min_{S\subseteq\mathbf{DU}}\left\{\frac{|S|\mu + \sqrt{|S|}\Gamma - \boldsymbol{e}_S^T\boldsymbol{b}}{-|S|\mu + \sqrt{|S|}\Gamma + \boldsymbol{e}_S^T\boldsymbol{b}}, \frac{-|S|\mu + \sqrt{|S|}\Gamma + \boldsymbol{e}_S^T\boldsymbol{b}}{|S|\mu + \sqrt{|S|}\Gamma - \boldsymbol{e}_S^T\boldsymbol{b}}\right\}.$$

Therefore, $\mathbf{sym}(\mu\boldsymbol{e},\mathbf{DU}) = 1$, which proves (4.15).

For $\frac{1}{\sqrt{m}}\Gamma < \mu < \Gamma$, μ is in one of the intervals of $[\frac{1}{\sqrt{2}}\Gamma,\Gamma), [\frac{1}{\sqrt{3}}\Gamma, \frac{1}{\sqrt{2}}\Gamma], \ldots, (\frac{1}{\sqrt{m}}\Gamma, \frac{1}{\sqrt{m-1}}\Gamma]$. We show that (4.16) holds for $\mu \in [\frac{1}{\sqrt{2}}\Gamma,\Gamma]$. A similar argument works for other intervals.

Let us assume $\mu \in [\frac{1}{\sqrt{2}}\Gamma,\Gamma]$. Notice that $|S|\mu - \sqrt{|S|}\Gamma \geq 0$ for all $|S| \geq 2$. Therefore, the constraints in the definition of $\mathbf{DU}$ (4.14) can be written as,

$$|S|\mu - \sqrt{|S|}\Gamma \leq \sum_{i\in S} b_i \leq |S|\mu + \sqrt{|S|}\Gamma, \quad \forall |S| \geq 2.$$

But for all $|S| = 1$, since $\mu - \Gamma < 0$, we will have $0 \leq b_i \leq \mu + \Gamma$ for all $i = 1,\ldots,m$. Thus, according to Proposition A.1, we can compute the following quantities for all $|S| \geq 2$,

$$\delta^*(S_+) := \max_{\boldsymbol{b}\in\mathbf{DU}} -\boldsymbol{e}_S^T\boldsymbol{b} = -|S|\mu + \sqrt{|S|}\Gamma,$$

$$\delta^*(S_-) := \max_{\boldsymbol{b}\in\mathbf{DU}} \boldsymbol{e}_S^T\boldsymbol{b} = |S|\mu + \sqrt{|S|}\Gamma,$$

and for $|S| = 1$, we have $\delta^*(i_+) := \max_{\boldsymbol{b}\in\mathbf{DU}} -b_i = 0$, $\delta^*(i_-) := \max_{\boldsymbol{b}\in\mathbf{DU}} b_i = \mu + \Gamma$ for all $i = 1,\ldots,m$. Since $\mathbf{DU}$ is symmetric about the direction $\boldsymbol{e}$, the point of symmetry has the form $\boldsymbol{b}_0 = \alpha\boldsymbol{e}$. The range of possible α is determined by the constraint $m\mu - \sqrt{m}\Gamma \leq m\alpha \leq m\mu + \sqrt{m}\Gamma$. Now we can compute the symmetry of $\alpha\boldsymbol{e}$,

$$\mathbf{sym}(\alpha\boldsymbol{e},\mathbf{DU}) = \min\left\{\frac{(\mu+\Gamma)-\alpha}{\alpha}, \frac{\alpha}{(\mu+\Gamma)-\alpha}, \frac{(|S|\mu + \sqrt{|S|}\Gamma) - |S|\alpha}{(-|S|\mu + \sqrt{|S|}\Gamma) + |S|\alpha}, \frac{(-|S|\mu + \sqrt{|S|}\Gamma) + |S|\alpha}{(|S|\mu + \sqrt{|S|}\Gamma) - |S|\alpha}, \forall |S| \geq 2\right\}. \tag{A.3}$$

The symmetry of $\mathbf{DU}$ is given as,

$$\mathbf{sym}(\mathbf{DU}) = \max\left\{\mathbf{sym}(\alpha\boldsymbol{e},\mathbf{DU}) : \alpha \in [\mu - \frac{1}{\sqrt{m}}\Gamma, \mu + \frac{1}{\sqrt{m}}\Gamma]\right\}.$$

Observe that $\frac{(|S|\mu+\sqrt{|S|}\Gamma)-|S|\alpha}{(-|S|\mu+\sqrt{|S|}\Gamma)+|S|\alpha}$ is a decreasing function in α, and $\frac{(-|S|\mu+\sqrt{|S|}\Gamma)+|S|\alpha}{(|S|\mu+\sqrt{|S|}\Gamma)-|S|\alpha}$ is an increasing function in α. Both functions have value 1 when $\alpha = \mu$, and decrease when $|S|$ increases. The formula (A.3) can be simplified as,

$$\mathbf{sym}(\alpha\boldsymbol{e},\mathbf{DU}) = \min\left\{\frac{(\mu+\Gamma)-\alpha}{\alpha}, \frac{\alpha}{(\mu+\Gamma)-\alpha}, \frac{(m\mu + \sqrt{m}\Gamma) - m\alpha}{(-m\mu + \sqrt{m}\Gamma) + m\alpha}, \frac{(-m\mu + \sqrt{m}\Gamma) + m\alpha}{(m\mu + \sqrt{m}\Gamma) - m\alpha}\right\}. \tag{A.4}$$

The maximum of $\mathbf{sym}(\alpha \boldsymbol{e}, \mathbf{DU})$ is achieved at the following intersection,

$$\frac{\alpha}{(\mu+\Gamma)-\alpha} = \frac{(m\mu+\sqrt{m}\Gamma)-m\alpha}{(-m\mu+\sqrt{m}\Gamma)+m\alpha},$$

which gives the symmetry and the symmetry point in (4.16). Since (A.4) also holds when μ is in any other interval, it implies that (4.16) is true for all $\frac{1}{\sqrt{m}}\Gamma < \mu < \Gamma$.

When $0 \le \mu \le \frac{1}{\sqrt{m}}\Gamma$, following a similar argument, we have,

$$\mathbf{sym}(\alpha \boldsymbol{e}, \mathbf{DU}) = \min_{|S|\ge 1}\left\{\frac{(|S|\mu+\sqrt{|S|}\Gamma)-|S|\alpha}{|S|\alpha}, \frac{|S|\alpha}{(|S|\mu+\sqrt{|S|}\Gamma)-|S|\alpha}\right\}.$$

Then, the maximum is achieved when $\dfrac{\alpha}{(\mu+\Gamma)-\alpha} = \dfrac{(m\mu+\sqrt{m}\Gamma)-m\alpha}{m\alpha}$, which gives the results in (4.17). □

A.7 Parallel slabs.

A parallel slab is defined as $\mathbf{PS} := \{\boldsymbol{b} \in \mathbb{R}^m_+ \mid L \le \boldsymbol{e}^T\boldsymbol{b} \le U\}$ for $0 \le L \le U$ and $U > 0$. The symmetry and symmetry point of a parallel slab is given as,

Proposition A.4

$$\mathbf{sym}(\mathbf{PS}) = \frac{1}{m - \frac{L}{U}}, \tag{A.5}$$

$$\boldsymbol{b}_0(\mathbf{PS}) = \frac{U^2}{(m+1)U - L}\boldsymbol{e}. \tag{A.6}$$

Proof. Define the following quantities,

$$\delta^*_L := \max_{\boldsymbol{b}\in\mathbf{PS}} \boldsymbol{e}^T\boldsymbol{b} = U, \quad \delta^*_U := \max_{\boldsymbol{b}\in\mathbf{PS}} -\boldsymbol{e}^T\boldsymbol{b} = -L,$$

and for each $k = 1, \dots, m$,

$$\delta^*(\boldsymbol{e_k}) := \max_{\boldsymbol{b}\in\mathbf{PS}} \boldsymbol{e_k}^T\boldsymbol{b} = U.$$

According to the geometry of the set, the point of symmetry is of the form $\boldsymbol{b}_0 = \alpha\boldsymbol{e}$. Thus, the symmetry can be written as,

$$\mathbf{sym}(\mathbf{PS}) = \max_{\alpha\in[\frac{L}{m},\frac{U}{m}]} \min\left\{\frac{m\alpha - L}{U - m\alpha}, \frac{U - m\alpha}{m\alpha - L}, \frac{\alpha}{U-\alpha}\right\}.$$

The first two terms are inverse of each other, therefore one increasing and the other decreasing in α; the third term is increasing in α. Thus, the maximum is attained when the second term is equal to the third, i.e.,

$$\frac{U - m\alpha}{m\alpha - L} = \frac{\alpha}{U-\alpha},$$

which gives the results. □

A.8 $\mathcal{U} = \mathbf{Conv}(\Delta_1, \{e\})$. The convex hull of the standard m-simplex $\Delta_1 = \{\boldsymbol{b} \in \mathbb{R}^m \mid \boldsymbol{e}^T\boldsymbol{b} \leq 1, \boldsymbol{b} \geq \mathbf{0}\}$ and a point $\boldsymbol{e} = (1, \ldots, 1)^T$ has a slightly improved symmetry comparing with the simplex itself, as shown below.

Proposition A.5 *The symmetry and point of symmetry of $\mathcal{U} = \mathbf{Conv}(\Delta_1, \{\boldsymbol{e}\})$ are given as,*

$$\mathbf{sym}(\mathcal{U}) = \frac{1}{m-1}, \quad \boldsymbol{b}_0(\mathcal{U}) = \frac{1}{m}\boldsymbol{e}.$$

Proof. The set $\mathcal{U}$ can be represented by halfspaces as

$$\mathcal{U} = \{\boldsymbol{b} \in \mathbb{R}^m \mid -\boldsymbol{b} \leq \mathbf{0}, \boldsymbol{\pi}_i^T\boldsymbol{b} \leq 1, \forall i = 1, \ldots, m\},$$

where $\boldsymbol{\pi}_i$ has $(2-m)$ in the i-th entry and 1 elsewhere. For each i, define,

$$\delta_i^* = \max_{\boldsymbol{b}\in\mathcal{U}} \boldsymbol{b}_i = 1, \quad \delta(\boldsymbol{\pi}_i)^* = \max_{\boldsymbol{b}\in\mathcal{U}} -\boldsymbol{\pi}_i^T\boldsymbol{b} = m-2.$$

By the symmetry of $\mathcal{U}$, we know the symmetry point should be on the $\boldsymbol{e}$ direction. Therefore, we have,

$$\mathbf{sym}(\mathcal{U}) = \max_{\alpha\in[0,1]} \min\left\{\frac{1-\alpha\boldsymbol{\pi}_i^T\boldsymbol{e}}{(m-2)+\alpha\boldsymbol{\pi}_i^T\boldsymbol{e}}, \frac{\alpha}{1-\alpha}\right\}, \tag{A.7}$$

$$= \max_{\alpha\in[0,1]} \min\left\{\frac{1-\alpha}{m-2+\alpha}, \frac{\alpha}{1-\alpha}\right\}. \tag{A.8}$$

The maximum is achieved when $\frac{1-\alpha}{m-2+\alpha} = \frac{\alpha}{1-\alpha}$, which gives $\alpha = \frac{1}{m}$ and symmetry is $\frac{1}{m-1}$. □

A.9 The matrix simplex. Define the uncertainty set $\mathcal{U}$ as follows,

$$\mathcal{U} = \{\boldsymbol{B} \in \mathbb{S}_+^m \mid \boldsymbol{I} \bullet \boldsymbol{B} \leq 1\}, \tag{A.9}$$

where $\mathbb{S}_+^m$ is the cone of positive semidefinite matrices, $\boldsymbol{I}$ is the identity matrix. We use it as an example of uncertainty sets for robust SDP problems. The following proposition shows that $\mathcal{U}$ has a similar symmetry property as a simplex in $\mathbb{R}^m$.

Proposition A.6 *Let $\mathcal{U}$ be defined in (A.9) and $\boldsymbol{B}_0$ be the center of symmetry of $\mathcal{U}$. We have,*

$$\mathbf{sym}(\mathcal{U}) = \frac{1}{m}, \quad \boldsymbol{B}_0 = \frac{1}{m+1}\boldsymbol{I}.$$

Proof. The set $\mathcal{U}$ can be written in the form of intersection of halfspaces,

$$\mathcal{U} = \{\boldsymbol{B} \in \mathbb{S}^m \mid \boldsymbol{I} \bullet \boldsymbol{B} \leq 1, (\boldsymbol{q}\boldsymbol{q}^T) \bullet \boldsymbol{B} \leq 1, \forall \|\boldsymbol{q}\|_2 = 1\}.$$

Then, we can compute the following quantities,

$$\delta^*(\boldsymbol{q}) = \max_{\boldsymbol{B}\in\mathcal{U}} \quad (\boldsymbol{q}\boldsymbol{q}^T) \bullet \boldsymbol{B} = \max_{\boldsymbol{p}:\|\boldsymbol{p}\|=1} (\boldsymbol{q}^T\boldsymbol{p})^2 = 1, \tag{A.10}$$

$$\delta^*(\boldsymbol{I}) = -\min_{\boldsymbol{B}\in\mathcal{U}} \quad \boldsymbol{I} \bullet \boldsymbol{B} = 0. \tag{A.11}$$

The second equality in (A.10) comes from the fact that the extreme points of $\mathcal{U}$ are rank-1 matrices $\boldsymbol{p}\boldsymbol{p}^T$ with $\boldsymbol{p}^T\boldsymbol{p} = 1$. (A.11) follows from $\boldsymbol{I} \bullet \boldsymbol{B} \geq 0$ and the minimum is achieved at $\boldsymbol{B} = \boldsymbol{0}$. Now we can compute the symmetry of $\mathcal{U}$ as,

$$\textbf{sym}(\mathcal{U}) = \max_{\boldsymbol{B}\in\mathcal{U}} \min \left\{ \frac{1 - \boldsymbol{I} \bullet \boldsymbol{B}}{\boldsymbol{I} \bullet \boldsymbol{B}}, \inf_{\boldsymbol{q}:\|\boldsymbol{q}\|=1} \frac{(\boldsymbol{q}\boldsymbol{q}^T) \bullet \boldsymbol{B}}{1 - (\boldsymbol{q}\boldsymbol{q}^T) \bullet \boldsymbol{B}} \right\}. \tag{A.12}$$

To compute the inf term is equivalent to solving the following problem, which admits a closed form solution by Courant minimax theorem: $\lambda_1(\boldsymbol{B}) = \min_{\boldsymbol{q}:\|\boldsymbol{q}\|=1} \boldsymbol{q}^T\boldsymbol{B}\boldsymbol{q}$, where $\lambda_1(\boldsymbol{B})$ is the smallest eigenvalue of $\boldsymbol{B}$, denoted as $(\lambda_1, \ldots, \lambda_m)$ in ascending order. Then, (A.12) can be rewritten as,

$$\textbf{sym}(\mathcal{U}) = \max_{\boldsymbol{e}^T\boldsymbol{\lambda}\leq 1, \boldsymbol{\lambda}\geq \boldsymbol{0}} \min \left\{ \frac{1 - \boldsymbol{e}^T\boldsymbol{\lambda}}{\boldsymbol{e}^T\boldsymbol{\lambda}}, \frac{\lambda_1}{1 - \lambda_1} \right\}, \tag{A.13}$$

which can be reformulated as,

$$\textbf{sym}(\mathcal{U}) = \max_{z,\boldsymbol{\lambda}} \{ z \ : \ \boldsymbol{e}^T\boldsymbol{\lambda} \leq \frac{1}{z+1}, \boldsymbol{e}^T\boldsymbol{\lambda} \leq 1, \ \boldsymbol{\lambda} \geq \boldsymbol{0}, \lambda_i \geq \frac{z}{z+1}, \forall i = 1, \ldots, m. \}$$

From the constraints, we have

$$\frac{1}{z+1} \geq \boldsymbol{e}^T\boldsymbol{\lambda} \geq \frac{mz}{z+1} \Rightarrow z \leq \frac{1}{m}.$$

In fact, $z = 1/m$ can be achieve by the feasible solution $\lambda_i = 1/(m+1)$ for all i. This completes the proof. □

Appendix B. Proof of Theorem 7.1.

Proof of Theorem 7.1. Let $\mathcal{U}'$ be the translation of the uncertainty set $\mathcal{U}$ with the translation factor ρ, i.e. $\mathcal{U}' = \mathcal{U} - (1-\rho)\boldsymbol{b}_0$. Let $\boldsymbol{b}_1 := \boldsymbol{b}_0 - (1-\rho)\boldsymbol{b}_0$. Also let $\boldsymbol{z} := \boldsymbol{b}_0 - \boldsymbol{b}_1 = (1-\rho)\boldsymbol{b}_0$. Following a similar argument as in the proof of Theorem 4.1, we have

$$\left(1 + \frac{1}{s}\right)\boldsymbol{b}_1 \succeq_{\boldsymbol{\mathcal{K}}} \boldsymbol{b'}, \quad \forall \boldsymbol{b'} \in \mathcal{U}'.$$

Translating back to the original uncertainty set $\mathcal{U}$ by adding $\boldsymbol{z}$ on both sides, we have,

$$\begin{aligned}
&\left(1 + \frac{1}{s}\right)\boldsymbol{b}_1 + \boldsymbol{z} \succeq_{\boldsymbol{\mathcal{K}}} \boldsymbol{b}, \quad \forall \boldsymbol{b} \in \mathcal{U},\\
\Rightarrow &\left(1 + \frac{1}{s}\right)\rho\boldsymbol{b}_0 + (1-\rho)\boldsymbol{b}_0 \succeq_{\boldsymbol{\mathcal{K}}} \boldsymbol{b}, \quad \forall \boldsymbol{b} \in \mathcal{U},\\
\Rightarrow &\left(1 + \frac{\rho}{s}\right)\boldsymbol{b}_0 \succeq_{\boldsymbol{\mathcal{K}}} \boldsymbol{b}, \quad \forall \boldsymbol{b} \in \mathcal{U}.
\end{aligned} \tag{B.1}$$

For brevity, let $\tau := (1 + \rho/s)$. Suppose $(\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{b}), \mu\text{-a.e. } \boldsymbol{b} \in \mathcal{U})$ is an optimal solution of the stochastic optimization problem (7.1). We have,

$$\boldsymbol{A}(\tau\boldsymbol{x}) + \boldsymbol{B}(\tau\boldsymbol{y}(\boldsymbol{b})) \succeq_{\boldsymbol{\mathcal{K}}} \tau\boldsymbol{b}, \ \mu\text{-a.e. } \boldsymbol{b} \in \mathcal{U}.$$

Equivalently,

$$\boldsymbol{A}(\tau\boldsymbol{x}) + \boldsymbol{B}(\tau\boldsymbol{y}(\boldsymbol{b})) - \tau\boldsymbol{b} \in \boldsymbol{\mathcal{K}}, \ \mu\text{-a.e. } \boldsymbol{b} \in \mathcal{U}.$$

Therefore, the expectation with respect to the probability measure μ satisfies,

$$\boldsymbol{A}(\tau \boldsymbol{x}) + \boldsymbol{B}(\mathbb{E}_\mu[\tau \boldsymbol{y}(\boldsymbol{b})]) - \mathbb{E}_\mu[\tau \boldsymbol{b}] \in \boldsymbol{\mathcal{K}}.$$

Thus, we have

$$\boldsymbol{A}(\tau \boldsymbol{x}) + \boldsymbol{B}(\mathbb{E}_\mu[\tau \boldsymbol{y}(\boldsymbol{b})]) \succeq_{\boldsymbol{\mathcal{K}}} \mathbb{E}_\mu[\tau \boldsymbol{b}] \succeq_{\boldsymbol{\mathcal{K}}} \tau \boldsymbol{b}_0 \succeq_{\boldsymbol{\mathcal{K}}} \boldsymbol{b}, \ \forall \boldsymbol{b} \in \mathcal{U}.$$

The last inequality uses (B.1). Therefore, $(\tau \boldsymbol{x}, \mathbb{E}_\mu[\tau \boldsymbol{y}(\boldsymbol{b})])$ is a feasible static robust solution. Therefore, the cost of an optimal static robust solution is at most,

$$\boldsymbol{c}^T(\tau \boldsymbol{x}) + \boldsymbol{d}^T \mathbb{E}_\mu[\tau \boldsymbol{y}(\boldsymbol{b})] = \tau(\boldsymbol{c}^T \boldsymbol{x} + \boldsymbol{d}^T \mathbb{E}_\mu[\boldsymbol{y}(\boldsymbol{b})]).$$

Furthermore, $z^{\boldsymbol{\mathcal{K}}}_{\text{Stoch}} = \boldsymbol{c}^T \boldsymbol{x} + \boldsymbol{d}^T \mathbb{E}_\mu[\boldsymbol{y}(\boldsymbol{b})]$, which implies that the cost of an optimal static solution is at most $\tau \cdot z^{\boldsymbol{\mathcal{K}}}_{\text{Stoch}}$. □