

Distributionally Robust Optimization and Its Tractable Approximations

Joel Goh

Stanford Graduate School of Business, and NUS Business School, National University of Singapore, Singapore
joelgoh@stanford.edu

Melvyn Sim

NUS Business School and NUS Risk Management Institute, National University of Singapore, Singapore,
dscsim@nus.edu.sg

In this paper we focus on a linear optimization problem with uncertainties, having expectations in the objective and in the set of constraints. We present a modular framework to obtain an approximate solution to the problem that is distributionally robust and more flexible than the standard technique of using linear rules. Our framework begins by first affinely extending the set of primitive uncertainties to generate new linear decision rules of larger dimensions and is therefore more flexible. Next, we develop new piecewise-linear decision rules that allow a more flexible reformulation of the original problem. The reformulated problem will generally contain terms with expectations on the positive parts of the recourse variables. Finally, we convert the uncertain linear program into a deterministic convex program by constructing distributionally robust bounds on these expectations. These bounds are constructed by first using different pieces of information on the distribution of the underlying uncertainties to develop separate bounds and next integrating them into a combined bound that is better than each of the individual bounds.

Subject classifications: programming: stochastic.

Area of review: Optimization.

History: Received April 2009; revision received July 2009; accepted October 2009. Published online in *Articles in Advance* April 23, 2010.

1. Introduction

Traditionally, robust optimization has been used to immunize deterministic optimization problems against infeasibility caused by perturbations in model parameters, while simultaneously preserving computational tractability. The general approach involves reformulation of the original uncertain optimization problem into a deterministic convex program, such that each feasible solution of the new program is feasible for all allowable realizations of the model uncertainties. The deterministic program is therefore “robust” against perturbations in the model parameters. This approach dates back to Soyster (1973), who considered a deterministic linear optimization model that is feasible for all data lying in a convex set. Recent works using this general approach include Ben-Tal and Nemirovski (1998, 1999, 2000), Bertsimas and Sim (2004), and El Ghaoui and Lebret (1997), and El Ghaoui et al. (1998).

Ben-Tal et al. (2004) noted that the traditional robust optimization approach was limited in the sense that it only allowed for all decisions to be made before the realization of the underlying uncertainties. They noted that in the modeling of real-world problems with multiple stages it might be permissible for a subset of the decisions to be made after the realization of all or part of the

underlying uncertainties. To overcome this limitation, the authors introduced the Affinely Adjustable Robust Counterpart (AARC), which allowed for delayed decisions that are affinely dependent upon the primitive uncertainties. Chen and Zhang (2009) also introduced the Extended Affinely Adjustable Robust Counterpart (EAARC) as an extension of the AARC by an affine reparameterization the primitive uncertainties. In a related work, Chen et al. (2008) introduced several piecewise-linear decision rules which are more flexible than (and improve upon) regular LDRs, and they show that under their new rules computational tractability is preserved.

Typically, robust optimization problems do not require specifications of the exact distribution of the model uncertainties. This is the general distinction between the approaches of robust optimization and stochastic programming toward modeling problems with uncertainties. In the latter, uncertainties are typically modeled as random variables with known distributions and have been used to obtain analytic solutions to important classes of problems (see, e.g., Birge and Louveaux 1997, Ruszczyński and Shapiro 2003). In the framework of robust optimization, however, uncertainties are usually modeled as random variables with true distributions that are unknown to the modeler but are constrained to lie within a known support. Each approach

has its advantages: if the exact distribution of uncertainties is precisely known, optimal solutions to the robust problem would be overly and unnecessarily conservative. Conversely, if the assumed distribution of uncertainties is in fact different from the actual distribution, the optimal solution using a stochastic programming approach may perform poorly. Bertsimas and Thiele (2006) reported computational results for an inventory model showing that even in the case when the assumed demand distribution had identical first and second moments to the actual demand distribution, an inventory policy which is heavily tuned to the assumed distribution might perform poorly when used against the true distribution.

A body which aims to bridge the gap between the conservatism of robust optimization and the specificity of stochastic programming is the minimax stochastic programming approach, where optimal decisions are sought for the *worst-case* probability distributions within a family of possible distributions, defined by certain properties such as their support and moments. This approach was pioneered by Žáčková (1966) and studied in many other works (e.g., Dupačová 1987, Breton and El Hachem 1995, Shapiro and Kleywegt 2002). This approach has seen numerous applications, dating back to the study by Scarf (1958) of an optimal single-product newsvendor problem under an unknown distribution with known mean and variance, as well as the subsequent simplification of Scarf's proof by Gallego and Moon (1993) and their extensions to include recourse and fixed cost. El Ghaoui et al. (2003) developed worst-case Value-at-Risk bounds for a robust portfolio selection problem when only the bounds on the means and covariance matrix of the assets are known. Chen et al. (2007) introduced directional deviations as an additional means to characterize a family of distributions that were applied by Chen and Sim (2009) to a goal-driven optimization problem. In a recent work, Delage and Ye (2010) study distributionally robust stochastic programs where the mean and covariance of the primitive uncertainties are themselves subject to uncertainty.

Our paper aims to extend this body of work in a similar direction: for a linear optimization problem with partially characterized uncertainties, we seek a solution that is *distributionally robust*, i.e., feasible for the worst-case probability distribution within the family of distributions. The model that we study is different from most minimax stochastic programs in that we allow for expectations of recourse variables in the constraint specifications. In addition, our model allows for nonanticipativity requirements, which occurs in many practical problems (e.g., multistage problems). Such problems are known to be difficult to solve exactly (see Shapiro and Nemirovski 2005), but are important in practice. We approach the problem by first using a simple LDR model of recourse to tractably approximate the problem, and subsequently build more complex piecewise-linear decision rules to improve the quality of the approximation. The overarching motivation for our work is to

design a general framework for modeling and solving linear robust optimization problems, which can then be automated. In an ongoing parallel work (Goh and Sim 2009), we are concurrently designing software to model robust problems within this framework. The key contributions of this paper are summarized below:

1. We present a new flexible nonanticipative decision rule, which we term the bideflected LDR, which generalizes both the previous deflected LDR of Chen et al. (2008) and the truncated LDR of See and Sim (2009). Furthermore, being nonanticipative, our new decision rule is also suitable for multistage modeling. We show that our new decision rule is an improvement over the original deflected LDR as well as the standard LDR.

2. We discuss a technique of segregating the primitive uncertainties to obtain a new set of uncertainties. We show that by applying LDRs on the new segregated uncertainties, we obtain decision rules (which we term Segregated LDRs, or SLDRs) that are more flexible than the original LDRs and which preserve the nonanticipativity requirements of the original LDRs. We study how these SLDRs can be used in conjunction with other partially known characteristics of the original uncertainty distribution (such as its mean and covariance) to construct distributionally ambiguous bounds on the expected positive part of an SLDR.

This paper is structured as follows: In §2 we present the general optimization problem that we attempt to solve, discussing our motivation and some applications. In §3 we discuss the model of uncertainty, which we will use for the rest of the paper, and highlight the distributional properties of the model uncertainties which we assume we have knowledge of. In §4 we discuss a tractable linear approximation to the general problem and how segregated uncertainties can be used to improve the quality of the approximation. In §5 we present and extend existing known bounds on the expected value of the positive part of a random variable. This is used in §6, where we introduce the two-stage and nonanticipative bideflected LDR and discuss its properties. Section 7 concludes. The mathematical proofs in this paper are relegated to Appendices A, B, and C, which can be found in the electronic companion to this paper.

Notation. We denote a random variable by the tilde sign, i.e., \tilde{x} . Bold lower case letters such as \mathbf{x} represent vectors and the upper case letters such as \mathbf{A} denote matrices. In addition, $x^+ = \max\{x, 0\}$ and $x^- = \max\{-x, 0\}$. The same notation can be used on vectors, such as \mathbf{y}^+ and \mathbf{z}^- , which denotes that the corresponding operations are performed componentwise. For any set S we will denote by $\mathbb{1}_S$ the indicator function on the set. Also we will denote by $[N]$ the set of positive running indices to N , i.e., $[N] = \{1, 2, \dots, N\}$, for some positive integer N . For completeness, we assume $[0] = \emptyset$. We also denote with a superscripted letter “ c ” the complement of a set, e.g., I^c . We denote by \mathbf{e} the vector of all ones and by \mathbf{e}^i the i th standard basis vector. In addition, we denote the identity matrix

using the upper case boldface symbol \mathbf{I} and, for brevity, we will omit specifying its dimension where it is contextually clear. We will use superscripted indices on vectors to index members of a collection of vectors, while subscripted indices on a vector denotes its components, i.e., $x_i = \mathbf{e}^i \mathbf{x}$. Finally, we distinguish between models which are intractable against those that are computationally tractable, by denoting the optimal objectives of intractable functions with a superscripted asterisk, i.e., Z^* .

2. Linear Optimization with Expectation Constraints

2.1. General Model

Let $\tilde{\mathbf{z}} = (\tilde{z}_1, \dots, \tilde{z}_N)$ be a vector of N random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. $\tilde{\mathbf{z}}$ represents the primitive uncertainties of our model. We do not presume knowledge of the actual joint distribution of $\tilde{\mathbf{z}}$. Instead, we shall assume that the true joint distribution \mathbb{P} lies in some family of distributions \mathbb{F} . We shall denote by $\mathbf{x} \in \mathfrak{R}^n$ the vector of decision variables, representing the *here-and-now* decisions which are unaffected by realizations of the primitive uncertainties. We also optimize over a set of K *wait-and-see* decision rules (also known as recourse variables), denoted by $\mathbf{y}^k(\cdot) \in \mathfrak{R}^{m_k}$, which are functions of the primitive uncertainties. In general, each decision rule may only depend on a subset of the primitive uncertainties. For each $k \in [K]$, we denote by $I_k \subseteq [N]$ the index set of dependent uncertainties for $\mathbf{y}^k(\cdot)$. Furthermore, for any index set $I \subseteq [N]$, we denote by $\mathcal{Y}(m, N, I)$ the space of allowable recourse decisions, which are measurable functions, defined as

$$\mathcal{Y}(m, N, I) \triangleq \left\{ \mathbf{f}: \mathfrak{R}^N \rightarrow \mathfrak{R}^m: \mathbf{f}\left(\mathbf{z} + \sum_{i \notin I} \lambda_i \mathbf{e}^i\right) = \mathbf{f}(\mathbf{z}), \forall \boldsymbol{\lambda} \in \mathfrak{R}^N \right\}, \quad (1)$$

and $\mathbf{y}^k \in \mathcal{Y}(m_k, N, I_k), \forall k \in [K]$. For example, if $I = \{1, 2\}$ and $\mathbf{y} \in \mathcal{Y}(m, N, I)$, then \mathbf{y} only depends on the first two components of the primitive uncertainty vector $\tilde{\mathbf{z}}$. From a practical perspective, the specific structure of $\{I_k\}_{k=1}^K$ often translates naturally into practical modeling phenomena. For example, the condition $I_1 \subseteq I_2 \subseteq \dots \subseteq I_K$ reflects the nonanticipativity requirement in information-dependent modeling (of which multistage problems are a special case) where we have successive revelation of information at each stage. We consider the ambiguity-averse minimization of a linear expected cost, with a finite set of M linear expectation constraints. The general problem can be expressed as

$$\begin{aligned} Z_{GEN}^* &= \min_{\mathbf{x}, \{\mathbf{y}^k(\cdot)\}_{k=1}^K} \mathbf{c}^{0'} \mathbf{x} + \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(\sum_{k=1}^K \mathbf{d}^{0, k'} \mathbf{y}^k(\tilde{\mathbf{z}}) \right) \\ \text{s.t. } & \mathbf{c}^{l'} \mathbf{x} + \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(\sum_{k=1}^K \mathbf{d}^{l, k'} \mathbf{y}^k(\tilde{\mathbf{z}}) \right) \leq b_l \\ & \forall l \in [M], \\ & T(\tilde{\mathbf{z}}) \mathbf{x} + \sum_{k=1}^K \mathbf{U}^k \mathbf{y}^k(\tilde{\mathbf{z}}) = \mathbf{v}(\tilde{\mathbf{z}}), \\ & \underline{\mathbf{y}}^k \leq \mathbf{y}^k(\tilde{\mathbf{z}}) \leq \bar{\mathbf{y}}^k \quad \forall k \in [K], \\ & \mathbf{x} \geq \mathbf{0}, \\ & \mathbf{y}^k \in \mathcal{Y}(m_k, N, I_k) \quad \forall k \in [K], \end{aligned} \quad (2)$$

where the model data $(\mathbf{c}^l, \mathbf{d}^l, \mathbf{U}^k, \underline{\mathbf{y}}^k, \bar{\mathbf{y}}^k, I_k)$ is deterministic and we assume that the model data $\mathbf{T}(\tilde{\mathbf{z}}), \mathbf{v}(\tilde{\mathbf{z}})$ are affinely dependent on $\tilde{\mathbf{z}}$, given by $\mathbf{T}(\tilde{\mathbf{z}}) = \mathbf{T}^0 + \sum_{j=1}^N \tilde{z}_j \mathbf{T}^j$ and $\mathbf{v}(\tilde{\mathbf{z}}) = \mathbf{v}^0 + \sum_{j=1}^N \tilde{z}_j \mathbf{v}^j$, and are also given by deterministic parameters $\{\mathbf{T}^j\}_{j=0}^N$ and $\{\mathbf{v}^j\}_{j=0}^N$. In our model, the matrices \mathbf{U}^k are unaffected by the uncertainties, corresponding to the case of *fixed recourse* in the stochastic programming literature. Uncertainty in the values of \mathbf{c}^l or b_l can be handled by reformulating the problem and adding slack variables; however, our model does not handle the cases when $\mathbf{U}^k, \mathbf{d}^{l, k}$ are uncertain. In such cases, Ben-Tal et al. (2004, §4) showed that even using LDRs for recourse decisions can result in intractability.

The bounds on the recourse variables, $\underline{\mathbf{y}}^k$ and $\bar{\mathbf{y}}^k$, for each k , are specified constants which can be infinite. Explicitly, $\underline{\mathbf{y}}^k \in (\mathfrak{R} \cup \{-\infty\})^{m_k}$ and $\bar{\mathbf{y}}^k \in (\mathfrak{R} \cup \{+\infty\})^{m_k}$. For ease of exposition later, we will denote the index sets for the noninfinite bounds as follows:

$$\begin{aligned} \underline{J}^k &= \{i \in [m_k]: \underline{y}_i^k > -\infty\}, \\ \bar{J}^k &= \{i \in [m_k]: \bar{y}_i^k < +\infty\}. \end{aligned} \quad (3)$$

For brevity, we adopt the convention here and throughout this paper that (in)equalities involving recourse variables hold almost surely for all probability distributions \mathbb{P} in the family \mathbb{F} , i.e., $\mathbf{y}(\tilde{\mathbf{z}}) \leq \mathbf{u} \Leftrightarrow \mathbb{P}(\mathbf{y}(\tilde{\mathbf{z}}) \leq \mathbf{u}) = 1, \forall \mathbb{P} \in \mathbb{F}$.

2.2. Motivation

The general model (2) that we consider has a linear structure, which may appear overly restrictive. In this section, however, we will proceed to show how (2) can be used to model important classes of problems with piecewise-linear structures.

2.2.1. Piecewise-Linear Utility Functions. In the modeling of certain problems, such as newsvendor-type models, it is common to encounter constraints of the form

$\sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}}(\mathbf{y}(\tilde{\mathbf{z}})^+) \leq \mathbf{b}$, which can be modeled in the form of (2) using a slack decision rule $\mathbf{s}(\tilde{\mathbf{z}})$ as follows:

$$\begin{aligned} \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}}(\mathbf{s}(\tilde{\mathbf{z}})) &\leq \mathbf{b}, \\ \mathbf{s}(\tilde{\mathbf{z}}) &\geq \mathbf{0}, \\ \mathbf{s}(\tilde{\mathbf{z}}) &\geq \mathbf{y}(\tilde{\mathbf{z}}), \\ \mathbf{s}, \mathbf{y} &\in \mathcal{Y}(m, N, I). \end{aligned} \quad (4)$$

2.2.2. CVaR Constraints. The Conditional Value-at-Risk (CVaR) risk metric was popularized by Rockafellar and Uryasev (2000) and is the smallest law-invariant convex risk measure which is continuous from above that dominates Value-at-Risk (VaR) (Föllmer and Schied 2002, Theorem 4.61). CVaR is a coherent measure of risk (as axiomatized by Artzner et al. 1999) and is typically parameterized by a level $\beta \in (0, 1)$. Furthermore, the β -CVaR can be derived as a special case of the negative optimized certainty equivalent (OCE) introduced by Ben-Tal and Teboulle (1986, 2007). The worst-case β -CVaR, when the actual uncertainty distribution \mathbb{P} lies in a family of distributions \mathbb{F} , can be expressed as

$$\beta\text{-CVaR}_{\mathbb{F}}(\tilde{x}) \triangleq \inf_{v \in \mathbb{R}} \left\{ v + \frac{1}{1 - \beta} \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}}((\tilde{x} - v)^+) \right\}. \quad (5)$$

The β -CVaR can be used to approximate chance constraints by using the relation $\beta\text{-CVaR}_{\mathbb{F}}(\tilde{x}) \leq b \Rightarrow \mathbb{P}(\tilde{x} \geq b) \leq 1 - \beta$, which holds for any distribution \mathbb{P} in the family of distributions \mathbb{F} . Using a similar argument as (4), the constraint $\beta\text{-CVaR}_{\mathbb{F}}(\mathbf{y}(\tilde{\mathbf{z}})) \leq b$, for a scalar-valued decision rule $\mathbf{y}(\tilde{\mathbf{z}})$, can therefore be expressed as

$$\begin{aligned} v + \frac{1}{1 - \beta} \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}}(\mathbf{s}(\tilde{\mathbf{z}})) &\leq b, \\ \mathbf{s}(\tilde{\mathbf{z}}) &\geq \mathbf{0}, \\ \mathbf{s}(\tilde{\mathbf{z}}) &\geq \mathbf{y}(\tilde{\mathbf{z}}) - v, \\ \mathbf{s}, \mathbf{y} &\in \mathcal{Y}(1, N, I). \end{aligned} \quad (6)$$

3. Model of Uncertainty, \mathbf{U}

In the modeling of most problems, even though the problem data contains elements of uncertainty, the modeler may have access to some crude or partial information about the data. We assume that we may have knowledge of certain descriptive statistics of the primitive uncertainty vector $\tilde{\mathbf{z}}$, as follows.

Support. We denote by $\mathcal{W} \subseteq \mathbb{R}^N$ the smallest convex set containing the support of $\tilde{\mathbf{z}}$, which can also be unbounded. For example, if the actual support of $\tilde{\mathbf{z}}$ is nonconvex, we can take \mathcal{W} as its convex hull. We further assume that \mathcal{W} is a full-dimensional tractable conic representable set, which we take to mean a set that can be represented (exactly or approximately) by a polynomial number of linear and/or second order conic constraints.¹

Mean. We denote by $\hat{\mathbf{z}}$ the mean of $\tilde{\mathbf{z}}$. Instead of modeling the mean as a precisely known quantity, we consider a generalization in which the mean $\hat{\mathbf{z}}$ is itself uncertain, with corresponding (possibly unbounded) support contained in a set \mathcal{W} . We again assume that \mathcal{W} is a tractable conic representable set. This includes the case of a known mean, which corresponds to \mathcal{W} being a singleton set.

Covariance. We denote by Σ the covariance of $\tilde{\mathbf{z}}$. Unlike the mean, which we assume to be known to lie within a set, we assume that the covariance is precisely known.²

Directional Deviations. While $\tilde{\mathbf{z}}$ may not have stochastically independent components, we may be able to find a linear transformation of $\tilde{\mathbf{z}}$, parameterized by a matrix $\mathbf{H}_{\sigma} \in \mathbb{R}^{N_{\sigma} \times N}$, generally with $N_{\sigma} \leq N$, that yields a vector $\mathbf{H}_{\sigma} \tilde{\mathbf{z}} = \tilde{\mathbf{z}}_{\sigma}$, which has stochastically independent components. We denote by $\hat{\mathbf{z}}_{\sigma}$ the mean of $\tilde{\mathbf{z}}_{\sigma}$, which lies in a set $\mathcal{W}_{\sigma} \triangleq \{\mathbf{H}_{\sigma} \hat{\mathbf{z}} : \hat{\mathbf{z}} \in \mathcal{W}\}$. We denote by $\sigma_{\mathbf{f}}$ and $\sigma_{\mathbf{b}}$ the upper bounds of the forward and backward deviations of $\tilde{\mathbf{z}}_{\sigma}$, i.e., $\sigma_{\mathbf{f}} \geq \sigma_{\mathbf{f}\mathbb{P}}(\tilde{\mathbf{z}}_{\sigma})$ and $\sigma_{\mathbf{b}} \geq \sigma_{\mathbf{b}\mathbb{P}}(\tilde{\mathbf{z}}_{\sigma})$, where $\sigma_{\mathbf{f}, \mathbb{P}}(\cdot)$, $\sigma_{\mathbf{b}, \mathbb{P}}(\cdot)$ are defined componentwise by Chen et al. (2007) as

$$\begin{aligned} \sigma_{\mathbf{f}\mathbb{P}}(\tilde{\mathbf{z}}_{\sigma}) \mathbf{e}^j &= \sigma_{\mathbf{f}\mathbb{P}}(\tilde{\mathbf{z}}_{\sigma, j}) \\ &\triangleq \sup_{\theta > 0} \left\{ \sqrt{2 \ln \mathbb{E}_{\mathbb{P}}(\exp(\theta(\tilde{\mathbf{z}}_{\sigma, j} - \hat{\mathbf{z}}_{\sigma, j})))/\theta^2} \right\}, \\ \sigma_{\mathbf{b}\mathbb{P}}(\tilde{\mathbf{z}}_{\sigma}) \mathbf{e}^j &= \sigma_{\mathbf{b}\mathbb{P}}(\tilde{\mathbf{z}}_{\sigma, j}) \\ &\triangleq \sup_{\theta > 0} \left\{ \sqrt{2 \ln \mathbb{E}_{\mathbb{P}}(\exp(-\theta(\tilde{\mathbf{z}}_{\sigma, j} - \hat{\mathbf{z}}_{\sigma, j})))/\theta^2} \right\}, \end{aligned}$$

for $j \in [N_{\sigma}]$. We consider upper bounds of directional deviations in order to characterize a family of distributions. We note that numerical values of these bounds can be estimated from empirical data, and we refer interested readers to Chen et al. (2007), Natarajan et al. (2008), or See and Sim (2009) for examples of how directional deviations can be estimated and used.

These distributional properties characterize the family of distributions \mathbb{F} . We will see in §5 how each property can be used to construct bounds to approximately solve the general problem (2), which will, in turn, be used in §6 where we introduce the deflected linear decision rules.

4. Linear Approximations of the General Model

Solving model (2) exactly is generally a computationally intractable endeavor. For instance, when the family \mathbb{F} contains a single distribution, Dyer and Stougie (2006) formally showed that a two-stage problem is $\#P$ -hard to solve. In the robust case, i.e., when \mathbb{F} is solely defined by a support set, a two-stage problem can be NP-hard (see the Adjustable Robust Counterpart of Ben-Tal et al. 2004). However, by applying a suitable restriction to the space of allowable decision rules, we can obtain a tractable approximation to the problem. Instead of considering all possible

choices of $\mathbf{y}^k(\cdot)$ from $\mathcal{Y}(m, N, I)$, we restrict ourselves to Linear Decision Rules (LDRs), where each $\mathbf{y}^k(\cdot)$ is instead chosen from the space of *affine* functions of $\tilde{\mathbf{z}}$, denoted by $\mathcal{L}(m, N, I) \subset \mathcal{Y}(m, N, I)$, and defined as follows:

$$\mathcal{L}(m, N, I) \triangleq \left\{ \mathbf{f}: \mathfrak{R}^N \rightarrow \mathfrak{R}^m: \exists (\mathbf{y}^0, \mathbf{Y}) \in \mathfrak{R}^m \times \mathfrak{R}^{m \times N}: \right. \\ \left. \begin{aligned} \mathbf{f}(\mathbf{z}) &= \mathbf{y}^0 + \mathbf{Y}\mathbf{z} \\ \mathbf{Y}\mathbf{e}^i &= \mathbf{0}, \forall i \notin I \end{aligned} \right\}. \quad (7)$$

We notice that the final condition, $\mathbf{Y}\mathbf{e}^i = \mathbf{0}$ enforces the information dependency upon the index set I . Therefore, using a linear model of recourse, the recourse decision can be written explicitly as $\mathbf{y}^k(\tilde{\mathbf{z}}) = \mathbf{y}^{0,k} + \mathbf{Y}^k\tilde{\mathbf{z}}$, for each $k \in [K]$.

Using LDRs as our model of recourse and denoting by \mathbb{F} the family of distributions \mathbb{P} with distributional properties as specified in the Model of Uncertainty, problem (2) is approximated as

$$\begin{aligned} & Z_{\text{LDR}} \\ &= \min_{\mathbf{x}, \{\mathbf{y}^{0,k}, \mathbf{Y}^k\}_{k=1}^K} \mathbf{c}^{0'}\mathbf{x} + \sum_{k=1}^K \mathbf{d}^{0,k'}\mathbf{y}^{0,k} + \sup_{\hat{\mathbf{z}} \in \hat{\mathcal{W}}} \left(\sum_{k=1}^K \mathbf{d}^{0,k'}\mathbf{Y}^k\hat{\mathbf{z}} \right) \\ & \text{s.t. } \mathbf{c}^{l'}\mathbf{x} + \sum_{k=1}^K \mathbf{d}^{l,k'}\mathbf{y}^{0,k} \\ & \quad + \sup_{\hat{\mathbf{z}} \in \hat{\mathcal{W}}} \left(\sum_{k=1}^K \mathbf{d}^{l,k'}\mathbf{Y}^k\hat{\mathbf{z}} \right) \leq b_l \quad \forall l \in [M], \\ & \mathbf{T}^0\mathbf{x} + \sum_{k=1}^K \mathbf{U}^k\mathbf{y}^{0,k} = \mathbf{v}^0, \\ & \mathbf{T}^j\mathbf{x} + \sum_{k=1}^K \mathbf{U}^k\mathbf{Y}^k\mathbf{e}^j = \mathbf{v}^j \quad \forall j \in [N], \\ & \underline{\mathbf{y}}^k \leq \mathbf{y}^{0,k} + \mathbf{Y}^k\mathbf{z} \leq \bar{\mathbf{y}}^k \quad \forall \mathbf{z} \in \mathcal{W}, \forall k \in [K], \\ & \mathbf{Y}^k\mathbf{e}^j = \mathbf{0} \quad \forall j \notin I_k, \forall k \in [K], \\ & \mathbf{x} \geq \mathbf{0}. \end{aligned} \quad (8)$$

We formalize this in the following proposition:

PROPOSITION 1. *If $\mathcal{L}(m_k, N, I_k)$ is used to approximate $\mathcal{Y}(m_k, N, I_k)$, then under the approximation, Problem (2) is equivalent to Problem (8).*

PROOF. Please see Appendix A.1 of the electronic companion, which is available as part of the online version that can be found at <http://or.journal.informs.org/>.

In the transformed problem, we notice that the constraints $\underline{\mathbf{y}}^k \leq \mathbf{y}^{0,k} + \mathbf{Y}^k\mathbf{z} \leq \bar{\mathbf{y}}^k$ over all $\mathbf{z} \in \mathcal{W}$ as well as the suprema over $\hat{\mathbf{z}} \in \hat{\mathcal{W}}$ in the objective and first M constraints can be converted into their robust counterparts. This will render Problem (8) as a tractable conic optimization problem, since \mathcal{W} and $\hat{\mathcal{W}}$ are tractable conic representable sets (see Ben-Tal and Nemirovski 1998, Bertsimas and Sim 2004). In particular, if \mathcal{W} and $\hat{\mathcal{W}}$ are polyhedral, then Problem (8) becomes a linear program.

While the LDR approximation is computationally tractable, the quality of the approximation can be very poor, even being infeasible for very simple constraints (see, e.g., the discussion by Chen et al. 2008). In the remainder of this section, we discuss how we may segregate the primitive uncertainty vector to obtain a more flexible model of recourse. In §6, we further discuss how to exploit the structure of the constraints to construct even more flexible piecewise-linear decision rules.

4.1. Remapping the Primitive Uncertainty Vector

In the positive-and-negative Segregated LDR introduced by Chen et al. (2008), the authors split the original uncertainty vector into positive and negative half spaces and applied LDRs to the split uncertainties in order to increase the flexibility of the LDR. We extend their result and discuss here how we may segregate primitive uncertainties into intervals and, more importantly, how we can use the segregated uncertainties in our modeling framework. To begin, we consider a functional mapping $\mathbf{M}: \mathfrak{R}^N \rightarrow \mathfrak{R}^{N_E}$, where $N_E \geq N$, that satisfies the following relationship for any $\mathbf{z} \in \mathfrak{R}^N$, for some given matrix $\mathbf{F} \in \mathfrak{R}^{N \times N_E}$ and vector $\mathbf{g} \in \mathfrak{R}^{N_E}$, such that

$$\mathbf{z} = \mathbf{F}\mathbf{M}(\mathbf{z}) + \mathbf{g}. \quad (9)$$

We notice that \mathbf{F} has to be full rank since \mathcal{W} is assumed to have a nonempty interior. We denote by \mathcal{V}^* the image of $\mathbf{M}(\cdot)$ corresponding to the domain \mathcal{W} , i.e.,

$$\mathcal{V}^* \triangleq \{\mathbf{M}(\mathbf{z}): \mathbf{z} \in \mathcal{W}\}, \quad (10)$$

and similarly by $\hat{\mathcal{V}}^*$ the image of $\mathbf{M}(\cdot)$ corresponding to the domain $\hat{\mathcal{W}}$, i.e., $\hat{\mathcal{V}}^* \triangleq \{\mathbf{M}(\hat{\mathbf{z}}): \hat{\mathbf{z}} \in \hat{\mathcal{W}}\}$.

Again, we require $\mathbf{M}(\cdot)$ to be such that the convex hull of \mathcal{V}^* is full dimensional. We notice that while $\mathbf{M}(\cdot)$ is invertible by an affine mapping, $\mathbf{M}(\cdot)$ itself is not required to be affine in its argument. Indeed, in the example which follows, we present a piecewise-affine $\mathbf{M}(\cdot)$, and show how it can be used to segregate a scalar primitive uncertainty into different regions of interest.

4.2. Example: Segregating a Scalar Uncertainty

Suppose we have a scalar primitive uncertainty, \tilde{z} with support \mathfrak{R} , that we wish to segregate into three regions, $(-\infty, -1]$, $[-1, 1]$, and $[1, +\infty)$. Denoting the points, $(p_1, p_2, p_3, p_4) = (-\infty, -1, 1, +\infty)$, we can construct the segregated uncertainty by applying the following nonlinear mapping: $\tilde{\zeta} = \mathbf{M}(\tilde{z})$, where

$$\tilde{\zeta}_i = \begin{cases} \tilde{z} & \text{if } p_i \leq \tilde{z} \leq p_{i+1}, \\ p_i & \text{if } \tilde{z} \leq p_i, \\ p_{i+1} & \text{if } \tilde{z} \geq p_{i+1}, \end{cases}$$

for $i \in \{1, 2, 3\}$. We notice that

$$\tilde{\zeta}_1 + \tilde{\zeta}_2 + \tilde{\zeta}_3 = \begin{cases} \tilde{z} - 1 + 1 = \tilde{z} & \text{if } \tilde{z} \leq -1, \\ -1 + \tilde{z} + 1 = \tilde{z} & \text{if } -1 < \tilde{z} < 1, \\ -1 + 1 + \tilde{z} = \tilde{z} & \text{if } \tilde{z} \geq 1, \end{cases}$$

and therefore Equation (9) holds with $\mathbf{F} = [1, 1, 1]$ and $\mathbf{g} = 0$.

4.3. Mappings That Represent Segregations

In the preceding example, we notice that the segregation resulted in a new uncertainty vector $\tilde{\zeta}$, the components of which provide local information of the original scalar uncertainty \tilde{z} . In general, the purpose of segregating uncertainties into intervals is to obtain a finer resolution of the original uncertainty. In §4.4, we will use segregated uncertainties to define more flexible decision rules. To better understand how we can construct such segregations, we proceed to characterize the mapping functions $\mathbf{M}(\cdot)$ which represents a segregation of a primitive uncertainty vector.

For some positive integer L , we consider a collection of $N \times (L + 1)$ distinct points on the extended real line, denoted by ξ_{ij} for some $i \in [N], j \in [L + 1]$, with the following properties $\forall i \in [N]$:

$$\begin{aligned} \xi_{i,1} &= -\infty, \\ \xi_{i,L+1} &= +\infty, \\ \xi_{i,j_1} &< \xi_{i,j_2} \quad \text{iff } j_1 < j_2. \end{aligned} \quad (11)$$

Furthermore, we denote by ξ the $N \times (L + 1)$ matrix which collects these points. We call $\mathbf{M}(\cdot)$ a segregation if $\forall \zeta = \mathbf{M}(\mathbf{z})$, its components $\forall j \in [N_E]$, are given by

$$\zeta_j = \begin{cases} z_i & \text{if } \xi_{i,k} \leq z_i \leq \xi_{i,k+1}, \\ \xi_{i,k} & \text{if } z_i \leq \xi_{i,k}, \\ \xi_{i,k+1} & \text{if } z_i \geq \xi_{i,k+1}, \end{cases} \quad (12)$$

where $j = i + (k - 1)N$ for some $i \in [N], k \in [L]$. We notice that for a particular $j \in [N_E]$, i and k can be uniquely obtained as

$$i = ((j - 1) \bmod N) + 1,$$

$$k = \lceil j/N \rceil.$$

In the following proposition, we prove that if $\mathbf{M}(\cdot)$ is a segregation, it is also affinely invertible by proper choice of \mathbf{F} and \mathbf{g} .

PROPOSITION 2. *If $\mathbf{M}(\cdot)$ is a segregation and if we choose \mathbf{F} and \mathbf{g} to be*

$$\mathbf{F} = [\mathbf{I} \quad \mathbf{I} \quad \cdots \quad \mathbf{I}], \quad \mathbf{g} = -\sum_{i=2}^L \xi \mathbf{e}^i,$$

then $\mathbf{z} = \mathbf{F}\mathbf{M}(\mathbf{z}) + \mathbf{g} \quad \forall \mathbf{z} \in \mathfrak{R}^N$, where $\mathbf{F} \in \mathfrak{R}^{N \times LN}$, $\mathbf{g} \in \mathfrak{R}^N$, and \mathbf{I} is the $N \times N$ identity matrix.

PROOF. Please see Appendix A.2 in the electronic companion.

REMARK. Although we consider a segregation $\mathbf{M}(\cdot)$ which segments each component of $\tilde{\mathbf{z}}$ uniformly into L parts, it is clear from the proof that Proposition 2 still holds even if each component of $\tilde{\mathbf{z}}$ was segmented into a different number of positive integer parts, $\{L_i\}_{i=1}^N$, albeit with a different choice of \mathbf{F} and \mathbf{g} . This would come at the expense of more notation and bookkeeping. Hence, for simplicity, here and for the rest of this paper, we will discuss only the case of uniform segregation.

4.4. Segregated Linear Decision Rules

As seen above, the segregation $\mathbf{M}(\cdot)$ can be used to define a new uncertainty vector $\tilde{\zeta} \in \mathfrak{R}^{N_E}$, which we will term the segregated uncertainty vector. By considering LDRs on the segregated uncertainty vector $\tilde{\zeta}$, we obtain a new set of decision rules, which we term segregated LDRs or SLDRs for short. By using the SLDRs, the recourse decisions become

$$\mathbf{y}^k(\tilde{\mathbf{z}}) = \mathbf{r}^k(\mathbf{M}(\tilde{\mathbf{z}})) = \mathbf{r}^{0,k} + \mathbf{R}^k \mathbf{M}(\tilde{\mathbf{z}}), \quad \forall k \in [K],$$

which is effectively the composition of an affine functional with the segregating mapping $\mathbf{M}(\cdot)$. Under these SLDRs, Problem (2) becomes

$$\begin{aligned} Z_{SLDR}^* &= \min_{\mathbf{x}, \{\mathbf{r}^k(\cdot)\}_{k=1}^K} \mathbf{c}'\mathbf{x} + \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}} \left(\sum_{k=1}^K \mathbf{d}^{0,k'} \mathbf{r}^k(\mathbf{M}(\tilde{\mathbf{z}})) \right) \\ \text{s.t.} \quad &\mathbf{c}'\mathbf{x} + \sup_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}} \left(\sum_{k=1}^K \mathbf{d}^{l,k'} \mathbf{r}^k(\mathbf{M}(\tilde{\mathbf{z}})) \right) \leq b_l \\ &\quad \forall l \in [M], \\ &\mathbf{T}(\tilde{\mathbf{z}})\mathbf{x} + \sum_{k=1}^K \mathbf{U}^k \mathbf{r}^k(\mathbf{M}(\tilde{\mathbf{z}})) = \mathbf{v}(\tilde{\mathbf{z}}), \quad (13) \\ &\underline{\mathbf{y}}^k \leq \mathbf{r}^k(\mathbf{M}(\tilde{\mathbf{z}})) \leq \bar{\mathbf{y}}^k \quad \forall k \in [K], \\ &\mathbf{x} \geq \mathbf{0}, \\ &\mathbf{r}^k(\mathbf{M}(\tilde{\mathbf{z}})) = \mathbf{r}^{0,k} + \mathbf{R}^k \mathbf{M}(\tilde{\mathbf{z}}) \quad \forall k \in [K], \\ &\mathbf{r}^k \circ \mathbf{M} \in \mathcal{Y}(m_k, N, I_k) \quad \forall k \in [K]. \end{aligned}$$

Problem (A.4) of Appendix A in the electronic companion presents an equivalent, and more explicit formulation of Problem (13). We notice that Problem (A.4) and, equivalently (13), is generally intractable, since \mathcal{V}^* and $\hat{\mathcal{V}}^*$ are generally nonconvex. Furthermore, it is not obvious how to handle the non-anticipativity constraints. We therefore aim to construct an approximation of Problem (13) that would still improve upon the standard LDR.

4.5. Approximating \mathcal{V}^* and $\hat{\mathcal{V}}^*$

We begin by constructing sets $\mathcal{V}' \supseteq \mathcal{V}^*$ and $\hat{\mathcal{V}}' \supseteq \hat{\mathcal{V}}^*$ which approximate the sets \mathcal{V}^* and $\hat{\mathcal{V}}^*$. Both \mathcal{V}' and $\hat{\mathcal{V}}'$ should

be tractable conic representable sets and should satisfy the implicit set relations

$$\begin{aligned} \mathcal{W} &= \{\mathbf{F}\boldsymbol{\zeta} + \mathbf{g} : \boldsymbol{\zeta} \in \mathcal{V}\}, \\ \widehat{\mathcal{W}} &= \{\mathbf{F}\widehat{\boldsymbol{\zeta}} + \mathbf{g} : \widehat{\boldsymbol{\zeta}} \in \widehat{\mathcal{V}}\}. \end{aligned} \tag{14}$$

To begin explicit construction of \mathcal{V} , we note that the best convex approximant of \mathcal{V}^* is its convex hull, $\text{conv}(\mathcal{V}^*)$ and, ideally, we would choose $\mathcal{V} = \text{conv}(\mathcal{V}^*)$. Indeed, Wang et al. (2008) show how to describe $\text{conv}(\mathcal{V}^*)$ when \mathcal{W} has a special structure, which they term an *absolute set*. However, in general, for \mathcal{W} of arbitrary structures, it may not be easy to describe $\text{conv}(\mathcal{V}^*)$. Instead, we construct \mathcal{V} using a proxy set \mathcal{H} by defining

$$\mathcal{V} \triangleq \{\boldsymbol{\zeta} \in \mathcal{H} : \mathbf{F}\boldsymbol{\zeta} + \mathbf{g} \in \mathcal{W}\}, \tag{15}$$

where \mathcal{H} is explicitly defined as

$$\begin{aligned} \mathcal{H} \triangleq \{\boldsymbol{\zeta} \in \mathfrak{R}^{N_E} : \xi_{i,k} \leq \zeta_j \leq \xi_{i,k+1} \ \forall j \in [N_E] : \\ i \in [N], k \in [L] : j = i + (k - 1)N\}. \end{aligned} \tag{16}$$

We observe that from the definition of a segregation (12), \mathcal{H} satisfies the property that $\mathbf{z} \in \mathcal{W} \Rightarrow \mathbf{M}(\mathbf{z}) \in \mathcal{H}$ and, clearly, \mathcal{V} defined in this manner will satisfy (14). The motivation for this seemingly extraneous construction of \mathcal{V} is that \mathcal{H} depends only on the segregation $\mathbf{M}(\cdot)$ and is decoupled from \mathcal{W} , making it easy to specify in practice. In particular, for the example presented in §4.2, \mathcal{H} can be represented by $\mathcal{H} = \{\boldsymbol{\zeta} \in \mathfrak{R}^3 : \zeta_1 \leq -1, -1 \leq \zeta_2 \leq 1, \zeta_3 \geq 1\}$. Using a similar argument, we define $\widehat{\mathcal{V}}$ by the same set \mathcal{H} as $\widehat{\mathcal{V}} \triangleq \{\widehat{\boldsymbol{\zeta}} \in \mathcal{H} : \mathbf{F}\widehat{\boldsymbol{\zeta}} + \mathbf{g} \in \widehat{\mathcal{W}}\}$.

4.6. Approximating Problem (13)

After constructing \mathcal{V} and $\widehat{\mathcal{V}}$, we now discuss how we can approximate Problem (13). For convenience, we define the collection of index sets for each $k \in [K]$ as

$$\Phi_k = \{j \in [N_E] : \exists i \in I_k : (i - 1) \equiv (j - 1) \pmod{N}\}. \tag{17}$$

We then define the tractable SLDR approximation to Problem (13) as the following:

$$\begin{aligned} Z_{SLDR} &= \min_{\mathbf{x}, \{\mathbf{r}^{0,k}, \mathbf{R}^k\}_{k=1}^K} \mathbf{c}^0 \mathbf{x} + \sum_{k=1}^K \mathbf{d}^{0,k'} \mathbf{r}^{0,k} + \sup_{\widehat{\boldsymbol{\zeta}} \in \widehat{\mathcal{V}}} \left(\sum_{k=1}^K \mathbf{d}^{0,k'} \mathbf{R}^k \widehat{\boldsymbol{\zeta}} \right) \\ \text{s.t.} \quad &\mathbf{c}^l \mathbf{x} + \sum_{k=1}^K \mathbf{d}^{l,k'} \mathbf{r}^{0,k} + \sup_{\widehat{\boldsymbol{\zeta}} \in \widehat{\mathcal{V}}} \left(\sum_{k=1}^K \mathbf{d}^{l,k'} \mathbf{R}^k \widehat{\boldsymbol{\zeta}} \right) \leq b_l \\ &\forall l \in [M], \\ &\mathcal{J}^0 \mathbf{x} + \sum_{k=1}^K \mathbf{U}^k \mathbf{r}^{0,k} = \mathbf{v}^0, \\ &\mathcal{J}^j \mathbf{x} + \sum_{k=1}^K \mathbf{U}^k \mathbf{R}^k \mathbf{e}^j = \mathbf{v}^j \quad \forall j \in [N_E], \\ &\mathbf{y}^k \leq \mathbf{r}^{0,k} + \mathbf{R}^k \boldsymbol{\zeta} \leq \bar{\mathbf{y}}^k \quad \forall \boldsymbol{\zeta} \in \mathcal{V} \ \forall k \in [K], \\ &\mathbf{R}^k \mathbf{e}^j = \mathbf{0} \quad \forall j \notin \Phi_k, \forall k \in [K], \\ &\mathbf{x} \geq \mathbf{0}, \end{aligned} \tag{18}$$

where the transformed model data is defined as

$$\begin{aligned} \mathbf{v}^0 &= \mathbf{v}^0 + \sum_{i=1}^N g_i \mathbf{v}^i, & \mathcal{J}^0 &= \mathbf{T}^0 + \sum_{i=1}^N g_i \mathbf{T}^i, \\ \mathbf{v}^j &= \sum_{i=1}^N F_{ij} \mathbf{v}^i, & \mathcal{J}^j &= \sum_{i=1}^N F_{ij} \mathbf{T}^i \quad \forall j \in [N_E]. \end{aligned} \tag{19}$$

The following proposition relates the objectives under the exact (intractable) SLDR, the approximate SLDR, and the LDR models of recourse.

PROPOSITION 3. *The following inequality holds: $Z_{SLDR}^* \leq Z_{SLDR} \leq Z_{LDR}$.*

PROOF. Please see Appendix A.3 in the electronic companion.

REMARK 1. Proposition 3 shows that irrespective of how crudely ($\mathcal{V}, \widehat{\mathcal{V}}$) approximates ($\mathcal{V}^*, \widehat{\mathcal{V}}^*$), using the approximate SLDR will nonetheless not be worse than using the original LDR. Furthermore, using the SLDR retains the linear structure of the problem. Specifically, if \mathcal{V} and $\widehat{\mathcal{V}}$ are polyhedral, the SLDR approximation (Problem (18)) reduces to a linear program.

REMARK 2. A key difference in the SLDR that we describe here and the SLDR of Chen et al. (2008) is that they assume precise knowledge of the mean and covariance of the segregated uncertainty vector. In our tractable SLDR model, we only exploit the support information of the segregated uncertainty vector, which is captured by the set \mathcal{H} . Our SLDR however, does include their model as a special case, since, if we did have knowledge of the segregated moments, we could simply reformulate the problem, expressing what they term as the segregated uncertainty vector as our primitive uncertainty vector.

4.7. Interpreting the SLDR Approximation (18)

We notice that the structure of the SLDR approximation (18) above closely resembles the form of the LDR approximation (8). Indeed, we can interpret the SLDR as a linear approximation of the following uncertain optimization problem, defined over a different uncertainty vector $\tilde{\boldsymbol{\zeta}}$:

$$\begin{aligned} Z_{GEN,2}^* &= \min_{\mathbf{x}, \{\mathbf{r}^k(\cdot)\}_{k=1}^K} \mathbf{c}^0 \mathbf{x} + \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(\sum_{k=1}^K \mathbf{d}^{0,k'} \mathbf{r}^k(\tilde{\boldsymbol{\zeta}}) \right) \\ \text{s.t.} \quad &\mathbf{c}^l \mathbf{x} + \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(\sum_{k=1}^K \mathbf{d}^{l,k'} \mathbf{r}^k(\tilde{\boldsymbol{\zeta}}) \right) \leq b_l \\ &\forall l \in [M], \\ &\mathcal{J}(\tilde{\boldsymbol{\zeta}}) \mathbf{x} + \sum_{k=1}^K \mathbf{U}^k \mathbf{r}^k(\tilde{\boldsymbol{\zeta}}) = \mathbf{v}(\tilde{\boldsymbol{\zeta}}), \\ &\mathbf{y}^k \leq \mathbf{r}^k(\tilde{\boldsymbol{\zeta}}) \leq \bar{\mathbf{y}}^k \quad \forall k \in [K], \\ &\mathbf{x} \geq \mathbf{0}, \\ &\mathbf{r}^k \in \mathcal{Y}(m_k, N_E, \Phi_k) \quad \forall k \in [K], \end{aligned} \tag{20}$$

where

$$\mathcal{J}(\tilde{\zeta}) = \mathcal{J}^0 + \sum_{j=1}^N \tilde{\zeta}_j \mathcal{J}^j, \quad \mathbf{v}(\tilde{\zeta}) = \mathbf{v}^0 + \sum_{j=1}^N \tilde{\zeta}_j \mathbf{v}^j.$$

After approximating $\mathcal{Y}(m_k, N_E, \Phi_k)$ with $\mathcal{L}(m_k, N_E, \Phi_k)$ and following the same steps as the LDR approximation (see the online Appendix A.1), (20) will reduce to (18). When we interpret (18) as an LDR approximation of Problem (20) above, the approximate sets \mathcal{V} and $\hat{\mathcal{V}}$ should therefore be interpreted respectively as supersets of the support and mean support of the new uncertainty vector $\tilde{\zeta}$. Furthermore, the collection of index sets $\{\Phi_k\}_{k=1}^K$, which was somewhat arbitrarily defined before, now has the natural interpretation as the information index sets of the new decision rules, $\mathbf{r}^k(\cdot)$.

4.8. Example: Specifying Distributional Properties for Segregated Uncertainties

We provide a concrete example of how various distributional properties can be specified for segregated uncertainties to illustrate how segregations might work in practice. Consider a primitive uncertainty vector $\tilde{\mathbf{z}} \in \mathfrak{R}^4$, where only \tilde{z}_2 and \tilde{z}_3 are stochastically independent. The distributional properties \mathcal{W} , \mathcal{M} , and Σ of the primitive uncertainty vector $\tilde{\mathbf{z}}$ can be specified directly. Furthermore, we wish to segregate each component of $\tilde{\mathbf{z}}$ into three regions, $(-\infty, -1]$, $[-1, 1]$, and $[1, \infty)$ as in the earlier example of §4.2. This results in a segregated uncertainty vector $\tilde{\zeta} \in \mathfrak{R}^{12}$, which obeys the relation $\tilde{\mathbf{z}} = \mathbf{F}\tilde{\zeta} + \mathbf{g}$ for parameters $\mathbf{F} \in \mathfrak{R}^{4 \times 12}$, $\mathbf{g} \in \mathfrak{R}^4$, where

$$\mathbf{F} = [\mathbf{I} \quad \mathbf{I} \quad \mathbf{I}] \quad \text{and} \quad \mathbf{g} = \mathbf{0},$$

and \mathbf{I} represents the 4-by-4 identity matrix. Based on the segregation, we can choose \mathcal{H} as

$$\mathcal{H} = \left\{ \begin{array}{ll} \zeta_i \leq -1 & \text{if } i = 1, 2, 3, 4, \\ \zeta \in \mathfrak{R}^{12}: -1 \leq \zeta_i \leq 1 & \text{if } i = 5, 6, 7, 8, \\ \zeta_i \geq 1 & \text{if } i = 9, 10, 11, 12 \end{array} \right\}.$$

Finally, we notice that expressing $\tilde{\mathbf{z}}_\sigma = \mathbf{F}_\sigma \tilde{\zeta} + \mathbf{g}_\sigma$, where

$$\mathbf{F}_\sigma = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{F} \quad \text{and} \quad \mathbf{g}_\sigma = \mathbf{0},$$

we can construct an uncertainty vector $\tilde{\mathbf{z}}_\sigma$ with independent components. We can therefore specify the directional deviations (σ_f, σ_b) of $\tilde{\mathbf{z}}_\sigma$.

5. Distributionally Ambiguous Bounds for $E_{\mathbb{P}}((\cdot)^+)$

When we specify partial distributional information of the model uncertainties, we effectively characterize a family

of distributions \mathbb{F} , which contains the true uncertainties distribution \mathbb{P} . In this section, we discuss how we may evaluate the supremum of the expected positive part of an SLDR (recall that according to the discussion in §4.7, the SLDR can be interpreted as an LDR on the segregated uncertainties), i.e.,

$$\sup_{\mathbb{P} \in \mathbb{F}} E_{\mathbb{P}}((\mathbf{r}(\tilde{\zeta}))^+),$$

$$\mathbf{r} \in \mathcal{L}(m, N_E, \Phi),$$

where the family \mathbb{F} is partially characterized by the distributional properties as laid out in the Model of Uncertainty in §3. This is in anticipation of our discussion of deflected linear decision rules (DLDRs) in §6, where we will use these bounds. We will show how each pair of distributional properties

- Mean and (segregated) Support,
- Mean and Covariance,
- Mean and Directional Deviation

establishes a distinct bound, and we conclude this section by showing how to combine these separate bounds when we have access to a combination of distributional information from these three categories.

We show that each bound can be constructed by solving a deterministic optimization problem, and we adopt the standard convention in convex programming that an infeasible minimization (maximization) problem has an optimal value of $+\infty$ ($-\infty$). Also, we will only present results for $E_{\mathbb{P}}((\cdot)^+)$, since the results for the negative case can be derived easily by applying the identity $x^- = (-x)^+$. For generality, we will discuss the bound of $E_{\mathbb{P}}((r^0 + \mathbf{r}'\tilde{\zeta})^+)$, applied to the segregated LDR as defined in §4, since bounding the expectation of the positive part of a standard LDR, i.e., $E_{\mathbb{P}}((y^0 + \mathbf{y}'\tilde{\mathbf{z}})^+)$, can be derived as a special case when $\mathbf{M}(\mathbf{z}) = \mathbf{z}$.

5.1. Mean and Support Information

THEOREM 1. *Let \mathbb{F}_1 be the family of all distributions \mathbb{P} such that the random variable $\tilde{\zeta}$ has support in \mathcal{V} and its mean $\hat{\zeta}$ has support in $\hat{\mathcal{V}}$, i.e.,*

$$\mathbb{F}_1 = \{\mathbb{P}: \hat{\zeta} = E_{\mathbb{P}}(\tilde{\zeta}) \in \hat{\mathcal{V}}, \mathbb{P}(\tilde{\zeta} \in \mathcal{V}) = 1\}.$$

Then $\pi^1(r^0, \mathbf{r})$ is a tight upper bound for $E_{\mathbb{P}}((r^0 + \mathbf{r}'\tilde{\zeta})^+)$ over all distributions $\mathbb{P} \in \mathbb{F}_1$, i.e.,

$$\sup_{\mathbb{P} \in \mathbb{F}_1} E_{\mathbb{P}}((r^0 + \mathbf{r}'\tilde{\zeta})^+) = \pi^1(r^0, \mathbf{r}),$$

where

$$\pi^1(r^0, \mathbf{r}) \triangleq \inf_{\mathbf{s} \in \mathfrak{R}^{N_E}} (\sup_{\tilde{\zeta} \in \hat{\mathcal{V}}} \{\mathbf{s}'\tilde{\zeta}\} + \sup_{\tilde{\zeta} \in \mathcal{V}} (\max\{r^0 + \mathbf{r}'\tilde{\zeta} - \mathbf{s}'\tilde{\zeta}, -\mathbf{s}'\tilde{\zeta}\})). \quad (21)$$

PROOF. Please see the online Appendix B.1.

REMARK. First, we notice that for given (r^0, \mathbf{r}) ,

$$r^0 + \mathbf{r}'\boldsymbol{\zeta} \geq 0, \quad \forall \boldsymbol{\zeta} \in \mathcal{V}$$

$$\Rightarrow \sup_{\mathbb{P} \in \mathbb{F}_1} \mathbb{E}_{\mathbb{P}}((r^0 + \mathbf{r}'\tilde{\boldsymbol{\zeta}})^+) = r^0 + \sup_{\hat{\boldsymbol{\zeta}} \in \hat{\mathcal{V}}} \mathbf{r}'\hat{\boldsymbol{\zeta}},$$

which is attained by $\pi^1(r^0, \mathbf{r})$ by choosing $\mathbf{s} = \mathbf{r}$. Also,

$$r^0 + \mathbf{r}'\boldsymbol{\zeta} \leq 0, \quad \forall \boldsymbol{\zeta} \in \mathcal{V} \Rightarrow \sup_{\mathbb{P} \in \mathbb{F}_1} \mathbb{E}_{\mathbb{P}}((r^0 + \mathbf{r}'\tilde{\boldsymbol{\zeta}})^+) = 0,$$

which is attained by $\pi^1(r^0, \mathbf{r})$ by choosing $\mathbf{s} = \mathbf{0}$. Second, we note that the epigraph of $\pi^1(r^0, \mathbf{r})$ can be reformulated as a set of robust constraints over the support sets \mathcal{V} and $\hat{\mathcal{V}}$. Since \mathcal{V} and $\hat{\mathcal{V}}$ are tractable conic representable sets by assumption, the bound $\pi^1(r^0, \mathbf{r})$ is consequently computationally tractable.

5.2. Mean and Covariance Information

THEOREM 2. Let \mathbb{F}_2 be the family of all distributions \mathbb{P} such that the mean of the segregated uncertainties, $\hat{\boldsymbol{\zeta}}$, has support in $\hat{\mathcal{V}}$, and the primitive uncertainty vector has covariance matrix $\boldsymbol{\Sigma}$, i.e.,

$$\mathbb{F}_2 = \{\mathbb{P}: \hat{\boldsymbol{\zeta}} = \mathbb{E}_{\mathbb{P}}(\tilde{\boldsymbol{\zeta}}) \in \hat{\mathcal{V}}, \mathbb{E}_{\mathbb{P}}(\mathbf{F}(\tilde{\boldsymbol{\zeta}} - \hat{\boldsymbol{\zeta}})(\tilde{\boldsymbol{\zeta}} - \hat{\boldsymbol{\zeta}})' \mathbf{F}') = \boldsymbol{\Sigma}\}.$$

Then $\pi^2(r^0, \mathbf{r})$ is a tight upper bound for $\mathbb{E}_{\mathbb{P}}((r^0 + \mathbf{r}'\tilde{\boldsymbol{\zeta}})^+)$ over all distributions $\mathbb{P} \in \mathbb{F}_2$, i.e.,

$$\sup_{\mathbb{P} \in \mathbb{F}_2} \mathbb{E}_{\mathbb{P}}((r^0 + \mathbf{r}'\tilde{\boldsymbol{\zeta}})^+) = \pi^2(r^0, \mathbf{r}),$$

where

$$\pi^2(r^0, \mathbf{r}) \triangleq \inf_{\mathbf{y} \in \{\mathbf{y}: \mathbf{F}'\mathbf{y} = \mathbf{r}\}} \left\{ \sup_{\hat{\boldsymbol{\zeta}} \in \hat{\mathcal{V}}} \left\{ \frac{1}{2}(r^0 + \mathbf{r}'\hat{\boldsymbol{\zeta}}) + \frac{1}{2}\sqrt{(r^0 + \mathbf{r}'\hat{\boldsymbol{\zeta}})^2 + \mathbf{y}'\boldsymbol{\Sigma}\mathbf{y}} \right\} \right\}. \quad (22)$$

PROOF. Please see the online Appendix B.2.

REMARK. We observe that the function $f(u) = \frac{1}{2}u + \frac{1}{2}\sqrt{u^2 + \mathbf{y}'\boldsymbol{\Sigma}\mathbf{y}}$ is everywhere nondecreasing in u , which allows us to express $\pi^2(r^0, \mathbf{r})$ as the following tractable conic optimization problem:

$$\pi^2(r^0, \mathbf{r}) = \inf_{u, \mathbf{y}} \frac{1}{2}u + \frac{1}{2}\sqrt{u^2 + \mathbf{y}'\boldsymbol{\Sigma}\mathbf{y}},$$

$$r^0 + \sup_{\hat{\boldsymbol{\zeta}} \in \hat{\mathcal{V}}} \mathbf{r}'\hat{\boldsymbol{\zeta}} \leq u, \quad (23)$$

$$\mathbf{F}'\mathbf{y} = \mathbf{r}.$$

5.3. Mean and Directional Deviation Information

THEOREM 3. Let \mathbb{F}_3 be the family of all distributions \mathbb{P} such that the mean of the segregated uncertainties, $\hat{\boldsymbol{\zeta}}$, has support in $\hat{\mathcal{V}}$ and the projected uncertainty vector $\tilde{\mathbf{z}}_{\sigma} = \mathbf{F}_{\sigma}\hat{\boldsymbol{\zeta}} + \mathbf{g}_{\sigma}$ has independent components with directional deviations bounded above by $\sigma_{\mathbf{f}}$ and $\sigma_{\mathbf{b}}$ for known parameters $(\mathbf{F}_{\sigma}, \mathbf{g}_{\sigma})$, i.e.,

$$\mathbb{F}_3 = \{\mathbb{P}: \hat{\boldsymbol{\zeta}} = \mathbb{E}_{\mathbb{P}}(\tilde{\boldsymbol{\zeta}}) \in \hat{\mathcal{V}}, \sigma_{\mathbf{f}\mathbb{P}}(\tilde{\mathbf{z}}_{\sigma}) \leq \sigma_{\mathbf{f}}, \sigma_{\mathbf{b}\mathbb{P}}(\tilde{\mathbf{z}}_{\sigma}) \leq \sigma_{\mathbf{b}}\}.$$

Then $\pi^3(r^0, \mathbf{r})$ is an upper bound for $\mathbb{E}_{\mathbb{P}}((r^0 + \mathbf{r}'\tilde{\boldsymbol{\zeta}})^+)$ over all distributions $\mathbb{P} \in \mathbb{F}_3$, i.e.,

$$\sup_{\mathbb{P} \in \mathbb{F}_3} \mathbb{E}_{\mathbb{P}}((r^0 + \mathbf{r}'\tilde{\boldsymbol{\zeta}})^+) \leq \pi^3(r^0, \mathbf{r}),$$

where

$$\pi^3(r^0, \mathbf{r}) \triangleq \inf_{\substack{s^0, \mathbf{s}, x^0, \mathbf{x} \\ x^0 + \mathbf{x}'\mathbf{g}_{\sigma} = r^0 \\ \mathbf{F}_{\sigma}\mathbf{x} = \mathbf{r}}} \left\{ (r^0 - s^0 - \mathbf{s}'\mathbf{g}_{\sigma}) + \sup_{\hat{\boldsymbol{\zeta}} \in \hat{\mathcal{V}}} (\mathbf{r}' - \mathbf{s}'\mathbf{F}_{\sigma})\hat{\boldsymbol{\zeta}} + \psi(s^0 - x^0, \mathbf{s} - \mathbf{x}) + \psi(s^0, \mathbf{s}) \right\}, \quad (24)$$

and

$$\psi(x^0, \mathbf{x}) = \inf_{\lambda > 0} \left\{ \frac{\lambda}{e} \exp \left(\frac{1}{\lambda} \sup_{\tilde{\mathbf{z}}_{\sigma} \in \tilde{\mathcal{W}}_{\sigma}} \{x^0 + \mathbf{x}'\tilde{\mathbf{z}}_{\sigma}\} + \frac{\|\mathbf{u}\|_2^2}{2\lambda^2} \right) \right\},$$

and $u_j = \max\{x_j\sigma_{f,j}, -x_j\sigma_{b,j}\}$.

PROOF. Please see the online Appendix B.3.

REMARK. First, we notice that from the Model of Uncertainty in §3 we specify the linear transform parameter \mathbf{H}_{σ} mapping the primitive uncertainty vector $\tilde{\mathbf{z}}$ to $\tilde{\mathbf{z}}_{\sigma}$ instead of specifying the affine transform parameters mapping the segregated uncertainty vector $\tilde{\boldsymbol{\zeta}}$ to $\tilde{\mathbf{z}}_{\sigma}$ directly. Nevertheless, $(\mathbf{F}_{\sigma}, \mathbf{g}_{\sigma})$ is obtained as $\mathbf{F}_{\sigma} = \mathbf{H}_{\sigma}\mathbf{F}$ and $\mathbf{g}_{\sigma} = \mathbf{H}_{\sigma}\mathbf{g}$. Second, unlike $\pi^1(r^0, \mathbf{r})$ and $\pi^2(r^0, \mathbf{r})$, this bound, while valid, is not tight. However, it remains useful because it uses any componentwise independence of the uncertainties to form a bound, and it has been shown computationally by See and Sim (2009) to significantly improve the quality of the solution for a robust inventory problem. Finally, we note that evaluating $\pi^3(r^0, \mathbf{r})$ involves exponential functions with quadratic arguments. Chen and Sim (2009, Appendix B) showed that a small number of second-order conic constraints can be used to approximate such functions with good accuracy and can perform numerical studies showing the usefulness of the bound.

5.4. Unified Bounds

Each of the above functions separately bound $\mathbb{E}_{\mathbb{P}}((r^0 + \mathbf{r}'\tilde{\boldsymbol{\zeta}})^+)$ from above for \mathbb{P} belonging to a given family of distributions $\mathbb{F} \in \{\mathbb{F}_1, \mathbb{F}_2, \mathbb{F}_3\}$. We now consider whether we are able to construct a better bound if we know that the actual distribution \mathbb{P} lies in the intersection of these families. This can be done via a well-known technique in convex analysis known as infimal convolution (see Chen and Sim 2009). Due to its importance in our discussion, we reproduce it here in the following theorem.

THEOREM 4. Let $S \subseteq \{1, 2, 3\}$ be an index set of the bounds to be combined. Then the bound $\pi(r^0, \mathbf{r})$, defined as

$$\begin{aligned} \pi(r^0, \mathbf{r}) = \min & \sum_{s \in S} \pi^s(r^{0,s}, \mathbf{r}^s) \\ \text{s.t.} & r^0 = \sum_{s \in S} r^{0,s}, \\ & \mathbf{r} = \sum_{s \in S} \mathbf{r}^s, \end{aligned} \quad (25)$$

is a better bound for $\sup_{\mathbb{P} \in \mathbb{F}} E_{\mathbb{P}}((r^0 + \mathbf{r}'\tilde{\boldsymbol{\zeta}})^+)$ than $\pi^s(r^0, \mathbf{r})$, $\forall s \in S$, where $\mathbb{F} = \bigcap_{s \in S} \mathbb{F}_s$, such that

$$\sup_{\mathbb{P} \in \mathbb{F}} E_{\mathbb{P}}((r^0 + \mathbf{r}'\tilde{\boldsymbol{\zeta}})^+) \leq \pi(r^0, \mathbf{r}) \leq \pi^s(r^0, \mathbf{r}) \quad \forall s \in S. \quad (26)$$

PROOF. Please see the online Appendix B.4.2.

REMARK. On the theoretical front, we notice that the unified bound $\pi(r^0, \mathbf{r})$ has the mathematically desirable properties of convexity and positive homogeneity. From a practical perspective, $\pi(r^0, \mathbf{r})$ is able to selectively synthesize disparate pieces of information about the distribution of primitive uncertainties and present a combined bound which takes into account all pieces of information.

6. Deflected Linear Decision Rules

We earlier showed that SLDRs improve over LDRs, while retaining a linear model of recourse. We aim to investigate if we can do even better. The deflected linear decision rule (DLDR) proposed by Chen et al. (2008) exploited the *structure* of the model constraints (by solving a series of subproblems based on the model parameters) to obtain an even more flexible decision rule. We adapt this idea here, similarly solving a series of subproblems to exploit structural information within the model to generate a better decision rule, which we term the bideflected linear decision rule (BDLDR).

Although the original DLDR has been shown to be more flexible in comparison to LDRs, we will present an example in §6.1.1 where the DLDR can be further improved upon by an alternate piecewise-linear decision rule, which suggests that there is room for further improvement. In addition, the original DLDR of Chen et al. (2008) does not explicitly handle expectation constraints or nonanticipativity requirements. We seek to address these limitations in the BDLDR which we present here.

Since the basic LDR can be obtained from the SLDR by choosing $\mathbf{M}(\mathbf{z}) = \mathbf{z}$, our results in this section also hold if we choose to omit constructing the SLDR as an intermediate step. However, we choose to present the techniques in this section as an additional layer of improvement over the SLDR for greater generality.

6.1. DLDR of Chen et al. (2008)

We review the two-stage optimization problem as in Chen et al. (2008) under linear recourse with nonnegative constraints for a subset $J \subseteq [m]$ of indices as follows:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{r}(\cdot)} & \mathbf{c}'\mathbf{x} + \sup_{\mathbb{P} \in \mathbb{F}} E_{\mathbb{P}}(\mathbf{d}'\mathbf{r}(\tilde{\boldsymbol{\zeta}})) \\ \text{s.t.} & \mathcal{T}(\tilde{\boldsymbol{\zeta}})\mathbf{x} + \mathbf{U}\mathbf{r}(\tilde{\boldsymbol{\zeta}}) = \mathbf{v}(\tilde{\boldsymbol{\zeta}}), \\ & r_j(\tilde{\boldsymbol{\zeta}}) \geq 0 \quad \forall j \in J, \\ & \mathbf{r} \in \mathcal{L}(m, N_E, \Phi). \end{aligned} \quad (27)$$

We consider a series of subproblems for each $i \in J$:

$$\begin{aligned} \min_{\mathbf{p}} & \mathbf{d}'\mathbf{p} \\ \text{s.t.} & \mathbf{U}\mathbf{p} = \mathbf{0}, \\ & p_i = 1, \\ & p_j \geq 0 \quad \forall j \in J. \end{aligned} \quad (28)$$

Denoting by $J^\circ \subseteq J$ the set of indices where problem (28) has a feasible solution, and by $\bar{\mathbf{p}}^i$, the optimal solution to subproblem (28) for each $i \in J^\circ$, the DLDR is then defined from the SLDR by the relation

$$\hat{\mathbf{r}}_D(\tilde{\boldsymbol{\zeta}}) \triangleq \mathbf{r}(\tilde{\boldsymbol{\zeta}}) + \sum_{i \in J^\circ} (r_i(\tilde{\boldsymbol{\zeta}}))^- \bar{\mathbf{p}}^i, \quad (29)$$

where the SLDR satisfies

$$\begin{aligned} \mathcal{T}(\tilde{\boldsymbol{\zeta}})\mathbf{x} + \mathbf{U}\mathbf{r}(\tilde{\boldsymbol{\zeta}}) &= \mathbf{v}(\tilde{\boldsymbol{\zeta}}), \\ r_j(\tilde{\boldsymbol{\zeta}}) &\geq 0 \quad \forall j \in J \setminus J^\circ, \\ \mathbf{r} &\in \mathcal{L}(m, N_E, \Phi). \end{aligned} \quad (30)$$

Problem (27) under the DLDR can be approximated by

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{r}(\cdot)} & \mathbf{c}'\mathbf{x} + \sup_{\mathbb{P} \in \mathbb{F}} E_{\mathbb{P}}(\mathbf{d}'\mathbf{r}(\tilde{\boldsymbol{\zeta}})) + \sum_{i \in J_R^\circ} \sup_{\mathbb{P} \in \mathbb{F}} E_{\mathbb{P}}((r_i(\tilde{\boldsymbol{\zeta}}))^-) \mathbf{d}'\bar{\mathbf{p}}^i \\ \text{s.t.} & \mathcal{T}(\tilde{\boldsymbol{\zeta}})\mathbf{x} + \mathbf{U}\mathbf{r}(\tilde{\boldsymbol{\zeta}}) = \mathbf{v}(\tilde{\boldsymbol{\zeta}}), \\ & r_j(\tilde{\boldsymbol{\zeta}}) \geq 0 \quad \forall j \in J \setminus J^\circ, \\ & \mathbf{r} \in \mathcal{L}(m, N_E, \Phi), \end{aligned}$$

where we define the reduced index set as $J_R^\circ = \{i \in J^\circ: \mathbf{d}'\bar{\mathbf{p}}^i > 0\}$ to avoid nonconvexity in the objective. The objective is bounded from above by summing over only indices in J_R° , since the respective summation terms for $i \notin J_R^\circ$ are nonpositive. We notice that the objective involves summing over terms of the form $\sup_{\mathbb{P} \in \mathbb{F}} E_{\mathbb{P}}((\cdot)^-)$, which we can bound from above using the unified bound $\pi(\cdot)$ in §5. The two-stage DLDR model can therefore be expressed explicitly as

$$\begin{aligned} Z_{DLDR} = \min_{\mathbf{x}, \mathbf{r}^0, \mathbf{R}} & \mathbf{c}'\mathbf{x} + \mathbf{d}'\mathbf{r}^0 + \sup_{\hat{\boldsymbol{\zeta}} \in \hat{\mathcal{V}}} \{\mathbf{d}'\mathbf{R}\hat{\boldsymbol{\zeta}}\} \\ & + \sum_{i \in J_R^\circ} \pi(-r_i^0, -\mathbf{R}'\mathbf{e}^i) \mathbf{d}'\bar{\mathbf{p}}^i \\ \text{s.t.} & \mathcal{T}^0\mathbf{x} + \mathbf{U}\mathbf{r}^0 = \mathbf{v}^0, \\ & \mathcal{T}^j\mathbf{x} + \mathbf{U}\mathbf{R}\mathbf{e}^j = \mathbf{v}^j \quad \forall j \in [N_E], \\ & r_j^0 + \mathbf{e}^j\mathbf{R}\hat{\boldsymbol{\zeta}} \geq 0 \quad \forall \hat{\boldsymbol{\zeta}} \in \hat{\mathcal{V}} \quad \forall j \in J \setminus J^\circ. \end{aligned} \quad (31)$$

6.1.1. Example: Limitation of DLDR and LDR. We will now consider an example of the two-stage Problem (27), which will illustrate the limitation of the DLDR and motivate our subsequent exposition. For simplicity, we will discuss applying the DLDR to the LDR instead of the SLDR (i.e., using $\mathbf{M}(\mathbf{z}) = \mathbf{z}$). We consider the family \mathbb{F} of scalar ($N = 1$) uncertainty distributions with infinite support ($\mathcal{W} = \mathbb{R}$), zero mean ($\widehat{\mathcal{W}} = \{0\}$), and unit variance ($\sigma^2 = 1$). We consider the following uncertain optimization problem with scalar recourse variables $y(\tilde{z})$, $u(\tilde{z})$, and $v(\tilde{z})$:

$$\begin{aligned} \min_{y(\cdot), u(\cdot), v(\cdot)} \quad & \sup_{\mathbb{P} \in \mathbb{F}} E_{\mathbb{P}}(u(\tilde{z}) + v(\tilde{z})) \\ \text{s.t.} \quad & u(\tilde{z}) - v(\tilde{z}) = y(\tilde{z}) - \tilde{z}, \\ & 0 \leq y(\tilde{z}) \leq 1, \\ & u(\tilde{z}), v(\tilde{z}) \geq 0, \\ & y, u, v \in \mathcal{Y}(1, 1, \{1\}). \end{aligned} \tag{32}$$

Using LDRs as our model of recourse, (i.e., using \mathcal{L} to approximate \mathcal{Y}), we note that Problem (32) is infeasible (i.e., $Z_{\text{LDR}} = +\infty$), since the inequalities over the infinite support of \tilde{z} cannot simultaneously fulfill the equality constraint. Now suppose we attempt to apply the DLDR to improve the solution: We will need a slack variable to convert the problem into the form of (27) which has model parameters

$$\mathbf{U} = \begin{bmatrix} 1 & -1 & -1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{d} = [1 \quad 1 \quad 0 \quad 0],$$

$$J = \{1, 2, 3, 4\}.$$

Solving subproblem (28) leads to the piecewise-linear decision rules:

$$\begin{aligned} \hat{u}_D(z) &= (u^0 + uz)^+ + (v^0 + vz)^-, \\ \hat{v}_D(z) &= (v^0 + vz)^+ + (u^0 + uz)^-, \\ y(z) &= y^0 + yz. \end{aligned}$$

Applying these decision rules, we obtain the following reformulation

$$\begin{aligned} \min_{y^0, y, u^0, u, v^0, v} \quad & \sup_{\mathbb{P} \in \mathbb{F}} E_{\mathbb{P}}(+) \\ \text{s.t.} \quad & u^0 - v^0 = y^0, \\ & u - v = y - 1, \\ & 0 \leq y^0 + yz \leq 1 \quad \forall z \in \mathbb{R}. \end{aligned}$$

After applying the bounds in §5 and noticing that the last inequality over all of \mathbb{R} implies $y = 0$, we get the final deterministic formulation which determines Z_{DLDR} :

$$\begin{aligned} Z_{\text{DLDR}} &= \min_{u^0, u, v^0, v} \left\| \begin{pmatrix} u^0 \\ u \end{pmatrix} \right\|_2 + \left\| \begin{pmatrix} v^0 \\ v \end{pmatrix} \right\|_2 \\ \text{s.t.} \quad & u - v = -1, \\ & 0 \leq u^0 - v^0 \leq 1. \end{aligned} \tag{33}$$

Solving, we get $Z_{\text{DLDR}} = 1$, which is a significant improvement over the LDR solution. We notice that even after applying the DLDR, the decision rule $y(\cdot)$ remains as an LDR, and we would like to investigate whether we can further improve on this. Now, suppose we consider the following hypothetical piecewise-linear decision rule:

$$\begin{aligned} \hat{u}(z) &= (u^0 + uz)^+ + (v^0 + vz)^- + (y^0 + yz)^-, \\ \hat{v}(z) &= (v^0 + vz)^+ + (u^0 + uz)^- + (y^0 - 1 + yz)^+, \\ \hat{y}(z) &= (y^0 + yz)^+ - (y^0 - 1 + yz)^+. \end{aligned}$$

We notice that under these decision rules, Problem (32), after applying the bounds, can be reduced to

$$\begin{aligned} Z_0 &= \min_{y^0, y, u^0, u, v^0, v} \left\| \begin{pmatrix} u^0 \\ u \end{pmatrix} \right\|_2 + \left\| \begin{pmatrix} v^0 \\ v \end{pmatrix} \right\|_2 + \frac{1}{2} \left\| \begin{pmatrix} y^0 \\ y \end{pmatrix} \right\|_2 \\ &\quad + \frac{1}{2} \left\| \begin{pmatrix} y^0 - 1 \\ y \end{pmatrix} \right\|_2 - \frac{1}{2} \\ \text{s.t.} \quad & u^0 - v^0 = y^0, \\ & u - v = y - 1. \end{aligned} \tag{34}$$

Solving the problem above, the optimal value is given by $Z_0 = 1/\sqrt{2} < Z_{\text{DLDR}}$, a further improvement over the DLDR. We therefore seek a decision rule that would encompass our hypothetical piecewise-linear model of recourse.

6.2. Two-Stage Bideflected Linear Decision Rule

In this subsection, we will first introduce the BDLDR for a two-stage problem and later generalize it to a non-anticipative BDLDR (including the multistage as a special case) in the following subsection. We consider a two-stage problem similar to (27), as follows:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{r}(\cdot)} \quad & \mathbf{c}'\mathbf{x} + \sup_{\mathbb{P} \in \mathbb{F}} E_{\mathbb{P}}(\mathbf{d}'\mathbf{r}(\tilde{\xi})) \\ \text{s.t.} \quad & \mathcal{T}(\tilde{\xi})\mathbf{x} + \mathbf{U}\mathbf{r}(\tilde{\xi}) = \mathbf{v}(\tilde{\xi}), \\ & \underline{\mathbf{y}} \leq \mathbf{r}(\tilde{\xi}) \leq \bar{\mathbf{y}}, \\ & \mathbf{r} \in \mathcal{L}(m, N_E, \Phi). \end{aligned} \tag{35}$$

Notice that for a two-stage problem we should have $\Phi = [N_E]$. Similar to the definition (3) of noninfinite bounds in our general model, we denote the index sets of noninfinite bounds in our two-stage model:

$$\begin{aligned} \underline{J} &= \{i \in [m]: \underline{y}_i > -\infty\}, \\ \bar{J} &= \{i \in [m]: \bar{y}_i < +\infty\}. \end{aligned} \tag{36}$$

Notice that if we were to choose

$$\begin{aligned} \underline{y}_j &= \begin{cases} 0 & \forall j \in J, \\ -\infty & \forall j \in [m] \setminus J, \end{cases} \\ \bar{y}_j &= +\infty \quad \forall j \in [m], \end{aligned} \tag{37}$$

we obtain model (27) exactly. To construct the BDLDR, we consider the following pairs of optimization problems. First, for each $i \in \underline{J}$,

$$\begin{aligned} \min_{\mathbf{p}} \quad & \mathbf{d}'\mathbf{p} \\ \text{s.t.} \quad & \mathbf{U}\mathbf{p} = \mathbf{0}, \\ & p_i = 1, \\ & p_j \geq 0 \quad \forall j \in \underline{J}, \\ & p_j \leq 0 \quad \forall j \in \bar{J} \setminus \{i\}. \end{aligned} \quad (38)$$

Notice that for $j \in (\underline{J} \cap \bar{J}) \setminus \{i\}$, the constraints imply that $p_j = 0$. Similarly, for $i \in \bar{J}$,

$$\begin{aligned} \min_{\mathbf{q}} \quad & \mathbf{d}'\mathbf{q} \\ \text{s.t.} \quad & \mathbf{U}\mathbf{q} = \mathbf{0}, \\ & q_i = -1, \\ & q_j \leq 0 \quad \forall j \in \bar{J}, \\ & q_j \geq 0 \quad \forall j \in \underline{J} \setminus \{i\}. \end{aligned} \quad (39)$$

By defining $\underline{J}^\circ \subseteq \underline{J}$ as the index set of i such that problem (38) is feasible, $\bar{J}^\circ \subseteq \bar{J}$ as the index set of i such that problem (39) is feasible, and $\bar{\mathbf{p}}^i, \bar{\mathbf{q}}^i$ as the respective optimal solutions for each i in \underline{J}° and \bar{J}° respectively, we consider an SLDR satisfying

$$\begin{aligned} \mathcal{J}(\tilde{\zeta})\mathbf{x} + \mathbf{U}\mathbf{r}(\tilde{\zeta}) &= \nu(\tilde{\zeta}), \\ r_j(\tilde{\zeta}) &\geq \underline{y}_j \quad \forall j \in \underline{J} \setminus \underline{J}^\circ, \\ r_j(\tilde{\zeta}) &\leq \bar{y}_j \quad \forall j \in \bar{J} \setminus \bar{J}^\circ. \end{aligned} \quad (40)$$

The associated BDLDR is then defined from the SLDR by $\hat{\mathbf{r}}(\tilde{\zeta}) \triangleq \mathbf{r}(\tilde{\zeta}) + \sum_{i \in \underline{J}^\circ} (r_i(\tilde{\zeta}) - \underline{y}_i)^- \bar{\mathbf{p}}^i + \sum_{i \in \bar{J}^\circ} (r_i(\tilde{\zeta}) - \bar{y}_i)^+ \bar{\mathbf{q}}^i$. (41)

Some properties of the BDLDR are stated in Proposition 4, which follows.

PROPOSITION 4. *The BDLDR, $\hat{\mathbf{r}}(\tilde{\zeta})$, satisfies the following properties:*

1. $\mathbf{U}\hat{\mathbf{r}}(\tilde{\zeta}) = \mathbf{U}\mathbf{r}(\tilde{\zeta})$,
2. $\underline{\mathbf{y}} \leq \hat{\mathbf{r}}(\tilde{\zeta}) \leq \bar{\mathbf{y}}$.

PROOF. Please see the Appendix C.1.2.

This implies that as long as we have an SLDR that satisfies (40), we can find a feasible BDLDR. Under the BDLDR, problem (35) becomes

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{r}(\cdot)} \quad & \mathbf{c}'\mathbf{x} + \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}} \left(\mathbf{d}'\mathbf{r}(\tilde{\zeta}) + \sum_{i \in \underline{J}^\circ} (r_i(\tilde{\zeta}) - \underline{y}_i)^- \mathbf{d}'\bar{\mathbf{p}}^i \right. \\ & \left. + \sum_{i \in \bar{J}^\circ} (r_i(\tilde{\zeta}) - \bar{y}_i)^+ \mathbf{d}'\bar{\mathbf{q}}^i \right) \\ \text{s.t.} \quad & \mathcal{J}(\tilde{\zeta})\mathbf{x} + \mathbf{U}\mathbf{r}(\tilde{\zeta}) = \mathbf{v}(\tilde{\zeta}), \\ & r_j(\tilde{\zeta}) \geq \underline{y}_j \quad \forall j \in \underline{J} \setminus \underline{J}^\circ, \\ & r_j(\tilde{\zeta}) \leq \bar{y}_j \quad \forall j \in \bar{J} \setminus \bar{J}^\circ, \\ & \mathbf{r} \in \mathcal{L}(m, N_E, \Phi). \end{aligned} \quad (42)$$

When $\mathbf{d}'\bar{\mathbf{p}}^i$ or $\mathbf{d}'\bar{\mathbf{q}}^i$ is negative, the objective becomes nonconvex. Thus, we consider an approximation of Problem (42) from above by defining the reduced index sets:

$$\begin{aligned} \underline{J}_R^\circ &\triangleq \{i \in \underline{J}^\circ: \mathbf{d}'\bar{\mathbf{p}}^i > 0\}, \\ \bar{J}_R^\circ &\triangleq \{i \in \bar{J}^\circ: \mathbf{d}'\bar{\mathbf{q}}^i > 0\}. \end{aligned} \quad (43)$$

We then use the subadditivity of the supremum to obtain the formulation of the BDLDR problem:

$$\begin{aligned} Z_{BDLDR}^* &= \min_{\mathbf{x}, \mathbf{r}(\cdot)} \mathbf{c}'\mathbf{x} + \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}}(\mathbf{d}'\mathbf{r}(\tilde{\zeta})) \\ &\quad + \sum_{i \in \underline{J}_R^\circ} \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}}((r_i(\tilde{\zeta}) - \underline{y}_i)^-) \mathbf{d}'\bar{\mathbf{p}}^i \\ &\quad + \sum_{i \in \bar{J}_R^\circ} \sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}}((r_i(\tilde{\zeta}) - \bar{y}_i)^+) \mathbf{d}'\bar{\mathbf{q}}^i \\ \text{s.t.} \quad & \mathcal{J}(\tilde{\zeta})\mathbf{x} + \mathbf{U}\mathbf{r}(\tilde{\zeta}) = \nu(\tilde{\zeta}), \\ & r_j(\tilde{\zeta}) \geq \underline{y}_j \quad \forall j \in \underline{J} \setminus \underline{J}^\circ, \\ & r_j(\tilde{\zeta}) \leq \bar{y}_j \quad \forall j \in \bar{J} \setminus \bar{J}^\circ, \\ & \mathbf{r} \in \mathcal{L}(m, N_E, \Phi). \end{aligned}$$

Using the bounds developed in §5 to approximate $\sup_{\mathbb{P} \in \mathbb{F}} \mathbb{E}_{\mathbb{P}}((\cdot)^\pm)$, we obtain the explicit final form of the BDLDR model:

$$\begin{aligned} Z_{BDLDR} &= \min_{\mathbf{x}, \mathbf{r}^0, \mathbf{R}} \mathbf{c}'\mathbf{x} + \mathbf{d}'\mathbf{r}^0 + \sup_{\tilde{\zeta} \in \mathcal{V}} \{\mathbf{d}'\mathbf{R}\tilde{\zeta}\} \\ &\quad + \sum_{i \in \underline{J}_R^\circ} \pi(-r_i^0 + \underline{y}_i, -\mathbf{R}'\mathbf{e}^i) \mathbf{d}'\bar{\mathbf{p}}^i \\ &\quad + \sum_{i \in \bar{J}_R^\circ} \pi(r_i^0 - \bar{y}_i, \mathbf{R}'\mathbf{e}^i) \mathbf{d}'\bar{\mathbf{q}}^i \\ \text{s.t.} \quad & \mathcal{J}^0\mathbf{x} + \mathbf{U}\mathbf{r}^0 = \mathbf{v}^0, \\ & \mathcal{J}^j\mathbf{x} + \mathbf{U}\mathbf{R}\mathbf{e}^j = \mathbf{v}^j \quad \forall j \in [N_E], \\ & r_j^0 + \mathbf{e}^{j'}\mathbf{R}\tilde{\zeta} \geq \underline{y}_j \quad \forall \tilde{\zeta} \in \mathcal{V} \quad \forall j \in \underline{J} \setminus \underline{J}^\circ, \\ & r_j^0 + \mathbf{e}^{j'}\mathbf{R}\tilde{\zeta} \leq \bar{y}_j \quad \forall \tilde{\zeta} \in \mathcal{V} \quad \forall j \in \bar{J} \setminus \bar{J}^\circ. \end{aligned} \quad (44)$$

6.3. Comparison of BDLDR with DLDR and LDR

In this subsection we will proceed to show that the BDLDR improves upon the DLDR and SLDR for Problem (35). Without loss of generality, we can consider a simplified version of the problem, such that the lower recourse constraint has the structure,

$$\underline{y}_j = 0, \quad \forall j \in \underline{J},$$

since we can simply apply a change of variables if the above was not true. For reference, we begin by writing Problem (35) under the SLDR explicitly as

$$\begin{aligned}
 Z_{SLDR} = \min_{\mathbf{x}, \mathbf{r}^0, \mathbf{R}} \quad & \mathbf{c}'\mathbf{x} + \mathbf{d}'\mathbf{r}^0 + \sup_{\hat{\boldsymbol{\zeta}} \in \hat{\mathcal{V}}} \{\mathbf{d}'\mathbf{R}\hat{\boldsymbol{\zeta}}\} \\
 \text{s.t.} \quad & \mathcal{J}^0\mathbf{x} + \mathbf{U}\mathbf{r}^0 = \mathbf{v}^0, \\
 & \mathcal{J}^j\mathbf{x} + \mathbf{U}\mathbf{R}\mathbf{e}^j = \mathbf{v}^j \quad \forall j \in [N_E], \\
 & r_j^0 + \mathbf{e}^j\mathbf{R}\hat{\boldsymbol{\zeta}} \geq y_j \quad \forall \hat{\boldsymbol{\zeta}} \in \hat{\mathcal{V}} \quad \forall j \in \underline{J}, \\
 & r_j^0 + \mathbf{e}^j\mathbf{R}\hat{\boldsymbol{\zeta}} \leq \bar{y}_j \quad \forall \hat{\boldsymbol{\zeta}} \in \hat{\mathcal{V}} \quad \forall j \in \bar{J}.
 \end{aligned} \tag{45}$$

Next, we consider the DLDR. In order to apply the DLDR, we need to introduce slack linear recourse variables $s_j(\tilde{\boldsymbol{\zeta}})$, $\forall j \in \bar{J}$ to convert (35) into the form of (27). The constraint set becomes

$$\begin{aligned}
 \mathcal{J}(\tilde{\boldsymbol{\zeta}})\mathbf{x} + \mathbf{U}\mathbf{r}(\tilde{\boldsymbol{\zeta}}) &= \mathbf{v}(\tilde{\boldsymbol{\zeta}}), \\
 r_j(\tilde{\boldsymbol{\zeta}}) + s_j(\tilde{\boldsymbol{\zeta}}) &= \bar{y}_j \quad \forall j \in \bar{J}, \\
 r_j(\tilde{\boldsymbol{\zeta}}) &\geq 0 \quad \forall j \in \underline{J}, \\
 s_j(\tilde{\boldsymbol{\zeta}}) &\geq 0 \quad \forall j \in \bar{J}, \\
 \mathbf{r}, \mathbf{s} &\in \mathcal{L}(m, N_E, \Phi).
 \end{aligned}$$

Under the DLDR model of recourse, we need to solve the subproblem (28) for each $i \in \underline{J}$, which corresponds to the inequalities $r_i(\tilde{\boldsymbol{\zeta}}) \geq 0$. After eliminating the slack variables, the problem takes the form

$$\begin{aligned}
 \min_{\mathbf{p}} \quad & \mathbf{d}'\mathbf{p} \\
 \text{s.t.} \quad & \mathbf{U}\mathbf{p} = \mathbf{0}, \\
 & p_i = 1, \\
 & p_j \geq 0 \quad \forall j \in \underline{J}, \\
 & p_j \leq 0 \quad \forall j \in \bar{J}.
 \end{aligned} \tag{46}$$

Similarly, the inequalities $s_i(\tilde{\boldsymbol{\zeta}}) \geq 0$ require us to solve the following subproblem (again after eliminating the slacks) for each $i \in \bar{J}$:

$$\begin{aligned}
 \min_{\mathbf{q}} \quad & \mathbf{d}'\mathbf{q} \\
 \text{s.t.} \quad & \mathbf{U}\mathbf{q} = \mathbf{0}, \\
 & q_i = -1, \\
 & q_j \geq 0 \quad \forall j \in \underline{J}, \\
 & q_j \leq 0 \quad \forall j \in \bar{J}.
 \end{aligned} \tag{47}$$

We denote by \underline{J}_D^o the set of indices $i \in \underline{J}$ such that (46) has a feasible solution, with corresponding optimal solution $\bar{\mathbf{p}}_D^i$, and by \bar{J}_D^o the set of indices $i \in \bar{J}$ such that (47) is feasible, with corresponding optimal solution $\bar{\mathbf{q}}_D^i$. From

the DLDR formulation (31), after rearrangement of terms, Problem (35) under the DLDR reduces to:

$$\begin{aligned}
 Z_{DLDR} = \min_{\mathbf{x}, \mathbf{r}^0, \mathbf{R}} \quad & \mathbf{c}'\mathbf{x} + \mathbf{d}'\mathbf{r}^0 + \sup_{\hat{\boldsymbol{\zeta}} \in \hat{\mathcal{V}}} \{\mathbf{d}'\mathbf{R}\hat{\boldsymbol{\zeta}}\} \\
 & + \sum_{i \in \underline{J}_{D,R}^o} \pi(-r_i^0 + y_i, -\mathbf{R}'\mathbf{e}^i) \mathbf{d}'\bar{\mathbf{p}}_D^i \\
 & + \sum_{i \in \bar{J}_{D,R}^o} \pi(r_i^0 - \bar{y}_i, \mathbf{R}'\mathbf{e}^i) \mathbf{d}'\bar{\mathbf{q}}_D^i \\
 \text{s.t.} \quad & \mathcal{J}^0\mathbf{x} + \mathbf{U}\mathbf{r}^0 = \mathbf{v}^0, \\
 & \mathcal{J}^j\mathbf{x} + \mathbf{U}\mathbf{R}\mathbf{e}^j = \mathbf{v}^j \quad \forall j \in [N_E], \\
 & r_j^0 + \mathbf{e}^j\mathbf{R}\hat{\boldsymbol{\zeta}} \geq y_j \quad \forall \hat{\boldsymbol{\zeta}} \in \hat{\mathcal{V}} \quad \forall j \in \underline{J} \setminus \underline{J}_D^o, \\
 & r_j^0 + \mathbf{e}^j\mathbf{R}\hat{\boldsymbol{\zeta}} \leq \bar{y}_j \quad \forall \hat{\boldsymbol{\zeta}} \in \hat{\mathcal{V}} \quad \forall j \in \bar{J} \setminus \bar{J}_D^o,
 \end{aligned} \tag{48}$$

where we define the reduced index sets $\underline{J}_{D,R}^o \triangleq \{i \in \underline{J}_D^o : \mathbf{d}'\bar{\mathbf{p}}_D^i > 0\}$ and $\bar{J}_{D,R}^o \triangleq \{i \in \bar{J}_D^o : \mathbf{d}'\bar{\mathbf{q}}_D^i > 0\}$. We now summarize the result relating the optimal objectives to Problem (35) under the SLDR, DLDR, and BDLDR in the following proposition:

PROPOSITION 5. *The optimal objective to Problem (35) under the BDLDR, DLDR, and the SLDR are related by the inequality $Z_{BDLDR} \leq Z_{DLDR} \leq Z_{SLDR}$.*

PROOF. Please see the online Appendix C.2.

REMARK. The discussion above underscores an important distinction between deterministic linear optimization and robust linear optimization: In the deterministic case, it is trivial to include slack variables to convert the feasible set of a linear program into the standard form $\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$. The decision to convert or not is typically a result of the trade-off between storage space and solver performance, and it would not affect the final optimal solution. However, in the robust case, using the LP “standard form” obscures the information of the set \bar{J} , which is in turn required to use the BDLDR. As we have seen, this could potentially deteriorate the optimal solution of the robust linear program.

6.4. Nonanticipative Bideflected Linear Decision Rule

In the previous section, we introduced the two-stage BDLDR and showed that it generalizes and improves upon the DLDR and SLDR. We notice that each component of the two-stage BDLDR is a sum of its original underlying SLDR and piecewise-linear functions of other SLDRs, disregarding any information dependencies between the original SLDRs. Using the two-stage BDLDR could therefore violate nonanticipativity constraints present in the model and result in an optimal solution which nonetheless improves upon the SLDR but would otherwise be practically meaningless. In this section, we will further adapt the BDLDR for the case of nonanticipative recourse. This

includes, but is not restricted to, multistage models. We will introduce the following notation:

$$\begin{aligned} N^+(k) &= \{j \in [K]: \Phi_k \subseteq \Phi_j\}, \\ N^-(k) &= \{j \in [K]: \Phi_j \subseteq \Phi_k\}. \end{aligned} \quad (49)$$

Notice that $k \in N^\pm(k) \forall k \in [K]$, guaranteeing that $N^\pm(k)$ cannot be empty. Furthermore, the following property follows directly from the definitions above:

$$j \in N^+(k) \Leftrightarrow k \in N^-(j). \quad (50)$$

Now, for each $k \in [K]$, $i \in \underline{J}_k$, we consider the polyhedron $P(i, k)$ defined by the constraints

$$\begin{aligned} \sum_{j \in N^+(k)} \mathbf{U}^j \mathbf{p}^{i,k,j} &= \mathbf{0}, \\ p_l^{i,k,j} &\geq 0 \quad \forall l \in \underline{J}_j \quad \forall j \in N^+(k), \\ p_l^{i,k,j} &\leq 0 \quad \forall l \in \bar{J}_j \quad \forall j \in N^+(k) \setminus \{k\}, \\ p_l^{i,k,k} &\leq 0 \quad \forall l \in \bar{J}_k \setminus \{i\}, \\ p_i^{i,k,k} &= 1. \end{aligned} \quad (51)$$

Similarly, for each $k \in [K]$, $i \in \bar{J}_k$, we consider the polyhedron $Q(i, k)$ defined by the constraints

$$\begin{aligned} \sum_{j \in N^+(k)} \mathbf{U}^j \mathbf{q}^{i,k,j} &= \mathbf{0}, \\ q_l^{i,k,j} &\leq 0 \quad \forall l \in \bar{J}_j \quad \forall j \in N^+(k), \\ q_l^{i,k,j} &\geq 0 \quad \forall l \in \underline{J}_j \quad \forall j \in N^+(k) \setminus \{k\}, \\ q_l^{i,k,k} &\geq 0 \quad \forall l \in \underline{J}_k \setminus \{i\}, \\ q_i^{i,k,k} &= -1. \end{aligned} \quad (52)$$

For convenience, we collect the indices i which yield feasible instances of (51) and (52) in the following index sets $\forall k \in [K]$:

$$\begin{aligned} \underline{J}_k^o &= \{i \subseteq \underline{J}_k: P(i, k) \neq \emptyset\}, \\ \bar{J}_k^o &= \{i \subseteq \bar{J}_k: Q(i, k) \neq \emptyset\}. \end{aligned} \quad (53)$$

Now suppose that we have a set of SLDRs, $\mathbf{r}^k \in \mathcal{L}(m_k, N_E, \Phi_k)$, $\forall k \in [K]$, which satisfies

$$\begin{aligned} \mathcal{J}(\tilde{\zeta})\mathbf{x} + \sum_{k=1}^K \mathbf{U}^k \mathbf{r}^k(\tilde{\zeta}) &= \mathbf{v}(\tilde{\zeta}), \\ r_j^k(\tilde{\zeta}) &\geq \underline{y}_j^k \quad \forall k \in [K], \forall j \in \underline{J}_k \setminus \underline{J}_k^o, \\ r_j^k(\tilde{\zeta}) &\leq \bar{y}_j^k \quad \forall k \in [K], \forall j \in \bar{J}_k \setminus \bar{J}_k^o, \\ \mathbf{r}^k &\in \mathcal{L}(m_k, N_E, \Phi_k) \quad \forall k \in [K]. \end{aligned} \quad (54)$$

Based on the SLDR, we define the nonanticipative BDLDR, denoted by $\hat{\mathbf{r}}^k(\tilde{\zeta})$, $\forall k \in [K]$, as

$$\begin{aligned} \hat{\mathbf{r}}^k(\tilde{\zeta}) &\triangleq \mathbf{r}^k(\tilde{\zeta}) + \sum_{j \in N^-(k)} \left(\sum_{i \in \underline{J}_j^o} (r_i^j(\tilde{\zeta}) - \underline{y}_i^j)^- \mathbf{p}^{i,j,k} \right. \\ &\quad \left. + \sum_{i \in \bar{J}_j^o} (r_i^j(\tilde{\zeta}) - \bar{y}_i^j)^+ \mathbf{q}^{i,j,k} \right). \end{aligned} \quad (55)$$

6.5. Properties of BDLDR

PROPOSITION 6. *The information index set of the k th BDLDR, denoted by $\hat{\Phi}_k$, is contained in the information index set of its underlying SLDR, Φ_k , i.e., $\hat{\Phi}_k \subseteq \Phi_k$. Furthermore, if problem (18) has a feasible solution, then $\hat{\Phi}_k = \Phi_k$.*

PROOF. Please see the online Appendix C.3.

REMARK. This proposition implies that the nonanticipative BDLDR only uses information that is available to the the SLDR in forming the recourse decision. For example, if we are modeling a problem with a temporal revelation of information, the nonanticipative BDLDR uses information that has *already* been revealed.

PROPOSITION 7. *Each nonanticipative BDLDR, $\hat{\mathbf{r}}^k(\tilde{\zeta})$ satisfies the following properties:*

1. $\sum_{k=1}^K \mathbf{U}^k \hat{\mathbf{r}}^k(\tilde{\zeta}) = \sum_{k=1}^K \mathbf{U}^k \mathbf{r}^k(\tilde{\zeta})$,
2. $\underline{\mathbf{y}}^k \leq \hat{\mathbf{r}}^k(\tilde{\zeta}) \leq \bar{\mathbf{y}}^k, \forall k \in [K]$.

PROOF. Please see the online Appendix C.4.

6.6. Comparison of Nonanticipative BDLDR with SLDR

We define the reduced index sets $\forall l \in \{0\} \cup [M], \forall k \in [K], \forall j \in N^-(k)$ as

$$\begin{aligned} \underline{J}_{l,j,k}^o &\triangleq \{i \in \underline{J}_j^o: \mathbf{d}^{l,k'} \mathbf{p}^{i,j,k} > 0\}, \\ \bar{J}_{l,j,k}^o &\triangleq \{i \in \bar{J}_j^o: \mathbf{d}^{l,k'} \mathbf{q}^{i,j,k} > 0\}. \end{aligned}$$

Using Proposition 7 Problem (2) under the BDLDR is then approximated as

$$\begin{aligned} Z_{BDLDR} &= \min_{\mathbf{x}, \{\mathbf{r}^{0,k}, \mathbf{R}^k\}_{k=1}^K} \mathbf{c}^{0'} \mathbf{x} + \sum_{k=1}^K \mathbf{d}^{0,k'} \mathbf{r}^{0,k} + \sup_{\hat{\zeta} \in \hat{\mathcal{Z}}} \left\{ \sum_{k=1}^K \mathbf{d}^{0,k'} \mathbf{R}^k \hat{\zeta} \right\} \\ &\quad + \sum_{k=1}^K \sum_{j \in N^-(k)} \sum_{i \in \underline{J}_{0,j,k}^o} \pi(-r_i^{0,j} + \underline{y}_i^j, -\mathbf{R}^{j'} \mathbf{e}^i) \mathbf{d}^{0,k'} \mathbf{p}^{i,j,k} \\ &\quad + \sum_{k=1}^K \sum_{j \in N^-(k)} \sum_{i \in \bar{J}_{0,j,k}^o} \pi(r_i^{0,j} - \bar{y}_i^j, \mathbf{R}^{j'} \mathbf{e}^i) \mathbf{d}^{0,k'} \mathbf{q}^{i,j,k} \\ \text{s.t. } &\mathbf{c}^{l'} \mathbf{x} + \sum_{k=1}^K \mathbf{d}^{l,k'} \mathbf{r}^{0,k} + \sup_{\hat{\zeta} \in \hat{\mathcal{Z}}} \left\{ \sum_{k=1}^K \mathbf{d}^{l,k'} \mathbf{R}^k \hat{\zeta} \right\} \\ &\quad + \sum_{k=1}^K \sum_{j \in N^-(k)} \sum_{i \in \underline{J}_{l,j,k}^o} \pi(-r_i^{0,j} + \underline{y}_i^j, -\mathbf{R}^{j'} \mathbf{e}^i) \mathbf{d}^{l,k'} \mathbf{p}^{i,j,k} \\ &\quad + \sum_{k=1}^K \sum_{j \in N^-(k)} \sum_{i \in \bar{J}_{l,j,k}^o} \pi(r_i^{0,j} - \bar{y}_i^j, \mathbf{R}^{j'} \mathbf{e}^i) \mathbf{d}^{l,k'} \mathbf{q}^{i,j,k} \leq b_l \\ &\quad \forall l \in [M], \\ &\mathcal{J}^0 \mathbf{x} + \sum_{k=1}^K \mathbf{U}^k \mathbf{r}^{0,k} = \mathbf{v}^0, \end{aligned}$$

$$\begin{aligned}
\mathcal{J}^j \mathbf{x} + \sum_{k=1}^K \mathbf{U}^k \mathbf{R}^k \mathbf{e}^j &= \mathbf{v}^j \quad \forall k \in [K], \forall j \in [N_E], \\
r_j^{0,k} + \mathbf{e}^{j'} \mathbf{R}^k \boldsymbol{\zeta} &\geq \underline{y}_j^k \\
&\quad \forall \boldsymbol{\zeta} \in \mathcal{V} \quad \forall k \in [K], j \in \underline{J}_k \setminus \underline{J}_k^o, \\
r_j^{0,k} + \mathbf{e}^{j'} \mathbf{R}^k \boldsymbol{\zeta} &\leq \bar{y}_j^k \\
&\quad \forall \boldsymbol{\zeta} \in \mathcal{V} \quad \forall k \in [K], j \in \bar{J}_k \setminus \bar{J}_k^o, \\
\mathbf{x} &\geq \mathbf{0}.
\end{aligned} \tag{56}$$

PROPOSITION 8. Problem (56) has an optimal objective not worse than Problem (18), i.e., $Z_{BDLDR} \leq Z_{SLDR}$.

PROOF. Please see the online Appendix C.5.

REMARK. Although we have shown that the non-anticipative BDLDR improves upon the SLDR solution, we notice that the BDLDR requires a choice of feasible points $\{\mathbf{p}^{i,k,j}\}_{j \in N^+(k)}$ and $\{\mathbf{q}^{i,k,j}\}_{j \in N^+(k)}$ which satisfy the polyhedral constraint sets (51) and (52), respectively. We leave it as an open question how to optimally choose points within these feasible polyhedra in the general case. If the model (2) can be reformulated such that there are no expectation constraints, we have a similar situation to the simpler two-stage BDLDR, and the optimal choices $\{\bar{\mathbf{p}}^{i,k,j}\}_{j \in N^+(k)}$ and $\{\bar{\mathbf{q}}^{i,k,j}\}_{j \in N^+(k)}$ can be found from the solutions of the following pairs of optimization problems:

$$\begin{aligned}
\min \quad & \sum_{j \in N^+(k)} \mathbf{d}^{0,j} \mathbf{p}^{i,k,j} \\
\text{s.t.} \quad & \{\mathbf{p}^{i,k,j}\}_{j \in N^+(k)} \in P(i, k), \\
\min \quad & \sum_{j \in N^+(k)} \mathbf{d}^{0,j} \mathbf{q}^{i,k,j} \\
\text{s.t.} \quad & \{\mathbf{q}^{i,k,j}\}_{j \in N^+(k)} \in Q(i, k).
\end{aligned} \tag{57}$$

In the general case, due to the coupling between the expectation constraints and the objective, solving the pair of subproblems (57) will no longer guarantee optimal $\{\bar{\mathbf{p}}^{i,k,j}\}_{j \in N^+(k)}$ and $\{\bar{\mathbf{q}}^{i,k,j}\}_{j \in N^+(k)}$ for use in the BDLDR. However, since (57) explicitly decreases the objective of the original problem, we feel that it remains as a viable heuristic for choosing feasible points for use in the BDLDR, which will nonetheless be an improvement over the original SLDR.

7. Conclusions

We have presented a framework for the robust optimization of linear programs under uncertainty, by using linear-based decision rules to model the recourse variables. We have introduced SLDRs and BDLDRs, which are more flexible models of recourse decisions than LDRs, and have shown how they can be used in a nonanticipative modeling context. In a parallel work (Goh and Sim 2009), we

are developing a modeling language to model and solve the class of problems described in this paper. There we present modeling examples and comprehensive numerical studies for a service-constrained inventory management problem and a portfolio optimization problem. In particular, we demonstrate numerically that the nonanticipative BDLDR improves significantly over the LDR, verifying some of the theoretical results developed in this paper.

8. Electronic Companion

Appendices A, B, and C are available in the electronic companion to this paper, which is available as part of the online version at <http://or.journal.informs.org/>.

Endnotes

1. While the results in this paper will still hold even if the definition encompasses semidefinite cones, we focus on second-order conic programs (SOCP) because the study of semidefinite programming (SDP) is still an active area of research, while SOCPs can already be solved with high efficiency and SOCP solvers are even commercially available.
2. Although our analysis can be extended to the case of unknown covariance (using, e.g., techniques developed by Delage and Ye 2010), we do not consider this generalization for two reasons. First, this generally increases the computational complexity of the problem from an SOCP to an SDP, which we prefer to avoid (see Endnote 1). Second, there are important applications which motivate this assumption. For example, in the study of portfolio management, Chopra and Ziemba (1993) showed empirically that the impact of estimation errors in mean asset returns is about an order of magnitude more severe than the corresponding impact of estimation error in asset variances and covariances.

Acknowledgments

The authors thank the associate editor and two anonymous referees for their comments and critique of this paper. The suggestions have helped to improve the quality of this work. The research of the second author is supported by the Singapore-MIT Alliance and NUS academic research grant R-314-000-068-122.

References

- Artzner, P., F. Delbaen, J.-M. Eber, D. Heath. 1999. Coherent measures of risk. *Math. Finance* 9(3) 203–228.
- Ben-Tal, A., A. Nemirovski. 1998. Robust convex optimization. *Math. Oper. Res.* 23(4) 769–805.
- Ben-Tal, A., A. Nemirovski. 1999. Robust solutions of uncertain linear programs. *Oper. Res. Lett.* 25(1, August) 1–13.
- Ben-Tal, A., A. Nemirovski. 2000. Robust solutions of linear programming problems contaminated with uncertain data. *Math. Programming Ser. A* 88(3) 411–424.
- Ben-Tal, A., M. Teboulle. 1986. Expected utility, penalty functions, and duality in stochastic nonlinear programming. *Management Sci.* 32(11) 1445–1466.

- Ben-Tal, A., M. Teboulle. 2007. An old-new concept of convex risk measures: The optimized certainty equivalent. *Math. Finance* **17**(3) 449–476.
- Ben-Tal, A., A. Goryashko, E. Guslitzer, A. Nemirovski. 2004. Adjustable robust solutions of uncertain linear programs. *Math. Programming* **99** 351–376.
- Bertsimas, D., M. Sim. 2004. The price of robustness. *Oper. Res.* **52**(1) 35–53.
- Bertsimas, D., A. Thiele. 2006. A robust optimization approach to inventory theory. *Oper. Res.* **54**(1) 150–168.
- Birge, J. R., F. Louveaux. 1997. *Introduction to Stochastic Programming*. Springer, New York.
- Breton, M., S. El Hachem. 1995. Algorithms for the solution of stochastic dynamic minimax problems. *Comput. Optim. Appl.* **4** 317–345.
- Chen, W., M. Sim. 2009. Goal-driven optimization. *Oper. Res.* **57**(2) 342–357.
- Chen, X., Y. Zhang. 2009. Uncertain linear programs: Extended affinely adjustable robust counterparts. *Oper. Res.* **57**(6) 1469–1482.
- Chen, X., M. Sim, P. Sun. 2007. A robust optimization perspective on stochastic programming. *Oper. Res.* **55**(6) 1058–1071.
- Chen, X., M. Sim, P. Sun, J. Zhang. 2008. A linear decision-based approximation approach to stochastic programming. *Oper. Res.* **56**(2) 344–357.
- Chopra, V. K., W. T. Ziemba. 1993. The effects of errors in means, variances, and covariances on optimal portfolio choice. *J. Portfolio Management* **19**(2) 6–11.
- Delage, E., Y. Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Oper. Res.*, ePub ahead of print January 28, <http://or.journal.informs.org/cgi/content/abstract/opre/1090.0741v1>.
- Dupačová, J. 1987. The minimax approach to stochastic programming and an illustrative application. *Stochastics* **20**(1) 73–88.
- Dyer, M., L. Stougie. 2006. Computational complexity of stochastic programming problems. *Math. Programming Ser. A* **106** 423–432.
- El Ghaoui, L., H. Lebret. 1997. Robust solutions to least-squares problems with uncertain data. *SIAM J. Matrix Anal. Appl.* **18**(4) 1035–1064.
- El Ghaoui, L., M. Oks, F. Oustry. 2003. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Oper. Res.* **51**(4) 543–556.
- El Ghaoui, L., F. Oustry, H. Lebret. 1998. Robust solutions to uncertain semidefinite programs. *SIAM J. Optim.* **9** 33–52.
- Föllmer, H., A. Schied. 2002. Convex measures of risk and trading constraints. *Finance and Stochastics* **6**(4) 429–447.
- Gallego, G., I. Moon. 1993. The distribution free newsboy problem: Review and extensions. *J. Oper. Res. Soc.* **44**(8) 825–834.
- Goh, J., M. Sim. 2009. Robust optimization made easy with ROME. Working paper, National University of Singapore.
- Natarajan, K., D. Pachamanova, M. Sim. 2008. Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Sci.* **54**(3) 573–585.
- Rockafellar, R. T., S. Uryasev. 2000. Optimization of conditional value-at-risk. *J. Risk* **2** 493–517.
- Ruszczynski, A., A. Shapiro, eds. 2003. *Stochastic Programming. Handbooks in Operations Research and Management Science*, Vol. 10. Elsevier Science, Amsterdam.
- Scarf, H. 1958. A min-max solution of an inventory problem. K. Arrow, ed. *Studies in the Mathematical Theory of Inventory and Production*. Stanford University Press, Stanford, CA, 201–209.
- See, C.-T., M. Sim. 2009. Robust approximation to multiperiod inventory management. *Oper. Res.*, ePub ahead of print December 29, <http://or.journal.informs.org/cgi/content/abstract/opre/1090.0746v1>.
- Shapiro, A., A. Kleywegt. 2002. Minimax analysis of stochastic programs. *Optim. Methods and Software* **17**(3) 523–542.
- Shapiro, A., A. Nemirovski. 2005. On complexity of stochastic programming problems. V. Jeyakumar, A. Rubinov, eds. *Continuous Optimization*. Springer, New York, 111–146.
- Soyster, A. L. 1973. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Oper. Res.* **21**(5) 1154–1157.
- Wang, C., C.-J. Ong, M. Sim. 2008. Model predictive control using segregated disturbance feedback. *Proc. Amer. Control Conf.*, Seattle.
- Žáčková, J. 1966. On minimax solution of stochastic linear programming problems. *Časopis pro Pěstování Matematiky* **91** 423–430.