

# A Linear-Decision Based Approximation Approach to Stochastic Programming

Xin Chen\*, Melvyn Sim<sup>†</sup>, Peng Sun<sup>‡</sup> and Jiawei Zhang<sup>§</sup>

Feb 2006; Revised July 2006, February 2007.

## Abstract

Stochastic optimization, especially multistage models, is well known to be computationally ex-cruciating. Moreover, such models require exact specifications of the probability distributions of the underlying uncertainties, which are often unavailable. In this paper, we propose tractable methods of addressing a general class of multistage stochastic optimization problems, which assume only limited information of the distributions of the underlying uncertainties, such as known mean, support and covariance. One basic idea of our methods is to approximate the recourse decisions via decision rules. We first examine linear decision rules in detail and show that even for problems with complete recourse, linear decision rules can be inadequate and even lead to infeasible instances. Hence, we propose several new decision rules that improve upon linear decision rules, while keeping the approximate models computationally tractable. Specifically, our approximate models are in the forms of the so-called second order cone (SOC) programs, which could be solved efficiently both in theory and in practice. We also present computational evidence indicating that our approach is a viable alternative, and possibly advantageous, to existing stochastic optimization solution techniques in solving a two-stage stochastic optimization problem with complete recourse.

---

\*Department of Industrial and Enterprize Engineering Engineering, University of Illinois at Urbana-Champaign. Email: xinchen@uiuc.edu

<sup>†</sup>NUS Business School, National University of Singapore and Singapore MIT Alliance (SMA). Email: dcsimm@nus.edu.sg. The research of the author is supported by SMA, NUS Risk Management Institute, NUS academic research grants R-314-000-066-122 and R-314-000-068-122.

<sup>‡</sup>The Fuqua School of Business, Duke University Box 90120, Durham, NC 27708, USA. Email: psun@duke.edu

<sup>§</sup>Stern School of Business, New York University, New York, NY, 10012. Email: jzhang@stern.nyu.edu

# 1 Introduction

The study of stochastic programming dates back to Beale [2] and Dantzig [19]. In a typical two-stage stochastic program, decisions are made in the first stage in the face of uncertainty. Once the uncertainties are realized, the optimal second stage decisions or recourse decisions are carried out. Such “stochastic programs” attempt to integrate optimization with stochastic modelling that could potentially solve a large class of important practical problems, ranging from engineering control to supply chain management; see, e.g. Ruszczyński and Shapiro [29] and Birge and Louveaux [12]. Another class of stochastic optimization problems deals with chance constraints, which dates back to Charns and Cooper [16], and is the topic of a series of recent papers [13, 14, 15, 17, 20, 24, 25, 27, 28]. Despite the immense modeling potential, stochastic programs, especially multistage problems, are notoriously difficult to solve to optimality (see Shapiro and Nemirovski [31] and Dyer and Stougie [21]). Quite often, finding a feasible solution is already a hard problem. It is therefore important to develop a tractable and scalable methodology that could reasonably approximate stochastic programs.

One issue with stochastic optimization problems is the assumption of full distributional knowledge in each and every of the uncertain data. As such information may rarely be available in practice, it has rekindled recent interests in robust optimization as an alternative perspective of data uncertainty [3, 4, 5, 6, 8, 9, 10, 11, 22, 23, 26, 32]. Ben-Tal et al. [6] propose an adjustable robust counterpart to handle dynamic decision making under uncertainty, where the uncertainties addressed are non-stochastic. A different approach is suggested by Chen, Sim and Sun [17] for chance-constrained stochastic programs, which assumes only limited distributional information such as known mean, support, and some deviation measures of the random data. In both approaches, *linear decision rules* are the key enabling mechanism that permits scalability to multistage models. Interesting applications of such models include designing supplier-retailer flexible commitments contracts (Ben-Tal et al. [7]), network design under uncertainty (Atamturk and Zhang [1]), crashing projects with uncertain activity times (Chen, Sim and Sun [17]), and analyzing distribution systems with transshipment (Chou, Sim and So [18]).

Even though linear decision rules allow us to derive tractable formulations in a variety of applications, they may lead to infeasible instances even for problems with complete recourse. This fact motivates us to refine linear decision rules and develop a framework to approximate multistage stochastic optimization via the refined decision rules. Specifically, we propose two approaches to improve linear decision rules.

The first approach is called *deflected linear decision rules*. This new class of decisions rules is suitable for stochastic optimization problems with *semi-complete recourse* variables, a relaxation of complete recourse. The idea is to solve the stochastic programming model by ignoring certain constraints, while

appropriately penalizing the constraint violations in the objective function. Using linear decision rules and certain approximation of the new objective function (with the penalty term), the model is turned into a second order cone (SOC) program, which could be solved efficiently both in theory and in practice. The resulting linear decision implies a feasible decision rule for the original problem. Compared to using linear decision rules, such a deflected linear decision rule performs better in the sense that it provides a tighter approximation of the original objective function.

The second approach is called *segregated linear decision rules*, which are suitable for stochastic optimization problems with general recourse. The idea here is to introduce a decision rule that is a piece-wise linear function of the realizations of the primitive uncertainties. A segregated linear decision rule can be formulated in such a way that it can be combined with the first approach to generate a *segregated deflected linear decision rule* that can be proven to have better performance than both linear and deflected linear decision rules. One attractive aspect of our proposal is the scalability to multistage stochastic programs.

The structure of the paper is as follows. In Section 2, we introduce a general stochastic optimization model. Section 3 discusses several decision rules to approximate recourse decisions. In Section 4 we provide preliminary computational results. Finally, Section 5 concludes this paper.

**Notations** We denote a random variable,  $\tilde{x}$ , with the tilde sign. Bold face lower case letters such as  $\mathbf{x}$  represent vectors and the corresponding upper case letters such as  $\mathbf{A}$  denote matrices. In addition,  $x^+ = \max\{x, 0\}$  and  $x^- = \max\{-x, 0\}$ . The same operations can be used on vectors, such as  $\mathbf{y}^+$  and  $\mathbf{z}^-$  in which corresponding operations are performed componentwise.

## 2 A Stochastic Programming Model

A classical two-stage stochastic program with fixed recourse can be formulated as follows (see, e.g. Ruszczyński and Shapiro [29]):

$$\begin{aligned} \min \quad & \mathbf{c}'\mathbf{x} + \mathbb{E}(Q(\mathbf{x}, \tilde{\mathbf{z}})) \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0}, \end{aligned} \tag{1}$$

where

$$\begin{aligned} Q(\mathbf{x}, \mathbf{z}) = \min \quad & \mathbf{f}'\mathbf{w} \\ \text{s.t.} \quad & \mathbf{T}(\mathbf{z})\mathbf{x} + \mathbf{W}\mathbf{w} = \mathbf{h}(\mathbf{z}) \\ & w_i \geq 0 \quad \forall i \in I \subseteq \{1, \dots, n_2\} \end{aligned} \tag{2}$$

and  $\mathbf{c}$ ,  $\mathbf{f}$  and  $\mathbf{b}$  are known vectors in  $\mathfrak{R}^{n_1}$ ,  $\mathfrak{R}^{n_2}$  and  $\mathfrak{R}^{m_1}$  respectively. In this formulation,  $\tilde{\mathbf{z}} \in \mathfrak{R}^N$  is the vector of primitive uncertainties that consolidates all underlying uncertainties in the stochastic model, and  $\mathbf{E}$  is used to represent the expectation associated with the random variables  $\tilde{\mathbf{z}}$ . We assume the following affine data dependency for  $\mathbf{T}(\tilde{\mathbf{z}})$  and  $\mathbf{h}(\tilde{\mathbf{z}})$ ,

$$\mathbf{T}(\tilde{\mathbf{z}}) = \mathbf{T}^0 + \sum_{k=1}^N \mathbf{T}^k \tilde{z}_k, \quad \mathbf{h}(\tilde{\mathbf{z}}) = \mathbf{h}^0 + \sum_{k=1}^N \mathbf{h}^k \tilde{z}_k,$$

with  $\mathbf{T}^0, \mathbf{T}^1, \dots, \mathbf{T}^N \in \mathfrak{R}^{m_2 \times n_1}$  and  $\mathbf{h}^0, \mathbf{h}^1, \dots, \mathbf{h}^N \in \mathfrak{R}^{m_2}$ . Matrices  $\mathbf{A}$  and  $\mathbf{W}$  are known matrices in  $\mathfrak{R}^{m_1 \times n_1}$  and  $\mathfrak{R}^{m_2 \times n_2}$ , respectively. The stochastic model represents a sequence of events. Here vectors  $\mathbf{x}$  and  $\mathbf{w}$  are the first and the second stage decision variables, respectively. The first stage decision,  $\mathbf{x}$  has to be made before the actual value of  $\tilde{\mathbf{z}}$  is realized; after applying the decision  $\mathbf{x}$  and after the uncertainty is realized, the second stage decision (a.k.a recourse decision),  $\mathbf{w}$ , can be made. For a given  $(\mathbf{x}, \mathbf{z})$ , the second stage cost  $Q(\mathbf{x}, \mathbf{z})$  is set to be  $+\infty$  if the feasible set of (2) is empty, and  $-\infty$  if problem (2) is unbounded from below. It can be shown that (see, e.g., Ruszczyński and Shapiro (2003)), under very general conditions, problem (1) is equivalent to

$$\begin{aligned} Z_{STOC} = \quad & \min \quad \mathbf{c}'\mathbf{x} + \mathbf{E}(\mathbf{f}'\mathbf{w}(\tilde{\mathbf{z}})) \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{T}(\tilde{\mathbf{z}})\mathbf{x} + \mathbf{W}\mathbf{w}(\tilde{\mathbf{z}}) = \mathbf{h}(\tilde{\mathbf{z}}) \\ & w_i(\tilde{\mathbf{z}}) \geq 0 \quad \forall i \in I \\ & \mathbf{x} \geq 0 \\ & \mathbf{w}(\tilde{\mathbf{z}}) \in \mathcal{Y}, \end{aligned} \tag{3}$$

where  $\mathcal{Y}$  is a space of mappings from  $\mathfrak{R}^N$  to  $\mathfrak{R}^{n_2}$  that are measurable with respect to the probability space on which the random vector  $\tilde{\mathbf{z}}$  is defined. The functional  $\mathbf{w}(\cdot)$  is the vector of the second stage, or recourse, variables in response to the realization of  $\tilde{\mathbf{z}}$ .

There are several important special cases of recourses in the context of stochastic programming. In most problems we assume that the matrix  $\mathbf{W}$  is not subject to uncertainty. This is commonly referred to as problems with *fixed recourse*. The stochastic program (1) is said to have *relatively complete recourse* if for any  $\mathbf{x} \in \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0\}$ ,  $Q(\mathbf{x}, \tilde{\mathbf{z}}) < +\infty$  with probability one. In problems with relatively complete recourse, the second stage problem is almost surely feasible for any choice of feasible first stage decision vector,  $\mathbf{x}$ . It is generally not easy to identify conditions of relatively complete recourse (see Birge and Louveaux [12]). An important class of relatively complete recourse is known as *complete recourse*, which is defined on the matrix  $\mathbf{W}$  such that for any  $\mathbf{t}$ , there exists  $w_i \geq 0$ ,  $i \in I$  satisfying  $\mathbf{W}\mathbf{w} = \mathbf{t}$ . Hence, the definition of complete recourse depends only on the structure of the matrix

$\mathbf{W}$ , which makes it easier to verify. Moreover, many stochastic programming problems have complete recourse. A special case of complete recourse is simple recourse, where  $\mathbf{W} = [\mathbf{I}, -\mathbf{I}]$ .

The model (3) represents a rather general fixed recourse framework characterized in the classical stochastic optimization formulation. We note that it is widely accepted that a two-stage stochastic optimization problem with relatively complete recourse can be reasonably well approximated by random sampling approaches. In the absence of relatively complete recourse, however, the solution obtained from sampling approximations may not be meaningful. Even if the original problem is infeasible, the objective function value obtained from a sampling approximation could be still finite. This motivates us to consider solving stochastic programming problems using decision rules, which has the potential to address more general recourse problems and multi-period models.

In the rest of this paper, we do not assume knowledge of full distributional information of the primitive uncertainties. Rather, we assume that the primitive uncertainties  $\{\tilde{z}_j\}_{j=1:N}$  are zero mean random variables with covariance  $\Sigma$  and support  $\tilde{z} \in \mathcal{W} = [-\underline{z}, \bar{z}]$ , where some components of  $\underline{z}$  and  $\bar{z}$  could be infinite, reflecting unbounded supports.

### 3 Approximation via Decision Rules

Under the assumption that the stochastic parameters are independently distributed, Dyer and Stougie [21] show that two-stage stochastic programming problems are #P-hard. Under the same assumption they show that certain multi-stage stochastic programming problems are PSPACE-hard. Due to the astronomical number of possible scenarios, Monte Carlo sampling methods have been an important approximate solution approach to stochastic optimization problems. Despite the wide adoption of this approach, its performance has only been recently studied in theory, for instance, by Shapiro and Nemirovski [31]. They concluded that the number of samples required to approximate multistage stochastic programs to reasonable accuracy grows exponentially with the number of stages.

Another caveat with stochastic optimization models is the need to assume exact distributions for all the uncertain parameters in order to conduct random sampling. However, exact distributions may not be available in practice. In view of the hardness results, we propose a tractable approximation for model (3) by restricting the recourse decisions to specified decision rules. Ben-Tal et al. [7] use linear decision rules for adjustable robust counterpart, and Chen, Sim and Sun [17] use linear decision rules and report promising computational results for chance-constrained stochastic programs. In this section we introduce the notion of *semi-complete recourse* variables and propose more general decision rules that can tackle problems with semi-complete recourse. Specifically, in the following three subsections,

we first go over linear decision rules and point out their limitations. Then we introduce deflected linear decision rules and segregated linear decision rules, which extend linear decision rules. Finally, in the last subsection, we present the extension of our approach to multi-stage stochastic programming problems.

### 3.1 Linear decision rules

Using linear decision rules, we restrict recourse variables, say  $\mathbf{w}(\tilde{\mathbf{z}})$  to be affinely dependent on the primitive uncertainties. Of course, only in very rare occasions, linear decision rules are optimal. Indeed, the only motivation for linear decision rules is its tractability. As Shapiro and Nemirovski [31] put:

The only reason for restricting ourselves with affine (linear) decision rules stems from the desire to end up with a computationally tractable problem. We do not pretend that affine decision rules approximate well the optimal ones - whether it is so or not, it depends on the problem, and we usually have no possibility to understand how good in this respect a particular problem we should solve is. The rationale behind restricting to affine decision rules is the belief that in actual applications it is better to pose a modest and achievable goal rather than an ambitious goal which we do not know how to achieve.

We denote  $\mathcal{L}$  to be the space of linear functions. For instance,  $w(\cdot) \in \mathcal{L} \subseteq \mathcal{Y}$  implies that there exists a set of vectors  $\mathbf{w}^0, \dots, \mathbf{w}^N$  such that

$$\mathbf{w}(\tilde{\mathbf{z}}) = \mathbf{w}^0 + \sum_{k=1}^N \mathbf{w}^k \tilde{z}_k.$$

We can approximate the stochastic model (3) as follows:

$$\begin{aligned} Z_{LDR} = \quad & \min \quad \mathbf{c}'\mathbf{x} + \mathbf{f}'\mathbf{w}^0 \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{T}^k\mathbf{x} + \mathbf{W}\mathbf{w}^k = \mathbf{h}^k \quad \forall k \in \{0, \dots, N\} \\ & w_i(\mathbf{z}) \geq 0 \quad \forall \mathbf{z} \in \mathcal{W}, \forall i \in I \\ & \mathbf{x} \geq 0 \\ & \mathbf{w}(\cdot) \in \mathcal{L}. \end{aligned} \tag{4}$$

Since any feasible solution of model (4) is also feasible in (3), and the objectives coincide, we have  $Z_{STOC} \leq Z_{LDR}$ . With  $\mathcal{W} = [-\bar{\mathbf{z}}, \bar{\mathbf{z}}]$ , the semi-infinite constraint

$$w_i(\mathbf{z}) \geq 0 \quad \forall \mathbf{z} \in \mathcal{W}$$

is equivalent to

$$w_i^0 \geq \sum_{j=1}^N (\underline{z}_j s_j + \bar{z}_j t_j)$$

for some  $\mathbf{s}, \mathbf{t} \geq \mathbf{0}$  satisfying  $s_j - t_j = w_i^j$ ,  $j = 1, \dots, N$ . Hence, Model (4) is essentially a linear optimization problem.

Even though linear decision rules have been successfully used in a variety of applications (see, for instance, Ben-Tal et al. [7], and Chen, Sim and Sun [17]), they may perform poorly for some problem instances. As an illustration, suppose that the support of  $\tilde{\mathbf{z}}$  is  $\mathcal{W} = (-\infty, \infty)$ , then the following nonnegativity constraints

$$\mathbf{w}(\tilde{\mathbf{z}}) = \mathbf{w}^0 + \sum_{k=1}^N \mathbf{w}^k \tilde{z}_k \geq \mathbf{0}$$

imply that

$$\mathbf{w}^k = \mathbf{0} \quad \forall k \in \{1, \dots, N\},$$

and the decision rule is reduced to  $\mathbf{w}(\tilde{\mathbf{z}}) = \mathbf{w}^0$ , and hence independent of the primitive uncertainties. This may lead to an infeasible instance even in the case of complete recourse. For example, consider the following stochastic optimization model that determines  $\mathbb{E}(|b(\tilde{\mathbf{z}})|)$ :

$$\min \{ \mathbb{E}(w_1(\tilde{\mathbf{z}}) + w_2(\tilde{\mathbf{z}})) : w_1(\tilde{\mathbf{z}}) - w_2(\tilde{\mathbf{z}}) = b(\tilde{\mathbf{z}}), w_1(\tilde{\mathbf{z}}) \geq 0, w_2(\tilde{\mathbf{z}}) \geq 0 \}, \quad (5)$$

which is one with simple recourse. Suppose  $\tilde{\mathbf{z}}$  has infinite support, we must have  $w_1(\tilde{\mathbf{z}}) = w_1^0$  and  $w_2(\tilde{\mathbf{z}}) = w_2^0$ , and hence, it would be impossible to satisfy the equality constraint.

Our goal in the next subsections is to improve upon linear decision rules for variables with the semi-complete recourse property. As we shall see, the condition of semi-complete recourse implies that even if there is a constraint violation due to  $\mathbf{w}(\tilde{\mathbf{z}})$ , we can still steer back the solution towards feasibility by paying a finite price.

### 3.2 Deflected linear decision rule

In order to improve upon linear decision rule, we first explore the structure of the recourse matrix  $\mathbf{W}$  by considering the following linear optimization problem

$$\begin{aligned} \bar{f}_i &= \min \quad \mathbf{f}'\mathbf{p} \\ \text{s.t.} \quad & \mathbf{W}\mathbf{p} = 0 \\ & p_i = 1 \\ & p_j \geq 0 \quad j \in I, \end{aligned} \quad (6)$$

for each  $i \in I$ . If the corresponding optimization problem is infeasible, we have  $\bar{f}_i = \infty$ . For convenience, we define the following sets,

$$I_1 \triangleq \{\bar{f}_i < \infty : i \in I\} \quad I_2 \triangleq I \setminus I_1.$$

Whenever  $\bar{f}_i$  is finite, we denote  $\bar{\mathbf{p}}^i$  as the corresponding optimal solution. Moreover, if the original problem (3) is bounded from below, we have that  $\bar{f}_i \geq 0$ .

We now propose a new notion of *semi-complete recourse*.

**Definition 1** *Collect all the columns  $i \in I_1$  of the recourse matrix  $W$  to formulate a matrix  $W^s$ . We call  $W^s$  a semi-complete recourse matrix.*

We first show the relationship between semi-complete recourse and complete recourse.

**Proposition 1** *A matrix  $W$  that satisfies the condition of complete recourse is a semi-complete recourse matrix, i.e.,  $I_1 = I$ .*

**Proof :** Under complete recourse, it suffices to show that there exists a vector  $\mathbf{r}$  with  $r_i \geq 1$ , for all  $i \in I$ , such that  $W\mathbf{r} = \mathbf{0}$ . By definition of complete recourse, for any vector  $\mathbf{v}$  with  $v_i \geq 1$ , for all  $i \in I$ , we can find a vector  $\mathbf{s}$  with  $s_i \geq 0$  for all  $i \in I$  such that  $W\mathbf{s} = -W\mathbf{v}$ . Clearly,  $r_i = s_i + v_i \geq 1$ ,  $i \in I$  and  $W\mathbf{r} = \mathbf{0}$ . ■

Consider any decision rule  $\mathbf{r}(\tilde{\mathbf{z}})$  satisfying

$$\begin{aligned} T(\tilde{\mathbf{z}})\mathbf{x} + W\mathbf{r}(\tilde{\mathbf{z}}) &= \mathbf{h}(\tilde{\mathbf{z}}) \\ r_i(\tilde{\mathbf{z}}) &\geq 0 \quad i \in I_2, \end{aligned} \tag{7}$$

and not necessarily nonnegative for the semi-complete recourse variables,  $r_j(\tilde{\mathbf{z}})$ ,  $j \in I_1$ . We derive the following decision rule from  $\mathbf{r}(\tilde{\mathbf{z}})$

$$\mathbf{w}(\tilde{\mathbf{z}}) = \mathbf{r}(\tilde{\mathbf{z}}) + \sum_{i \in I_1} (r_i(\tilde{\mathbf{z}})^-)\bar{\mathbf{p}}^i, \tag{8}$$

where for a given scalar  $v$ ,  $v^- = \max(0, -v)$ . It can be easily verified that

$$\begin{aligned} w_i(\tilde{\mathbf{z}}) &\geq 0 \quad \forall i \in I \\ W\mathbf{w}(\tilde{\mathbf{z}}) &= W\mathbf{r}(\tilde{\mathbf{z}}). \end{aligned}$$

Therefore, for any given first stage decision  $\mathbf{x}$ , as long as there exists a second stage decision rule  $\mathbf{r}(\tilde{\mathbf{z}})$  satisfying (7), we can find a feasible decision rule (referred to as a *deflected decision rule*),  $\mathbf{w}(\tilde{\mathbf{z}})$ . We note that the feasibility of model (7) depends on the solution in the first stage  $\mathbf{x}$ .

For the case of complete recourse, we can obtain stronger results.

**Proposition 2** *If  $W$  is a complete recourse matrix, we have  $I_2 = \emptyset$  and that for any  $\mathbf{x}$ , there exists a linear decision rule  $\mathbf{r}(\cdot) \in \mathcal{L}$  such that*

$$\mathbf{T}(\tilde{\mathbf{z}})\mathbf{x} + \mathbf{W}\mathbf{r}(\tilde{\mathbf{z}}) = \mathbf{h}(\tilde{\mathbf{z}})$$

**Proof :** Let

$$\mathbf{r}(\tilde{\mathbf{z}}) = \mathbf{r}^0 + \sum_{k=1}^N \mathbf{r}^k \tilde{z}_k$$

By the assumption of complete recourse, there exist  $\mathbf{r}^0, \mathbf{r}^1, \dots, \mathbf{r}^N$  such that

$$\mathbf{T}^k \mathbf{x} + \mathbf{W}\mathbf{r}^k = \mathbf{h}^k \quad \forall k \in \{0, \dots, N\}.$$

This implies the desired result. ■

**Remark :** The above result implies that any stochastic programming problem with complete recourse is feasible with respect to deflected linear decision rules. The readers may ask why we introduce the notion of “semi-complete recourse”, under which the feasibility of deflected linear decisions is not guaranteed. The reason is that semi-complete recourse is easy to verify and deflected linear decision rules may still be feasible without complete recourse. And therefore one can try deflected linear decision rules as long as the problem contains semi-complete recourse submatrix. ■

From (8), we have

$$\mathbf{f}'\mathbf{w}(\tilde{\mathbf{z}}) = \mathbf{f}'\mathbf{r}(\tilde{\mathbf{z}}) + \sum_{i \in I_1} \bar{f}_i r_i(\tilde{\mathbf{z}})^-$$

Therefore, using the deflected linear decision rule  $\mathbf{w}(\cdot)$  based on linear decision rule  $\mathbf{r}(\cdot)$ , we can approximate problem (3) as

$$\begin{aligned} Z_{DLDR} = \min \quad & \mathbf{c}'\mathbf{x} + \mathbf{f}'\mathbf{r}^0 + \sum_{i \in I_1} \bar{f}_i \mathbb{E} [r_i(\tilde{\mathbf{z}})^-] \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{T}^k \mathbf{x} + \mathbf{W}\mathbf{r}^k = \mathbf{h}^k \quad \forall k \in \{0, \dots, N\} \\ & r_j(\tilde{\mathbf{z}}) \geq 0 \quad \forall j \in I_2 \\ & \mathbf{x} \geq 0 \\ & \mathbf{r}(\cdot) \in \mathcal{L}. \end{aligned} \tag{9}$$

Notice that in formulating and solving the above model, we do not directly need  $\bar{\mathbf{p}}^i$  defined in (6). In fact, what we really need is just  $\bar{f}_i, i \in I_1$ .

Since any feasible solution  $(\mathbf{x}, \mathbf{w}(\tilde{\mathbf{z}}))$  to model (9), in which  $\mathbf{w}(\tilde{\mathbf{z}}) = \mathbf{r}(\tilde{\mathbf{z}}) + \sum_{i \in I_1} (\mathbf{r}_i(\tilde{\mathbf{z}}))^- \bar{\mathbf{p}}^i$ , is feasible in (3), and the objectives coincide, we have  $Z_{STOC} \leq Z_{DLDR}$ . Moreover, given any feasible solution,  $(\mathbf{x}, \mathbf{w}(\tilde{\mathbf{z}}))$ , of problem (4), we observe that

$$\mathbb{E} [\bar{f}_i w_i(\tilde{\mathbf{z}})^-] = 0 \quad \forall i \in I_1.$$

Hence by letting  $\mathbf{r}(\tilde{\mathbf{z}}) = \mathbf{w}(\tilde{\mathbf{z}})$ , we obtain a feasible solution of Model (9) with the same objective. Therefore,  $Z_{STOC} \leq Z_{DLDR} \leq Z_{LDR}$ .

Unfortunately, model (9) is still hard to solve because of the nonlinear term  $\mathbb{E}[\bar{f}_i w_i(\tilde{\mathbf{z}})^-]$  in the objective function. In the sequel, we approximate (9) via techniques from robust optimization so that the resulting model is in the form of second-order cone programming, which can be solved efficiently both in theory and in practice.

### 3.2.1 Bound on objective function

Given a random variable  $\tilde{r}$  with mean  $\mu$  and standard deviation,  $\sigma$ , the following tight bound on  $\mathbb{E}(\tilde{r}^-)$  is well known,

$$\mathbb{E}(\tilde{r}^-) \leq \frac{1}{2} \left( -\mu + \sqrt{\mu^2 + \sigma^2} \right) \quad (10)$$

(see Scarf [30] for instance). Therefore, suppose  $y(\tilde{\mathbf{z}}) = y_0 + \mathbf{y}'\tilde{\mathbf{z}} \in \mathcal{L}$  with  $\mathbf{y} = (y_1, \dots, y_N)$ , we have

$$\mathbb{E}(y(\tilde{\mathbf{z}})^-) \leq \frac{1}{2} \left( -y_0 + \sqrt{y_0^2 + \|\boldsymbol{\Sigma}^{1/2}\mathbf{y}\|_2^2} \right),$$

where  $\boldsymbol{\Sigma}$  is the covariance matrix of  $\tilde{\mathbf{z}}$ . The bound does not take into account the distributional support, which could degrade the quality of the approximation. For instance, if  $y(\tilde{\mathbf{z}}) \geq 0$ , it follows trivially that  $\mathbb{E}(y(\tilde{\mathbf{z}})^-) = 0$ . Likewise, if  $y(\tilde{\mathbf{z}}) \leq 0$ , we have  $\mathbb{E}(y(\tilde{\mathbf{z}})^-) = -y_0$ . Under these circumstances, the bound would be weak. Hence, we propose the following tighter bound that resolves these issues while still preserving the benefits of being second order cone representable.

**Theorem 1** *Let  $\tilde{\mathbf{z}} \in \mathbb{R}^N$  be a vector of zero mean random variables with covariance matrix  $\boldsymbol{\Sigma}$  and support in  $\mathcal{W} = [-\underline{\mathbf{z}}, \bar{\mathbf{z}}]$ .*

(a)

$$\mathbb{E}((y_0 + \mathbf{y}'\tilde{\mathbf{z}})^-) \leq h(y_0, \mathbf{y})$$

where

$$h(y_0, \mathbf{y}) \triangleq \min_{\mathbf{s}, \mathbf{t}, \mathbf{u}, \mathbf{v} \geq \mathbf{0}} \left\{ \frac{1}{2} \left( -y_0 + (\mathbf{s} + \mathbf{u})'\bar{\mathbf{z}} + (\mathbf{t} + \mathbf{v})'\underline{\mathbf{z}} + \sqrt{(-y_0 + (\mathbf{s} - \mathbf{u})'\bar{\mathbf{z}} + (\mathbf{t} - \mathbf{v})'\underline{\mathbf{z}})^2 + \|\boldsymbol{\Sigma}^{1/2}(-\mathbf{y} - \mathbf{s} + \mathbf{t} + \mathbf{u} - \mathbf{v})\|_2^2} \right) \right\}. \quad (11)$$

(b) *Moreover,*

$$h(y_0, \mathbf{y}) \leq \frac{1}{2} \left( -y_0 + \sqrt{y_0^2 + \|\boldsymbol{\Sigma}^{1/2}\mathbf{y}\|_2^2} \right).$$

(c) *Suppose  $y(\mathbf{z}) \leq 0$ ,  $\forall \mathbf{z} \in \mathcal{W}$ , then  $\mathbb{E}((y_0 + \mathbf{y}'\tilde{\mathbf{z}})^-) = h(y_0, \mathbf{y}) = -y_0$ .*

*Likewise, if  $y(\mathbf{z}) \geq 0$ ,  $\forall \mathbf{z} \in \mathcal{W}$ , then  $\mathbb{E}((y_0 + \mathbf{y}'\tilde{\mathbf{z}})^-) = h(y_0, \mathbf{y}) = 0$ .*

**Proof :** (a) Since  $-\underline{\mathbf{z}} \leq \tilde{\mathbf{z}} \leq \bar{\mathbf{z}}$ , we observe that

$$(\bar{\mathbf{z}} - \tilde{\mathbf{z}})' \mathbf{s} \geq 0, \quad (\underline{\mathbf{z}} + \tilde{\mathbf{z}})' \mathbf{t} \geq 0, \quad (\bar{\mathbf{z}} - \tilde{\mathbf{z}})' \mathbf{u} \geq 0, \quad (\underline{\mathbf{z}} + \tilde{\mathbf{z}})' \mathbf{v} \geq 0.$$

for all  $\mathbf{s}, \mathbf{t}, \mathbf{u}, \mathbf{v} \geq \mathbf{0}$ . Therefore,

$$\begin{aligned} & \mathbb{E}((y_0 + \mathbf{y}' \tilde{\mathbf{z}})^-) \\ & \leq \mathbb{E}((y_0 + \mathbf{y}' \tilde{\mathbf{z}} - (\underline{\mathbf{z}} + \tilde{\mathbf{z}})' \mathbf{t} - (\bar{\mathbf{z}} - \tilde{\mathbf{z}})' \mathbf{s})^-) \\ & = \mathbb{E}((y_0 - (-\mathbf{y} + \mathbf{t} - \mathbf{s})' \tilde{\mathbf{z}} - \underline{\mathbf{z}}' \mathbf{t} - \bar{\mathbf{z}}' \mathbf{s})^-) \\ & = \mathbb{E}(-y_0 + (-\mathbf{y} + \mathbf{t} - \mathbf{s})' \tilde{\mathbf{z}} + \underline{\mathbf{z}}' \mathbf{t} + \bar{\mathbf{z}}' \mathbf{s} + (y_0 - (-\mathbf{y} + \mathbf{t} - \mathbf{s})' \tilde{\mathbf{z}} - \underline{\mathbf{z}}' \mathbf{t} - \bar{\mathbf{z}}' \mathbf{s})^+) \end{aligned} \quad (12)$$

$$\begin{aligned} & = \mathbb{E}(-y_0 + \underline{\mathbf{z}}' \mathbf{t} + \bar{\mathbf{z}}' \mathbf{s} + (y_0 - (-\mathbf{y} + \mathbf{t} - \mathbf{s})' \tilde{\mathbf{z}} - \underline{\mathbf{z}}' \mathbf{t} - \bar{\mathbf{z}}' \mathbf{s})^+) \\ & \leq \mathbb{E}(-y_0 + \underline{\mathbf{z}}' \mathbf{t} + \bar{\mathbf{z}}' \mathbf{s} + (y_0 - (-\mathbf{y} + \mathbf{t} - \mathbf{s})' \tilde{\mathbf{z}} - \underline{\mathbf{z}}' \mathbf{t} - \bar{\mathbf{z}}' \mathbf{s} + (\underline{\mathbf{z}} + \tilde{\mathbf{z}})' \mathbf{v} + (\bar{\mathbf{z}} - \tilde{\mathbf{z}})' \mathbf{u})^+) \\ & = \mathbb{E}(-y_0 + \underline{\mathbf{z}}' \mathbf{t} + \bar{\mathbf{z}}' \mathbf{s} + (y_0 - (-\mathbf{y} + \mathbf{t} - \mathbf{s} + \mathbf{u} - \mathbf{v})' \tilde{\mathbf{z}} - \underline{\mathbf{z}}' (\mathbf{t} - \mathbf{v}) - \bar{\mathbf{z}}' (\mathbf{s} - \mathbf{u}))^+) \\ & = \underline{\mathbf{z}}' \mathbf{v} + \bar{\mathbf{z}}' \mathbf{u} + \mathbb{E}((y_0 - (-\mathbf{y} + \mathbf{t} - \mathbf{s} + \mathbf{u} - \mathbf{v})' \tilde{\mathbf{z}} - \underline{\mathbf{z}}' (\mathbf{t} - \mathbf{v}) - \bar{\mathbf{z}}' (\mathbf{s} - \mathbf{u}))^-) \end{aligned} \quad (13)$$

$$\begin{aligned} & \leq \underline{\mathbf{z}}' \mathbf{v} + \bar{\mathbf{z}}' \mathbf{u} + \frac{1}{2} \left( -y_0 + \underline{\mathbf{z}}' (\mathbf{t} - \mathbf{v}) + \bar{\mathbf{z}}' (\mathbf{s} - \mathbf{u}) \right) + \\ & \quad \sqrt{(-y_0 + \underline{\mathbf{z}}' (\mathbf{t} - \mathbf{v}) + \bar{\mathbf{z}}' (\mathbf{s} - \mathbf{u}))^2 + \|\boldsymbol{\Sigma}^{1/2}(-\mathbf{y} + \mathbf{t} - \mathbf{s} + \mathbf{u} - \mathbf{v})\|_2^2} \end{aligned} \quad (14)$$

$$\begin{aligned} & = \frac{1}{2} \left( -y_0 + \underline{\mathbf{z}}' (\mathbf{t} + \mathbf{v}) + \bar{\mathbf{z}}' (\mathbf{s} + \mathbf{u}) \right) + \\ & \quad \sqrt{(-y_0 + \underline{\mathbf{z}}' (\mathbf{t} - \mathbf{v}) + \bar{\mathbf{z}}' (\mathbf{s} - \mathbf{u}))^2 + \|\boldsymbol{\Sigma}^{1/2}(-\mathbf{y} + \mathbf{t} - \mathbf{s} + \mathbf{u} - \mathbf{v})\|_2^2} \end{aligned}$$

where the equalities of (12) and (13) follow from the fact that  $x = x^+ - x^-$ . The inequality (14) is due to the bound (10).

(b) Note that with  $\mathbf{s}, \mathbf{t}, \mathbf{u}, \mathbf{v} = \mathbf{0}$ , we have

$$\begin{aligned} & \frac{1}{2} \left( -y_0 + (\mathbf{s} + \mathbf{u})' \bar{\mathbf{z}} + (\mathbf{t} + \mathbf{v})' \underline{\mathbf{z}} + \sqrt{(-y_0 + (\mathbf{s} - \mathbf{u})' \bar{\mathbf{z}} + (\mathbf{t} - \mathbf{v})' \underline{\mathbf{z}})^2 + \|\boldsymbol{\Sigma}^{1/2}(-\mathbf{y} - \mathbf{s} + \mathbf{t} + \mathbf{u} - \mathbf{v})\|_2^2} \right) \\ & \quad = \frac{1}{2} \left( -y_0 + \sqrt{y_0^2 + \|\boldsymbol{\Sigma}^{1/2} \mathbf{y}\|_2^2} \right) \end{aligned}$$

Therefore,

$$h(y_0, \mathbf{y}) \leq \frac{1}{2} \left( -y_0 + \sqrt{y_0^2 + \|\boldsymbol{\Sigma}^{1/2} \mathbf{y}\|_2^2} \right).$$

(c) Suppose

$$y_0 + \mathbf{y}' \mathbf{z} \leq 0 \quad \forall \mathbf{z} \in \mathcal{W}$$

then let  $\mathbf{s} = \mathbf{t} = \mathbf{0}$ ,  $u_k = (y_k)^+$ ,  $v_k = (-y_k)^+$  for  $k = 1, \dots, N$  and

$$z_k^* = \begin{cases} \bar{z}_k & \text{if } y_k > 0 \\ -\underline{z}_k & \text{otherwise} \end{cases}$$

Since  $\mathbf{z}^* \in \mathcal{W}$ , we have  $y_0 + \mathbf{y}'\mathbf{z}^* \leq 0$ . Furthermore, it is easy to verify that

$$\mathbf{y} = \mathbf{u} - \mathbf{v} \quad \text{and} \quad y_0 + \mathbf{u}'\bar{\mathbf{z}} + \mathbf{v}'\underline{\mathbf{z}} = y_0 + \mathbf{y}'\mathbf{z}^* \leq 0.$$

We have

$$\frac{1}{2} \left( -y_0 + (\mathbf{s} + \mathbf{u})'\bar{\mathbf{z}} + (\mathbf{t} + \mathbf{v})'\underline{\mathbf{z}} + \sqrt{(-y_0 + (\mathbf{s} - \mathbf{u})'\bar{\mathbf{z}} + (\mathbf{t} - \mathbf{v})'\underline{\mathbf{z}})^2 + \|\boldsymbol{\Sigma}^{1/2}(-\mathbf{y} - \mathbf{s} + \mathbf{t} + \mathbf{u} - \mathbf{v})\|_2^2} \right) = -y_0.$$

Hence,

$$-y_0 = \mathbb{E}((y_0 + \mathbf{y}'\bar{\mathbf{z}})^-) \leq h(y_0, \mathbf{y}) \leq -y_0.$$

Similarly, if  $y_0 + \mathbf{y}'\mathbf{z} \geq 0$ ,  $\forall \mathbf{z} \in \mathcal{W}$  then let  $\mathbf{v} = \mathbf{u} = \mathbf{0}$ ,  $s_k = (-y_k)^+$ ,  $t_k = (y_k)^+$  for  $k = 1, \dots, N$  and

$$z_k^* = \begin{cases} \bar{z}_k & \text{if } y_k < 0 \\ -\underline{z}_k & \text{otherwise} \end{cases}$$

Since  $\mathbf{z}^* \in \mathcal{W}$ , we have  $y_0 + \mathbf{y}'\mathbf{z}^* \geq 0$ . Furthermore, it is easy to verify that

$$\mathbf{y} = \mathbf{t} - \mathbf{s} \quad \text{and} \quad y_0 - \mathbf{s}'\bar{\mathbf{z}} - \mathbf{t}'\underline{\mathbf{z}} = y_0 + \mathbf{y}'\mathbf{z}^* \geq 0.$$

Hence, we have

$$\begin{aligned} & \frac{1}{2} \left( -y_0 + (\mathbf{s} + \mathbf{u})'\bar{\mathbf{z}} + (\mathbf{t} + \mathbf{v})'\underline{\mathbf{z}} + \sqrt{(-y_0 + (\mathbf{s} - \mathbf{u})'\bar{\mathbf{z}} + (\mathbf{t} - \mathbf{v})'\underline{\mathbf{z}})^2 + \|\boldsymbol{\Sigma}^{1/2}(-\mathbf{y} - \mathbf{s} + \mathbf{t} + \mathbf{u} - \mathbf{v})\|_2^2} \right) \\ &= \frac{1}{2} \left( -y_0 + \mathbf{s}'\bar{\mathbf{z}} + \mathbf{t}'\underline{\mathbf{z}} + | -y_0 + \mathbf{s}'\bar{\mathbf{z}} + \mathbf{t}'\underline{\mathbf{z}} | \right) \\ &= 0. \end{aligned}$$

Therefore,  $0 = \mathbb{E}((y_0 + \mathbf{y}'\bar{\mathbf{z}})^-) \leq h(y_0, \mathbf{y}) \leq 0$ . ■

### 3.2.2 Second Order Cone approximation with deflected linear decision rule

Summarizing the discussions in previous subsections, we propose the following approximation for problem (3).

$$\begin{aligned}
\bar{Z}_{DLDR} = \min \quad & \mathbf{c}'\mathbf{x} + \mathbf{f}'\mathbf{r}^0 + \sum_{i \in I_1} \bar{f}_i g_i \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\
& \mathbf{T}^k \mathbf{x} + \mathbf{W}\mathbf{r}^k = \mathbf{h}^k \quad \forall k \in \{0, \dots, N\} \\
& r_j(\mathbf{z}) \geq 0 \quad \forall \mathbf{z} \in \mathcal{W}, \forall j \in I_2 \\
& g_i \geq h(r_i^0, (r_i^1, \dots, r_i^N)) \quad \forall i \in I_1 \\
& \mathbf{x} \geq 0 \\
& \mathbf{r}(\cdot) \in \mathcal{L}.
\end{aligned} \tag{15}$$

More explicitly, we have the following second order cone optimization problem

$$\begin{aligned}
\bar{Z}_{DLDR} = \min \quad & \mathbf{c}'\mathbf{x} + \mathbf{f}'\mathbf{r}^0 + \sum_{i \in I_1} \bar{f}_i g_i \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\
& \mathbf{T}^k \mathbf{x} + \mathbf{W}\mathbf{r}^k = \mathbf{h}^k \quad \forall k \in \{0, \dots, N\} \\
& r_j^0 \geq \mathbf{t}^{j'} \bar{\mathbf{z}} + \mathbf{s}^{j'} \underline{\mathbf{z}} \quad \forall j \in I_2 \\
& \mathbf{t}^j - \mathbf{s}^j = \mathbf{r}_j \quad \forall j \in I_2 \\
& \mathbf{t}^j, \mathbf{s}^j \geq \mathbf{0} \quad \forall j \in I_2 \\
& 2g_i + r_i^0 - (\mathbf{s}^i + \mathbf{u}^i)' \bar{\mathbf{z}} - (\mathbf{t}^i + \mathbf{v}^i)' \underline{\mathbf{z}} \geq \\
& \quad \left( (-r_i^0 + (\mathbf{s}^i - \mathbf{u}^i)' \bar{\mathbf{z}} + (\mathbf{t}^i - \mathbf{v}^i)' \underline{\mathbf{z}})^2 + \right. \\
& \quad \left. \|\Sigma^{1/2}(-\mathbf{r}_i - \mathbf{s}^i + \mathbf{t}^i + \mathbf{u}^i - \mathbf{v}^i)\|_2^2 \right)^{\frac{1}{2}} \quad \forall i \in I_1 \\
& \mathbf{s}^i, \mathbf{t}^i, \mathbf{u}^i, \mathbf{v}^i \geq \mathbf{0} \quad \forall i \in I_1 \\
& \mathbf{x} \geq \mathbf{0},
\end{aligned} \tag{16}$$

where  $\mathbf{r}_i = (r_i^1, \dots, r_i^N)$ . Furthermore, we can show that the deflected linear decision rule improves the linear decision rule.

**Theorem 2** *Deflected linear decision rule gives an objective value that is at least as good as linear decision rule, that is,  $\bar{Z}_{DLDR} \leq Z_{LDR}$ .*

**Proof :** Given any feasible solution,  $(\mathbf{x}, \mathbf{w}(\tilde{\mathbf{z}}))$  of problem (4), we observe that for all  $i \in I$

$$w_i(\mathbf{z}) \geq 0 \quad \forall \mathbf{z} \in \mathcal{W}$$

Hence, from Theorem 1 (c), we have

$$h(w_i^0, (w_i^1, \dots, w_i^N)) = 0.$$

Therefore, by letting  $\mathbf{r}(\tilde{z}) = \mathbf{w}(\tilde{z})$ , we obtain a feasible solution of model (15) that yields the same objective as problem (4). Hence,  $\bar{Z}_{DLDR} \leq Z_{LDR}$ . ■

Next we provide a simple illustration of the modeling steps.

### A Newsvendor example

We can model the classical newsvendor problem as a stochastic programming problem (see, for instance, Birge and Louveaux [12]). A single retailer faces random demand. The retailer places an order to an outside supplier before knowing the actual demand. Per unit ordering cost is  $c$  and the selling price to the customers is  $p > c$ . Let the demand,  $d(\tilde{z}) = \mu + \tilde{z}$ , in which  $\tilde{z}$  has zero mean, standard deviation  $\sigma$  and unknown support. Such modeling ambiguity was first studied by Scarf [30]. For simplicity of exposition, we assume that unsatisfied demand is lost and leftover inventory has zero salvage value. Finally, we assume that the retailer's objective is to maximize the expected profit (or minimize expected negative profit). Let  $x$  denote the ordering quantity and  $\tilde{d}$  denote the random demand. The stochastic optimization formulation of the newsvendor model is as follows:

$$\begin{array}{ll} \min & cx + p\mathbf{E}(w(\tilde{z})) \\ \text{s.t.} & w(\tilde{z}) \geq -x \\ & w(\tilde{z}) \geq -d(\tilde{z}) \quad \text{or equivalently,} \\ & x \geq 0 \\ & w(\cdot) \in \mathcal{Y} \end{array} \qquad \begin{array}{ll} \min & cx + p\mathbf{E}(w_3(\tilde{z})) \\ \text{s.t.} & x + w_3(\tilde{z}) - w_1(\tilde{z}) = 0 \\ & w_3(\tilde{z}) - w_2(\tilde{z}) = -d(\tilde{z}) \\ & w_1(\tilde{z}), w_2(\tilde{z}) \geq 0 \\ & x \geq 0 \\ & w_1(\cdot), w_2(\cdot), w_3(\cdot) \in \mathcal{Y}, \end{array}$$

where  $-w_3(\tilde{z})$  is the amount of demand satisfied from the inventory,  $w_1(\tilde{z})$  is the inventory left and  $w_2(\tilde{z})$  is the unsatisfied demand. It is obvious that the associated recourse matrix

$$\mathbf{W} = \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix},$$

satisfies semi-complete recourse and that  $\bar{f}_1 = \bar{f}_2 = p$ . Using the approach of model (15), we solve the following problem

$$\begin{aligned}
\min \quad & cx + p(w_3^0 + g_1 + g_2) \\
\text{s.t.} \quad & x + w_3^0 - r_1^0 = 0 \\
& w_3^1 - r_1^1 = 0 \\
& w_3^0 - r_2^0 = -\mu \\
& w_3^1 - r_2^1 = -1 \\
& 2g_1 \geq -r_1^0 + \sqrt{(r_1^0)^2 + (\sigma r_1^1)^2} \\
& 2g_2 \geq -r_2^0 + \sqrt{(r_2^0)^2 + (\sigma r_2^1)^2} \\
& x \geq 0.
\end{aligned}$$

We will also show that the solution is identical to the famous result of Scarf [30]. After simplification, we have

$$\begin{aligned}
\min \quad & cx + \frac{1}{2}p \left( (-x - \mu) + \sqrt{(x + w_3^0)^2 + (\sigma w_3^1)^2} + \sqrt{(\mu + w_3^0)^2 + (\sigma(w_3^1 + 1))^2} \right) \\
\text{s.t.} \quad & x \geq 0.
\end{aligned}$$

By triangle inequality, we have

$$\sqrt{(x + w_3^0)^2 + (\sigma w_3^1)^2} + \sqrt{(\mu + w_3^0)^2 + (\sigma(w_3^1 + 1))^2} \geq \sqrt{(x - \mu)^2 + \sigma^2},$$

and the bound is achieved with  $w_3^0 = -\mu$  and  $w_3^1 = -1$ . Therefore, we end up with

$$\begin{aligned}
\min \quad & cx + \frac{1}{2}p \left( (-x - \mu) + \sqrt{(x - \mu)^2 + \sigma^2} \right) \\
\text{s.t.} \quad & x \geq 0,
\end{aligned}$$

which yields the same min-max solution of the newsvendor problem studied by Scarf [30].

### 3.3 Segregated Linear Decision Rule

One drawback on deflected linear decision rules is that it only works with semi-complete recourse decisions. In this section we provide another improvement to the linear decision rules that applies to general recourse problems.

We first define the segregated primitive uncertainties by dissecting each primitive uncertainty into its positive and negative components as follows:

$$\begin{aligned}
\tilde{z}_j^1 &\triangleq \tilde{z}_j^+ - \hat{z}_j \\
\tilde{z}_j^2 &\triangleq -\tilde{z}_j^- + \hat{z}_j
\end{aligned} \tag{17}$$

where  $\hat{z}_j = \mathbb{E}(\tilde{z}_j^+)$ . Since the primitive uncertainties have zero means, we have

$$\mathbb{E}(\tilde{z}_j^+) = \mathbb{E}(\tilde{z}_j^-),$$

and hence,  $\mathbb{E}(\tilde{z}_j^+) = \mathbb{E}(\tilde{z}_j^-) = 0$ . We denote  $\mathcal{S}$  as the space of segregated linear decision functions. That is,  $\mathbf{w}(\cdot) \in \mathcal{S} \subseteq \mathcal{Y}$  implies that there exists a set of vectors  $\mathbf{w}^0, \mathbf{w}^{11}, \dots, \mathbf{w}^{1N}$  and  $\mathbf{w}^{21}, \dots, \mathbf{w}^{2N}$  such that

$$\mathbf{w}(\tilde{\mathbf{z}}^1, \tilde{\mathbf{z}}^2) = \mathbf{w}^0 + \sum_{j=1}^N (\mathbf{w}^{1j} \tilde{z}_j^1 + \mathbf{w}^{2j} \tilde{z}_j^2).$$

When  $w^{1j} = w^{2j}$ , the decision rule  $\mathbf{w}(\tilde{\mathbf{z}}^1, \tilde{\mathbf{z}}^2)$  is a linear function of  $\tilde{\mathbf{z}}$ . Therefore  $\mathcal{L} \subseteq \mathcal{S}$ . Essentially, the segregated linear decision rule is a combination of linear decision rules on each of the segregated primitive uncertainties.

As we have pointed out, when the supports of  $\tilde{z}_j$  are unbounded both above and below, linear decision rules may be infeasible, or the feasible linear decision rule is trivial (a constant function). The next result indicates that in certain cases, segregated linear decision rules overcome the drawback of linear decision rules.

**Proposition 3** *If for a given  $\mathbf{x}$ , the following system*

$$\begin{aligned} \mathbf{T}(\mathbf{z})\mathbf{x} + \mathbf{W}\mathbf{w}(\mathbf{z}) &= \mathbf{h}(\mathbf{z}) \\ \mathbf{w}(\mathbf{z}) &\geq \mathbf{0} \end{aligned} \tag{18}$$

*has a feasible solution  $\mathbf{w}(\mathbf{z})$  for all  $\mathbf{z} \in \mathfrak{R}^N$ , then there exists a segregated linear decision rule which is feasible for the system (18).*

**Proof :** First, for any feasible decision rule, since

$$\mathbf{W}\mathbf{w}(\mathbf{z}) = \mathbf{h}^0 - \mathbf{T}^0\mathbf{x} + \sum_{j=1}^N (\mathbf{h}^j - \mathbf{T}^j\mathbf{x})z_j,$$

there exists  $\mathbf{w}^0 = \mathbf{w}(\mathbf{0}) \geq \mathbf{0}$  such that  $\mathbf{W}\mathbf{w}^0 = \mathbf{h}^0 - \mathbf{T}^0\mathbf{x}$ .

Next we focus on segregated linear decision rules of the following form:  $\mathbf{w}(\mathbf{z}) = \mathbf{w}^0 + \sum_{j=1}^N (\mathbf{u}^j z_j^+ + \mathbf{v}^j z_j^-)$ . Note that  $\mathbf{w}(\mathbf{z}) \geq \mathbf{0}$  for all  $\mathbf{z} \in \mathfrak{R}^N$  iff  $\mathbf{u}^j, \mathbf{v}^j \geq \mathbf{0}$ . We now show by contradiction that there exist  $\mathbf{u}^j, \mathbf{v}^j \geq \mathbf{0}$  such that

$$\mathbf{W}\mathbf{u}^j = -\mathbf{W}\mathbf{v}^j = \mathbf{h}^j - \mathbf{T}^j\mathbf{x}.$$

If there does not exist  $\mathbf{u}_j \geq \mathbf{0}$  such that  $\mathbf{W}\mathbf{u}^j = \mathbf{h}^j - \mathbf{T}^j\mathbf{x}$ , then from linear program strong duality theory, there exists a  $\mathbf{y}$  such that  $\mathbf{W}'\mathbf{y} \leq \mathbf{0}$  and  $(\mathbf{h}^j - \mathbf{T}^j\mathbf{x})'\mathbf{y} > 0$ . Notice that for any realization of  $\mathbf{z} = t\mathbf{e}_j$ , where  $\mathbf{e}_j$  is a unit vector at the  $j$ th row, there exists  $\hat{\mathbf{w}}(t) \geq \mathbf{0}$  such that

$$\mathbf{W} \frac{\hat{\mathbf{w}}(t)}{t} = \mathbf{h}^j - \mathbf{T}^j\mathbf{x} + \frac{1}{t}(\mathbf{h}^0 - \mathbf{T}^0\mathbf{x}).$$

However, although there exists a  $\mathbf{y}$  such that  $\mathbf{W}'\mathbf{y} \leq 0$  and  $(\mathbf{h}^j - \mathbf{T}^j\mathbf{x})'\mathbf{y} > 0$ , we also have

$$(\mathbf{h}^j - \mathbf{T}^j\mathbf{x})'\mathbf{y} = \mathbf{y}'\mathbf{W}\frac{\hat{\mathbf{w}}(t)}{t} - \frac{1}{t}\mathbf{y}'(\mathbf{h}^0 - \mathbf{T}^0\mathbf{x}) = \lim_{t \rightarrow \infty} \mathbf{y}'\mathbf{W}\frac{\hat{\mathbf{w}}(t)}{t} \leq 0,$$

which is a contradiction. Hence, there exists  $\mathbf{u}_j \geq \mathbf{0}$  such that  $\mathbf{W}\mathbf{u}^j = \mathbf{h}^j - \mathbf{T}^j\mathbf{x}$ . Similarly, there exists  $\mathbf{v}^j \geq \mathbf{0}$  such that  $\mathbf{W}\mathbf{v}^j = -(\mathbf{h}^j - \mathbf{T}^j\mathbf{x})$ . Thus, the segregated linear decision rule

$$\begin{aligned} \mathbf{w}(\mathbf{z}) &= \mathbf{w}^0 + \sum_{j=1}^N (\mathbf{u}^j z_j^+ + \mathbf{v}^j z_j^-) \\ &= \mathbf{w}^0 + \sum_{j=1}^N (\mathbf{u}^j + \mathbf{v}^j) \hat{z}_j + \sum_{j=1}^N (\mathbf{u}^j z_j^1 - \mathbf{v}^j z_j^2) \end{aligned}$$

is a feasible solution for the system (18). ■

**Model of Segregated Primitive Uncertainty,  $U_2$ :** We assume that the mean and mutual covariance of  $(\tilde{\mathbf{z}}^+, \tilde{\mathbf{z}}^-)$  are known. Equivalently, the segregated primitive uncertainties  $(\tilde{\mathbf{z}}^1, \tilde{\mathbf{z}}^2)$  are zero mean random variables, with covariance  $\hat{\Sigma}$  and support  $\tilde{\mathbf{z}}^1 \in \mathcal{W}^1 = [-\hat{\mathbf{z}}, \hat{\mathbf{z}} - \hat{\mathbf{z}}]$  and  $\tilde{\mathbf{z}}^2 \in \mathcal{W}^2 = [-\hat{\mathbf{z}} + \hat{\mathbf{z}}, \hat{\mathbf{z}}]$ , where  $\hat{\mathbf{z}} = \mathbf{E}(\tilde{\mathbf{z}}^+)$ .

Since  $\tilde{\mathbf{z}} = \tilde{\mathbf{z}}^1 + \tilde{\mathbf{z}}^2$ , we can express the data dependency in terms of the segregated primitive uncertainties as follows:

$$\begin{aligned} \mathbf{T}(\tilde{\mathbf{z}}^1, \tilde{\mathbf{z}}^2) &= \mathbf{T}^0 + \sum_{k=1}^N \mathbf{T}^k (\tilde{z}_k^1 + \tilde{z}_k^2), \\ \mathbf{h}(\tilde{\mathbf{z}}^1, \tilde{\mathbf{z}}^2) &= \mathbf{h}^0 + \sum_{k=1}^N \mathbf{h}^k (\tilde{z}_k^1 + \tilde{z}_k^2). \end{aligned}$$

Hence, using the segregated linear decision rule, we can approximate the stochastic model (3) as follows:

$$\begin{aligned} Z_{SLDR} = \min \quad & \mathbf{c}'\mathbf{x} + \mathbf{f}'\mathbf{w}^0 \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{T}^k\mathbf{x} + \mathbf{W}\mathbf{w}^{1k} = \mathbf{h}^k \quad \forall k \in \{0, \dots, N\} \\ & \mathbf{T}^k\mathbf{x} + \mathbf{W}\mathbf{w}^{2k} = \mathbf{h}^k \quad \forall k \in \{0, \dots, N\} \\ & w_i(\mathbf{z}^1, \mathbf{z}^2) \geq 0 \quad \forall (\mathbf{z}^1, \mathbf{z}^2) \in \mathcal{W}^1 \times \mathcal{W}^2, \forall i \in I \\ & \mathbf{x} \geq 0 \\ & \mathbf{w}(\cdot) \in \mathcal{S}. \end{aligned} \tag{19}$$

**Theorem 3** *Segregated linear decision rule gives an objective value that is at least as good as linear decision rule, that is,  $Z_{SLDR} \leq Z_{LDR}$ .*

**Proof :** Let  $\mathbf{x}$  and  $\mathbf{w}^k$   $k = 1, \dots, N$  be feasible to Problem (4). We will show that

$$\mathbf{w}^{1k} = \mathbf{w}^{2k} = \mathbf{w}^k, \quad k = 1, \dots, N$$

are feasible in Problem (19). It is trivial to see that the affine constraints in Problem (19) are satisfied. To show feasibility in the robust counterpart, it suffices to show that  $y_0, \mathbf{y}$  are feasible in

$$y_0 + \mathbf{y}'\mathbf{z} \geq 0 \quad \forall \mathbf{z} \in \mathcal{W} \quad (20)$$

then they are also feasible in

$$y_0 + \mathbf{y}'(\mathbf{z}^1 + \mathbf{z}^2) \geq 0 \quad \forall \mathbf{z}^1 \in \mathcal{W}^1, \mathbf{z}^2 \in \mathcal{W}^2. \quad (21)$$

Indeed, feasibility in the robust counterpart (20) implies

$$y_0 + \mathbf{y}'(\mathbf{z}^1 + \mathbf{z}^2) \geq 0 \quad \forall \mathbf{z}^1 + \mathbf{z}^2 \in \mathcal{W}.$$

However, for all  $\mathbf{z}^1 \in \mathcal{W}^1$  and  $\mathbf{z}^2 \in \mathcal{W}^2$ , we have  $\mathbf{z}^1 + \mathbf{z}^2 \in \mathcal{W}$ . Hence,  $y_0, \mathbf{y}$  are also feasible in the robust counterpart (20). ■

### 3.3.1 Segregated Deflected Linear Decision Rule

Although segregated linear decision rules perform better than linear decision rules, they are not necessarily better than deflected linear decision rules. Hence, in the spirit that deflected linear decision rules improve upon linear decision rules, we define the segregated deflected linear decision rule as follows:

$$\mathbf{w}(\tilde{\mathbf{z}}^1, \tilde{\mathbf{z}}^2) = \mathbf{r}(\tilde{\mathbf{z}}^1, \tilde{\mathbf{z}}^2) + \sum_{i \in I_1} (r_i(\tilde{\mathbf{z}}^1, \tilde{\mathbf{z}}^2)^-)\mathbf{p}^i, \quad (22)$$

in which  $\mathbf{w}(\cdot) \in \mathcal{S}$ .

In the same manner as we have derived earlier, we need to bound  $E((r_i(\tilde{\mathbf{z}}^1, \tilde{\mathbf{z}}^2)^-), i \in I_1$ . Following Theorem 1, and under the model of segregated primitive uncertainty  $U_2$ , we have

$$E\left((y_0 + \mathbf{y}^1'\tilde{\mathbf{z}}^1 + \mathbf{y}^2'\tilde{\mathbf{z}}^2)^-\right) \leq \hat{h}(y_0, \mathbf{y}^1, \mathbf{y}^2),$$

where

$$\begin{aligned} \hat{h}(y_0, \mathbf{y}^1, \mathbf{y}^2) \triangleq & \min_{\mathbf{s}^1, \mathbf{t}^1, \mathbf{u}^1, \mathbf{v}^1, \mathbf{s}^2, \mathbf{t}^2, \mathbf{u}^2, \mathbf{v}^2 \geq \mathbf{0}} \left\{ \frac{1}{2} \left( -y_0 + (\mathbf{s}^1 + \mathbf{u}^1)'(\bar{\mathbf{z}} - \hat{\mathbf{z}}) + (\mathbf{t}^1 + \mathbf{v}^1)'\hat{\mathbf{z}} + \right. \right. \\ & \left. \left. (\mathbf{s}^2 + \mathbf{u}^2)'\hat{\mathbf{z}} + (\mathbf{t}^2 + \mathbf{v}^2)'(\underline{\mathbf{z}} - \hat{\mathbf{z}}) + \right. \right. \\ & \left. \left. \left( -y_0 + (\mathbf{s}^1 - \mathbf{u}^1)'(\bar{\mathbf{z}} - \hat{\mathbf{z}}) + (\mathbf{t}^1 - \mathbf{v}^1)'\hat{\mathbf{z}} + (\mathbf{s}^2 - \mathbf{u}^2)'(\hat{\mathbf{z}}) + (\mathbf{t}^2 - \mathbf{v}^2)'(\underline{\mathbf{z}} - \hat{\mathbf{z}}) \right)^2 + \right. \right. \\ & \left. \left. \left\| \hat{\Sigma}^{1/2} \begin{pmatrix} -\mathbf{y}^1 - \mathbf{s}^1 + \mathbf{t}^1 + \mathbf{u}^1 - \mathbf{v}^1 \\ -\mathbf{y}^2 - \mathbf{s}^2 + \mathbf{t}^2 + \mathbf{u}^2 - \mathbf{v}^2 \end{pmatrix} \right\|_2^2 \right\}^{1/2}. \quad (23) \end{aligned}$$

Finally, we propose the following approximation for Problem (3) using the segregated deflected linear decision rule as the following

$$\begin{aligned}
\bar{Z}_{SDLDR} = \min \quad & \mathbf{c}'\mathbf{x} + \mathbf{f}'\mathbf{r}^0 + \sum_{i \in I_1} \bar{f}_i g_i \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\
& \mathbf{T}^k \mathbf{x} + \mathbf{W}\mathbf{r}^{1k} = \mathbf{h}^k & \forall k \in \{0, \dots, N\} \\
& \mathbf{T}^k \mathbf{x} + \mathbf{W}\mathbf{r}^{2k} = \mathbf{h}^k & \forall k \in \{0, \dots, N\} \\
& r_j(\mathbf{z}^1, \mathbf{z}^2) \geq 0 & \forall (\mathbf{z}^1, \mathbf{z}^2) \in \mathcal{W}^1 \times \mathcal{W}^2, \forall j \in I_2 \\
& g_i \geq \hat{h}(r_i^0, (r_i^{11}, \dots, r_i^{1N}), (r_i^{21}, \dots, r_i^{2N})) & \forall i \in I_1 \\
& \mathbf{x} \geq 0 \\
& \mathbf{r}(\cdot) \in \mathcal{S}.
\end{aligned} \tag{24}$$

**Theorem 4** *The segregated deflected linear decision rule gives an objective value that is at least as good as the deflected linear decision rule, that is,  $\bar{Z}_{SDLDR} \leq \bar{Z}_{DLDR}$ .*

**Proof :** Let  $\mathbf{x}, \mathbf{g}$  and  $\mathbf{r}^k$   $k = 0, \dots, N$  be feasible to Problem (15). We will show that

$$\mathbf{r}^{1k} = \mathbf{r}^{2k} = \mathbf{r}^k$$

$k = 0, \dots, N$ , are also feasible in Problem (24). We have addressed the feasibility of the affine constraint and the robust counterparts in the proof of Theorem 3. We need to show that

$$g_i \geq \hat{h}(r_i^0, (r_i^1, \dots, r_i^N), (r_i^1, \dots, r_i^N)) \quad \forall i \in I_1.$$

It suffices to show that

$$\hat{h}(y_0, \mathbf{y}, \mathbf{y}) \leq h(y_0, \mathbf{y}).$$

Let  $\mathbf{s}, \mathbf{t}, \mathbf{u}, \mathbf{v}$  be the optimal solution in the optimization of Problem (11). Indeed,

$$\begin{aligned}
& \hat{h}(y_0, \mathbf{y}, \mathbf{y}) \\
\leq & \frac{1}{2} \left( -y_0 + (\mathbf{s} + \mathbf{u})'(\bar{\mathbf{z}} - \hat{\mathbf{z}}) + (\mathbf{t} + \mathbf{v})'\hat{\mathbf{z}} + \right. \\
& (\mathbf{s} + \mathbf{u})'\hat{\mathbf{z}} + (\mathbf{t} + \mathbf{v})'(\underline{\mathbf{z}} - \hat{\mathbf{z}}) + \\
& \left. \left\{ \left( -y_0 + (\mathbf{s} - \mathbf{u})'(\bar{\mathbf{z}} - \hat{\mathbf{z}}) + (\mathbf{t} - \mathbf{v})'\hat{\mathbf{z}} + (\mathbf{s} - \mathbf{u})'(\hat{\mathbf{z}}) + (\mathbf{t} - \mathbf{v})'(\underline{\mathbf{z}} - \hat{\mathbf{z}}) \right)^2 + \right. \right. \\
& \left. \left. \left\| \hat{\Sigma}^{1/2} \begin{pmatrix} -\mathbf{y} - \mathbf{s} + \mathbf{t} + \mathbf{u} - \mathbf{v} \\ -\mathbf{y} - \mathbf{s} + \mathbf{t} + \mathbf{u} - \mathbf{v} \end{pmatrix} \right\|_2^2 \right\}^{1/2} \right) \\
= & \frac{1}{2} \left( -y_0 + (\mathbf{s} + \mathbf{u})'\bar{\mathbf{z}} + (\mathbf{t} + \mathbf{v})'\underline{\mathbf{z}} + \left\{ \left( -y_0 + (\mathbf{s} - \mathbf{u})'\bar{\mathbf{z}} + (\mathbf{t} - \mathbf{v})'\underline{\mathbf{z}} \right)^2 + \right. \right. \\
& \left. \left. \text{E} \left( \left( -\mathbf{y} - \mathbf{s} + \mathbf{t} + \mathbf{u} - \mathbf{v} \right)' \underbrace{(\tilde{\mathbf{z}}^1 + \tilde{\mathbf{z}}^2)}_{=\tilde{\mathbf{z}}} \right)^2 \right) \right\}^{1/2} \right) \\
= & h(y_0, \mathbf{y}).
\end{aligned}$$

■

Notice that we can essentially iterate the above process and partition the segregated primitive uncertainties to improve the decision rules at the expense of increased computational complexity.

### 3.4 Multiperiod modeling

Our approach can be easily extended to deal with multiperiod stochastic programming problems as the following.

$$\begin{aligned}
\min & \quad \mathbf{c}'\mathbf{x} + \text{E}(\sum_{t=1}^T \mathbf{f}'_t \mathbf{w}_t(\tilde{\boldsymbol{\xi}}_t)) \\
\text{s.t.} & \quad \mathbf{A}\mathbf{x} = \mathbf{b} \\
& \quad \mathbf{x} \geq 0 \\
& \quad \mathbf{T}_t(\tilde{\boldsymbol{\xi}}_t)\mathbf{x} + \sum_{\tau=1}^t \mathbf{W}_{t\tau} \mathbf{w}_\tau(\tilde{\boldsymbol{\xi}}_\tau) = \mathbf{b}_t(\tilde{\boldsymbol{\xi}}_t), \quad t = 1, 2, \dots, T \\
& \quad w_{ti}(\tilde{\boldsymbol{\xi}}_t) \geq 0, \quad t = 1, 2, \dots, T, i \in I_t,
\end{aligned} \tag{25}$$

where  $\tilde{\boldsymbol{\xi}}_t = (\tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_t)$ , the underlying uncertainties,  $\tilde{\mathbf{z}}_1 \in \mathfrak{R}^{N_1}, \dots, \tilde{\mathbf{z}}_T \in \mathfrak{R}^{N_T}$ , unfold progressively from the first period to the last period, and  $\tilde{\mathbf{z}}_t$  is the vector of primitive uncertainties that is only realized at the  $t^{\text{th}}$  period. We also assume that  $\mathbf{T}_t(\tilde{\boldsymbol{\xi}}_t)$  and  $\mathbf{b}_t(\tilde{\boldsymbol{\xi}}_t)$  are affine in  $\tilde{\boldsymbol{\xi}}_t$ .

It is easy to apply linear decision to obtain an upper bound to the multiperiod stochastic optimization problem. To improve upon the solution using deflected linear decision rule, similar to problem (6), we

can define for all  $t = 1, \dots, T$ ,  $i \in I_t$

$$\begin{aligned}
\bar{f}^{ti} &= \min \quad \sum_{\tau=t}^T \mathbf{f}'_{\tau} \mathbf{p}_{\tau}^{ti} \\
\text{s.t.} \quad & \sum_{\tau=t}^k \mathbf{W}_{k\tau} \mathbf{p}_{\tau}^{ti} = 0 \quad \forall k = t, \dots, T \\
& p_{ii}^{ti} = 1 \\
& p_{\tau j}^{ti} \geq 0 \quad \forall \tau = t, \dots, T, j \in I_{\tau},
\end{aligned} \tag{26}$$

where  $p_{\tau j}^{ti}$  represents the  $j$ th component of the vector  $\mathbf{p}_{\tau}^{ti}$ .

Likewise, we define

$$I_t^1 \triangleq \{\bar{f}^{ti} < \infty : i \in I_t\} \quad I_t^2 \triangleq I_t \setminus I_t^1$$

and  $(\bar{\mathbf{p}}_t^{ti}, \dots, \bar{\mathbf{p}}_T^{ti})$  represents the corresponding optimal solution for any  $t$  and  $i \in I_t^1$ . Similar to the two-stage model, the deflected linear decision rule can be defined as follows.

$$\begin{aligned}
\mathbf{r}_t(\tilde{\boldsymbol{\xi}}_t) &= \mathbf{r}_t^0 + \sum_{\tau=1}^t \sum_{j=1}^{N_{\tau}} \mathbf{r}_{\tau}^j z_{\tau}^j \\
\mathbf{w}_t(\tilde{\boldsymbol{\xi}}_t) &= \mathbf{r}_t(\tilde{\boldsymbol{\xi}}_t) + \sum_{k=1}^t \sum_{i \in I_k^1} (r_{ki}(\tilde{\boldsymbol{\xi}}_k))^{-\bar{p}_k^{ti}}.
\end{aligned}$$

Observe that the above decision rule fulfills the nonanticipativity requirement. Essentially, we end up with a formulation for the multi-period model similar to the one for the two-period model we have presented.

**Remark 1:** Notice that the number of possible scenarios in problem (25) is exponential in the number of time periods  $T$ . As a result the number of decision variables is also exponential in  $T$ . (Notice that even when  $T = 2$ , the possible scenarios in (25) could be huge, or even infinite. We have discussed this for the two-stage model. Here we focus on the dependency of the problem size on  $T$  only.) However, under the deflected or segregated linear decision rules, problem (25) is reduced to an optimization problem whose decisions variables are  $\mathbf{x}$  and  $\mathbf{r}_{\tau}^j$ ,  $j = 0, 1, \dots, N_{\tau}$  and  $\tau = 1, \dots, T$ . It is clear that, the size of this reduced optimization problem is polynomial in  $T$ , i.e., it does not explode exponentially with the number of periods. Notice that in this reduced optimization problem the size of the scenarios plays a role only in computing the objective value. However, as in the two-stage model, the expectation in the objective function will be approximated by a deterministic function of (polynomially many) decision variables.

**Remark 2:** Since we approximate a problem of exponential size by one of much smaller size, it is reasonable to question the quality of our approximation. As mentioned earlier, quoted from Shapiro and Nemirovski [31], we do not claim our approximation will work well for all multi-period models. Our

decision rules provide tractable approximations to an intractable problem. It may provide reasonably high quality solutions as well to certain classes of problems (some evidence can be found in [6, 7] where simple linear decision rules are applied to a class of inventory models). Further, our decision rules can be viewed as a first order approximation of the expected future costs, so that we can determine the first stage or the ‘here-and-now’ decision. In practice, we do not use the decision rule as the responding actions in the subsequent stages. Instead, we adopt the rolling horizon approach, that is, we resolve the subsequent stages upon realizations of uncertainties at earlier stages.

## 4 Computational Experiment

In this section, we illustrate our approach in a preliminary computational experiment. We demonstrate that for a two-stage stochastic optimization problem our proposed framework achieves a performance in objective values similar to what the Monte Carlo approach does. Yet, what is different is that the size of our model increases polynomially with the number of stages in the model, while the sample sizes in the Monte Carlo approach may increase exponentially (see, e.g., Shapiro and Nemirovski [31] and Dyer and Stougie [21]). Therefore, we believe our proposed model is promising in addressing large-scale multiperiod stochastic optimization models.

In our experiment, we consider a project management example with several activities. Project management problems can be represented by a directed graph. Each node on the graph represents an event marking the completion of a particular subset of activities. We denote the set of directed arcs on the graph as  $\mathcal{E}$ . Hence, an arc  $(i, j) \in \mathcal{E}$  is an activity that connects event  $i$  to event  $j$ . By convention, we use node 1 as the start event and the last node  $n$  as the end event.

Each activity has a random duration  $\tilde{t}_{ij}$ . The completion of activities must adhere to precedent constraints. For instance, activity  $e_1$  precedes activity  $e_2$  if activity  $e_1$  must be completed before activity  $e_2$ . In our computational experiments, we assume that the random duration  $\tilde{t}_{ij}$  is independent of the durations of other activities. In addition,  $\tilde{t}_{ij}$  also depends on the additional amount of resource,  $x_{ij} \in [0, 1]$ , committed on the activity as follows

$$\tilde{t}_{ij} = b_{ij} + a_{ij}(1 - x_{ij})\tilde{z}_{ij}, \quad (27)$$

where  $\tilde{z}_{ij} \in [-\bar{z}_{ij}, \bar{z}_{ij}]$ ,  $(i, j) \in \mathcal{E}$  are independent random variables with zero means and standard deviations  $\sigma_{ij}$ .

Let  $c_{ij}$  denote the cost of using each unit of resource for the activity on the arc  $(i, j)$ . Our goal is to minimize the expected completion time of the project subject to the constraint that the total resource available is no more than a budget  $C$ .

A stochastic programming model that addresses the above project management problem is as follows.

$$\begin{aligned}
\min \quad & E(y_n(\tilde{\mathbf{z}})) \\
\text{s.t.} \quad & \mathbf{c}'\mathbf{x} \leq C \\
& y_j(\tilde{\mathbf{z}}) - y_i(\tilde{\mathbf{z}}) - w_{ij}(\tilde{\mathbf{z}}) = b_{ij} + a_{ij}(1 - x_{ij})\tilde{z}_{ij} \quad \forall (i, j) \in \mathcal{E} \\
& y_1(\tilde{\mathbf{z}}) = 0 \\
& w_{ij}(\tilde{\mathbf{z}}) \geq 0 \quad \forall (i, j) \in \mathcal{E} \\
& \mathbf{0} \leq \mathbf{x} \leq \mathbf{1} \\
& \mathbf{x} \in \mathbb{R}^{|\mathcal{E}|} \\
& \mathbf{w}(\cdot), \mathbf{y}(\cdot) \in \mathcal{Y},
\end{aligned} \tag{28}$$

where  $y_j(\tilde{\mathbf{z}})$  represents the earliest completion time of event  $j$  and  $w_{ij}(\tilde{\mathbf{z}})$  is the slack variable. In the optimization formulation, the first inequality is the budget constraint, and the first equality together with the second inequality models the precedent constraints.

To check semi-complete recourse and model the deflected linear decision rules, we first take a look at the following problem

$$\begin{aligned}
\bar{f}_{kl} = \min \quad & y_n \\
\text{s.t.} \quad & y_j - y_i - w_{ij} = 0 \quad \forall (i, j) \in \mathcal{E} \\
& y_1 = 0 \\
& w_{kl} = 1 \\
& \mathbf{w} \geq \mathbf{0},
\end{aligned}$$

for each  $(k, l) \in \mathcal{E}$ . Here,  $\bar{f}_{kl}$  corresponds to the longest duration path from node 1 to node  $n$  in a network where one of the arcs  $(k, l)$  has a duration of one unit of time, while the rest of the arcs have zero durations. Clearly,  $\bar{f}_{kl} = 1$  for all  $(k, l) \in \mathcal{E}$ .

#### 4.1 Comparison between the deflected linear decision rule and the sampling approach

In this subsection we compare the deflected linear decision rule model with the sampling based stochastic programming solution approach. The sampling based stochastic programming formulation is as follows:

$$\begin{aligned}
Z_1(K) = \min \quad & \frac{1}{K} \sum_{k=1}^K y_n^k \\
\text{s.t.} \quad & \mathbf{c}'\mathbf{x} \leq C \\
& y_j^k - y_i^k \geq b_{ij} + a_{ij}(1 - x_{ij})z_{ij}^k \quad \forall (i, j) \in \mathcal{E}, k \in \{1, \dots, K\} \\
& y_1^k = 0 \quad k \in \{1, \dots, K\} \\
& \mathbf{0} \leq \mathbf{x} \leq \mathbf{1} \\
& \mathbf{x} \in \mathfrak{R}^{|\mathcal{E}|} \\
& \mathbf{y}^k \in \mathfrak{R}^n \quad \forall k \in \{1, \dots, K\},
\end{aligned} \tag{29}$$

where  $z_{ij}^k$ ,  $k = 1, \dots, K$  are independent samples of  $\tilde{z}_{ij}$ , and  $y_j^k$  is the earliest completion time of event  $j$  for scenario  $k$ , similar to the meaning of  $y_j$  in model (28).

For the deflected linear decision rule approach, we adopt the framework of (16) as follows:

$$\begin{aligned}
Z_2 = \min \quad & y_n^0 + \sum_{e \in \mathcal{E}} g_e \\
\text{s.t.} \quad & \mathbf{c}'\mathbf{x} \leq C \\
& y_j^0 - y_i^0 - r_{(i,j)}^0 = b_{ij} \quad \forall (i, j) \in \mathcal{E} \\
& y_j^{(i,j)} - y_i^{(i,j)} - r_{(i,j)}^{(i,j)} = a_{ij}(1 - x_{ij}) \quad \forall (i, j) \in \mathcal{E} \\
& y_j^e - y_i^e - r_{(i,j)}^e = 0 \quad \forall e, (i, j) \in \mathcal{E}, e \neq (i, j) \\
& y_1^e = 0 \quad \forall e \in \mathcal{E} \\
& 2g_e + r_e^0 - (\mathbf{s}^e + \mathbf{u}^e)' \bar{\mathbf{z}} - (\mathbf{t}^e + \mathbf{v}^e)' \underline{\mathbf{z}} \geq \\
& \quad \left( (-r_e^0 + (\mathbf{s}^e - \mathbf{u}^e)' \bar{\mathbf{z}} + (\mathbf{t}^e - \mathbf{v}^e)' \underline{\mathbf{z}})^2 + \right. \\
& \quad \left. \|\Sigma^{1/2}(-\mathbf{r}_e - \mathbf{s}^e + \mathbf{t}^e + \mathbf{u}^e - \mathbf{v}^e)\|_2^2 \right)^{\frac{1}{2}} \quad \forall e \in \mathcal{E} \\
& \mathbf{s}^e, \mathbf{t}^e, \mathbf{u}^e, \mathbf{v}^e \geq \mathbf{0} \quad \forall e \in \mathcal{E} \\
& \mathbf{0} \leq \mathbf{x} \leq \mathbf{1} \\
& \mathbf{r}_e, \mathbf{s}^e, \mathbf{t}^e, \mathbf{u}^e, \mathbf{v}^e \in \mathfrak{R}^{|\mathcal{E}|} \quad \forall e \in \mathcal{E} \\
& \mathbf{y}^0 \in \mathfrak{R}^n, \mathbf{x}, \mathbf{r}^0 \in \mathfrak{R}^{|\mathcal{E}|},
\end{aligned} \tag{30}$$

where  $\mathbf{r}_e = (r_e^a : a \in \mathcal{E})$ ,  $y_j^e$  is the coefficient of  $\tilde{z}_e$  in the deflected linear decision rule for  $y_j(\tilde{\mathbf{z}})$ , the earliest completion time of event  $j$ , and  $r_{(i,j)}^e$  is the coefficient of  $\tilde{z}_e$  in the deflected linear decision rule for the slack variable  $w_{ij}(\tilde{\mathbf{z}})$ .

For our computational experiment, we create a fictitious project with the activity network in the form of  $H$  by  $W$  grid (see Figure 1). There are a total of  $H \times W$  nodes, with the first node at

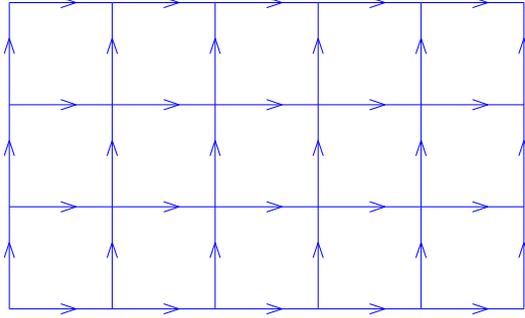


Figure 1: Project management “grid” with height,  $H = 4$  and width  $W = 6$ .

the left bottom corner and the last node at the right upper corner. Each arc on the graph proceeds either towards the right node or the upper node. We assume that the durations of the activities are independent and identical random variables. In particular, for every arc  $(i, j)$ ,

$$P(\tilde{z}_{ij} = z) = \begin{cases} \beta & \text{if } z = \frac{1}{2\beta} \\ (1 - \beta) & \text{if } z = -\frac{1}{2(1-\beta)}. \end{cases}$$

The parameter  $\beta$  controls the variance of the durations of the activities, which increases to  $\infty$  as  $\beta$  decreases to zero. We set  $b_{ij} = 3, a_{ij} = 3$  for all  $(i, j) \in \mathcal{E}$  and  $\mathbf{c} = \mathbf{1}$ . The project grid is fixed to  $H = 4$  by  $W = 6$ . We compare the performances of models (29) and (30) in Table 1. In the table,  $Z_1(K)$  and  $Z_2$  are the optimal objective values of Model (29) using  $K$  samples, and Model (30), respectively. We then use Monte Carlo simulation with 100,000 samples to estimate the actual objective function values achieved by the first stage solutions derived from Models (29) and (30). The corresponding estimated objective function values are recorded in columns  $\hat{Z}_1(K)$  and  $\hat{Z}_2$ . The computation experiment is conducted on an 800 MHz Labtop with 1G RAM using CPLEX version 9.1.

From Table 1, we can see how the performance of stochastic optimization of Model (29) changes as we change the sample sizes. We observe that the optimal objective value of Model (29) underestimates the expected completion time derived from the solutions of Model (29), which is due to the artifact of the sampling approximation of stochastic optimization. When the parameter  $\beta$  becomes very small, the variances of the primitive uncertainties  $\tilde{z}_{ij}$  increase dramatically, and the gap between  $Z_1$  and  $\hat{Z}_1$  increases significantly. On the other hand, under the same circumstances, Model (30) provides very consistent bounds of the expected completion times and significantly outperforms the sampling method

using 1000 samples. When the variances of  $\tilde{z}_{ij}$  are moderate, the sampling approach with 1000 samples slightly outperforms our approach. Given the approximation nature of our approach, the quality of the solutions generated by Model (30) is encouraging.

## 4.2 The comparison of various decision rules under ambiguous probability distributions

In this subsection, we consider the case that the probability distributions of the random variables are not fully specified. Instead, we assume that all random variables  $\tilde{z}_j$ s are independently distributed, each of which has a mean of zero, a standard deviation of

$$\sigma(\tilde{z}_j) = \frac{1}{2\sqrt{\beta(1-\beta)}},$$

and a support in

$$[\underline{z}_j, \bar{z}_j] = \left[ -\frac{1.2}{2(1-\beta)}, \frac{1.2}{2\beta} \right].$$

Furthermore, we also assume

$$\hat{z}_j = \mathbb{E}(\tilde{z}_j^+) = 1/2, \quad \sigma(\tilde{z}_j^+) = \frac{1}{2}\sqrt{\frac{1-\beta}{\beta}}, \quad \text{and} \quad \sigma(\tilde{z}_j^-) = \frac{1}{2}\sqrt{\frac{\beta}{1-\beta}}.$$

Let

$$\tilde{z}_j^1 = \tilde{z}_j^+ - \mathbb{E}(\tilde{z}_j^+) \quad \text{and} \quad \tilde{z}_j^2 = -\tilde{z}_j^- + \mathbb{E}(\tilde{z}_j^+).$$

We can determine the covariance of  $(\tilde{z}_j^1, \tilde{z}_j^2)$  as follows

$$\hat{\Sigma} = \begin{bmatrix} \sigma^2(\tilde{z}_j^+) & \mathbb{E}(\tilde{z}_j^+)^2 \\ \mathbb{E}(\tilde{z}_j^+)^2 & \sigma^2(\tilde{z}_j^-) \end{bmatrix}$$

Based on the above specifications, we adopt the deterministic optimization formulations (4), (16), (19) and (24) to model the stochastic project management problem by using linear, deflected linear, segregated linear, and segregated deflected linear decision rules, respectively. Each model provides an upper bound to the original stochastic optimization problem.

Similar to the computational study conducted in the previous subsection, we model the project as a grid presented in Figure 1.

Table 2 presents the objective function values according to the different decision rules. It is clear from the table that more sophisticated decision rules do improve upon the rather primitive linear decision rule.

C	$\beta$	$Z_2$	$Z_1(1000)$	$Z_1(500)$	$Z_1(100)$	$Z_1(50)$
8	0.0001	58.50	13.50	13.50	12.00	12.00
8	0.001	58.53	34.50	33.00	15.00	13.49
8	0.005	58.67	52.60	48.50	49.33	17.97
8	0.01	58.83	55.60	49.06	46.51	26.96
8	0.1	54.34	42.54	43.54	39.30	41.39
8	0.2	48.73	38.29	38.50	36.47	38.24
8	0.3	45.30	36.27	36.31	35.38	36.54
8	0.4	41.90	35.11	35.14	34.24	35.08
19	0.0001	44.25	13.50	13.50	12.00	12.00
19	0.001	44.27	22.49	19.50	14.99	13.49
19	0.005	44.35	39.03	34.75	20.99	17.97
19	0.01	44.45	41.42	36.38	29.92	20.97
19	0.1	42.67	35.73	36.25	32.85	33.43
19	0.2	39.32	33.01	32.99	31.42	32.60
19	0.3	36.26	31.59	31.59	30.87	31.65
19	0.4	33.38	30.88	30.82	30.14	30.86
C	$\beta$	$\hat{Z}_2$	$\hat{Z}_1(1000)$	$\hat{Z}_1(500)$	$\hat{Z}_1(100)$	$\hat{Z}_1(50)$
8	0.0001	59.06	69.60	70.50	70.36	70.36
8	0.001	58.22	59.58	59.65	65.72	68.99
8	0.005	56.83	58.39	58.99	59.53	63.95
8	0.01	55.59	58.32	57.53	58.27	59.64
8	0.1	43.95	42.99	43.74	44.68	44.75
8	0.2	38.98	38.55	38.75	39.42	40.25
8	0.3	37.09	36.50	36.52	36.94	37.39
8	0.4	35.97	35.40	35.45	35.92	36.07
19	0.0001	43.90	69.60	70.50	70.36	70.36
19	0.001	44.01	47.84	55.41	65.72	68.99
19	0.005	43.59	45.96	45.79	48.73	63.95
19	0.01	43.31	45.99	44.65	45.21	58.46
19	0.1	37.15	36.10	36.67	37.65	39.33
19	0.2	33.38	33.13	33.26	34.02	34.74
19	0.3	33.75	31.76	31.81	32.28	32.97
19	0.4	31.84	31.06	31.07	31.46	31.85

Table 1: Objective values of models (30) and (29) for  $K = 50, 100$  and  $500$ .

$C$	$\beta$	$Z_{LDR}$	$\bar{Z}_{DLDR}$	$Z_{SLDR}$	$\bar{Z}_{SDLDR}$
8	0.1	70	55.832	66.667	54.344
8	0.2	61.406	49.082	60.75	48.734
8	0.3	53.143	45.947	53.143	45.295
8	0.4	46.5	43.834	46.5	41.898
19	0.1	50.938	43.712	50.167	42.668
19	0.2	44.813	39.508	44.813	39.321
19	0.3	38.714	37.556	38.714	36.259
19	0.4	35.25	35.25	35.25	33.375

Table 2: Objective values of approximation models based on linear, deflected linear, segregated linear, and segregated deflected linear decision rules.

## 5 Conclusions

Although we only solve the stochastic optimization model approximately, we feel that the key advantage of our approach is the scalability to multistage models without suffering from the “curse of dimensionality” experienced by most dynamic stochastic programs.

We see various advantages in formulating a stochastic optimization problem as a standard mathematical programming model such as a second order cone program. It enables us to exploit specific structures for computational efficiency suited for large scale implementations. Moreover, the decision rule approaches blend well with recent development in robust optimization in approximating chance constraint problems, such as those by Chen, Sim and Sun [17], Bertsimas and Sim [9] and Ben-Tal and Nemirovski [5]. Hence, it is also possible to extend the model to consider multistage stochastic optimization problems with chance constraints.

When the distributions of the uncertainties are precisely known, we do not claim that our approach is superior to and can replace the sampling method. Instead, we believe both approaches (our approach and the sampling method) are valuable in certain domains. For instance, the preliminary computational study presented in this paper demonstrates that our approach may outperform the sampling method when the variances of the primitive uncertainties are huge. In other occasions, the sampling method may outperform our approach. In summary, the contribution of this paper is to provide a viable alternative to the sampling approach for solving stochastic programming problems.

**Acknowledgment:** We thank the AE and the referees for many useful comments on previous versions of the paper.

## References

- [1] Atamturk, A., Zhang, M. (2005) Two-Stage Robust Network Flow and Design under Demand Uncertainty, *Operations Research*, Forthcoming.
- [2] Beale, E.M.L. (1955): On minimizing a convex function subject to linear inequalities, *Journal of the Royal Statistical Society Series B*, 17, 173-184.
- [3] Ben-Tal, A., Nemirovski, A. (1998): Robust convex optimization, *Math. Oper. Res.*, 23, 769-805.
- [4] Ben-Tal, A., Nemirovski, A. (1999): Robust solutions to uncertain programs, *Oper. Res. Let.*, 25, 1-13.
- [5] Ben-Tal, A., Nemirovski, A. (2000): Robust solutions of Linear Programming problems contaminated with uncertain data, *Math. Progr.*, 88, 411-424.
- [6] Ben-Tal, A., A. Goryashko, E. Guslitzer and A. Nemirovski. (2004): Adjustable robust solutions of uncertain linear programs, *Math. Progr.*, 99, 351-376.
- [7] Ben-Tal, A., B. Golany, A Nemirovski., J. Vial. (2005): Supplier-Retailer Flexible Commitments Contracts: A Robust Optimization Approach. *Manufacturing & Service Operations Management*, 7(3), 248-273.
- [8] Bertsimas, D., Sim, M. (2003): Robust Discrete Optimization and Network Flows, *Math. Progr.*, 98, 49-71.
- [9] Bertsimas, D., Sim, M. (2004): Price of Robustness, *Oper. Res.*, 52(1), 35-53.
- [10] Bertsimas, D., Sim, M. (2006): Tractable Approximations to Robust Conic Optimization Problems, *Math. Progr.*, 107(1), 5 - 36.
- [11] Bertsimas, D., Sim, M. (2004): Robust Discrete Optimization and Downside Risk Measures, *Working Paper*, MIT.
- [12] Birge, J. R., Louveaux, F. (1997): Introduction to Stochastic Programming. *Springer*, New York.
- [13] Calafiore G. and M. C. Campi (2003): Uncertain convex programs: Randomized solutions and confidence levels, *Math. Prog.*, 102. 2546.
- [14] Calafiore G. and M. C. Campi (2004): Decision making in an uncertain environment: the scenariobased optimization approach, *Working Paper*.

- [15] Calafiore, G. and M. C. Campi (2006): The Scenario Approach to Robust Control Design, *IEEE Trans. on Automatic Control*, 51(5), 742-753.
- [16] Charnes, A. and Cooper, W.W. (1959): Uncertain Convex Programs: Randomize solutions and confidence level, *Management Science*, 6, 73-79.
- [17] Chen, X. Sim, M., Sun, P. (2005): A Robust Optimization Perspective of Stochastic Programming, *Operations Research*, Forthcoming.
- [18] Chou, M. Sim, M., So, R. (2005): A Robust Optimization Framework for Analyzing Distribution Systems with Transshipments, *Working Paper*, National University of Singapore.
- [19] Dantzig, G.B. (1955): Linear programming under uncertainty, *Management Science*, 1, 197-206.
- [20] de Farias D.P. and B. Van Roy (2004): On constraint sampling in the linear programming approach to approximate dynamic programming, *Math. Oper. Res.*, 29(3), 462-478.
- [21] Dyer, M., Stougie, L. (2006): Computational complexity of stochastic programming problems, *Mathematical Programming*, 106(3), 423-432.
- [22] El-Ghaoui, Lebret, H. (1997): Robust solutions to least-square problems to uncertain data matrices, *SIAM J. Matrix Anal. Appl.*, 18, 1035-1064.
- [23] El-Ghaoui, L., Oustry, F., Lebret, H. (1998): Robust solutions to uncertain semidefinite programs, *SIAM J. Optim.*, 9, 33-52.
- [24] Erdoğan E. and G. Iyengar (2004): Ambiguous chance constrained problems and robust optimization, *Math. Prog.*, 107, 37-61.
- [25] Erdoğan E. and G. Iyengar (2006): On two-stage convex chance constrained problems. *Mathematical Methods of Operations Research*, to appear.
- [26] Goldfarb, D., Iyengar, G.(2003): Robust quadratically constrained programs, *Math. Prog. Ser. B*, 97(3), 495-515
- [27] Henrion, R. (2005): Structural properties of linear probabilistic constraints, *Stochastic programming E-print series (SPEPS)*, 13.
- [28] Lagoa, C.M., X. Li, and M. Sznaier (2005): Probabilistically constrained linear programs and risk adjusted controller design, *SIAM Journal of Optimization*, 15:938-951.

- [29] Ruszczyński, A., Shapiro, A. (eds.) (2003): *Stochastic Programming, Handbook in Operations Research and Management Science*, Elsevier Science, Amsterdam.
- [30] Scarf, H. (1958): A min-max solution of an inventory problem, *Studies in The Mathematical Theory of Inventory and Production*, 201-209 Stanford University Press, California, 1958.
- [31] Shapiro, A. and Nemirovski, A. (2005): On Complexity of stochastic programming Problems, in *Continuous Optimization: Current Trends and Applications*, V. Jeyakumar and A.M. Rubinov (Eds.), Springer, pp. 111-144.
- [32] Soyster, A.L. (1973): Convex programming with set-inclusive constraints and applications to inexact linear programming, *Oper. Res.*, 21, 1154-1157.