Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 15, No. 2, pp. 219 – 236, 2019

A DISTRIBUTIONALLY ROBUST APPROACH TO A CLASS OF THREE-STAGE STOCHASTIC LINEAR PROGRAMS*

Bin Li, Jie Sun, Kok Lay Teo, Changjun Yu, Min Zhang†

Abstract: A common criterion in multi-stage stochastic programming is the expected total cost, which is risk-neutral and requires full knowledge on the joint distribution of random variables. These restrictions seriously affect the applicability of the multi-stage stochastic programs. By using a three-stage stochastic linear program as an example, we show how a distributionally robust approach could be used as an alternative, which is computationally more tractable. In particular, we show that if the problem is stagewise independent, then a multi-stage linear programming can be equivalent to a conic optimization problem under an affine decision rule. Moreover, this new problem does not require full information on the distribution of random variables; instead, it only requires partial statistical information such as the supporting sets and certain moments of these variables, specified by an ambiguity set. This set of distributions is specified by a very general form that can accommodate a wide class of applications. Our analysis is generally extendable to multi (>3) stage problems. A numerical example is provided to show the advantages of the distributionally robust approach.

Key words: *stochastic programming, conic optimization, duality*

Mathematics Subject Classification: *90B30; 90C15*

1 Introduction

Multi-stage stochastic linear programming is a classical model in operations research with important applications in areas such as production planning [7], finance [17], and others. As a special case, the solution methodology for the two-stage case has been studied. However, the solution methodology for three or more stages are relatively open. In this paper, we develop a distributionally robust approach to three-stage stochastic linear programming (TSSLP) as an example to show how the general multi-stage problems could be solved.

The format of three-stage model is as follows. Let $x_k \in \mathbb{R}^{d_k}, k = 1, 2, 3$, be the decision vectors to be chosen at the kth stage and let $\tilde{z}_k \in \mathbb{R}^{r_k}$ stand for the random vector representing the uncertainty at stage k, which is only revealed after x_k is chosen. Then a next decision $x_{k+1} \in \mathbb{R}^{d_{k+1}}$ is made, representing a recourse action in stage $k+1$. Starting from $k = 1$, this pattern is repeated twice until a final recourse decision x_3 is made. In the linear case, the recourse decision x_{k+1} is obtained by solving a linear program parameterized by all

*This work was partially supported by a grant from National Natural Science Foundation of China under number 61701124, a grant from Science and Technology on Space Intelligent Control Laboratory, No. KGJZDSYS-2018-03, a grant from Sichuan Province Government under application number 19YYJC2557, and a grant from Fundamental Research Funds for the Central Universities (China).

†Corresponding author.

previous x_k and $\tilde{z}_k$. Conceptually, a solution to TSSLP consists of a "decision-realization" chain in the order of

$$x_1, \tilde{z}_1, x_2(\tilde{z}_1), \tilde{z}_2, x_3(\tilde{z}_1, \tilde{z}_2)$$

for all possible realizations of $(\tilde{z}_1, \tilde{z}_2)$. The fact that the decision x_2 and x_3 are affected by all previous decisions and realizations, but not affected by any later decision and realization, is called the nonanticipativity constraints.

To simplify our analysis, we assume that $\tilde{z}_2$ is independent of $\tilde{z}_1$. Then by using expectation as the criterion of the decisions, the TSSLP can be formulated as

$$\min_{x_1 \in \mathcal{X}_1} \left\{ c_1^\top x_1 + \mathbb{E}_{\mathbb{P}_1} \min_{x_2 \in \mathcal{X}_2} \left[c_2^\top x_2 + \mathbb{E}_{\mathbb{P}_2} \left(\min_{x_3 \in \mathcal{X}_3} c_3^\top x_3 \right) \right] \right\} \tag{1.1}$$

where "$\top$" stands for the transpose, $\mathbb{E}$ stands for the expectation, and $\mathbb{P}_1$ and $\mathbb{P}_2$ are the distribution of $\tilde{z}_1$ and $\tilde{z}_2$, respectively. Let $\mathcal{X}_k$ be the feasible region of $x_k, k = 1, 2, 3$. We assume that

$$\begin{aligned}
\mathcal{X}_1 &= \left\{ x_1 \in \mathbb{R}^{d_1} : A_1 x_1 = b_1, x_1 \geq 0 \right\}, \text{ where } A_1 \in \mathbb{R}^{p_1} \times \mathbb{R}^{d_1}, b_1 \in \mathbb{R}^{p_1}, \\
\mathcal{X}_2 &= \left\{ x_2 \in \mathbb{R}^{d_2} : A_2(\tilde{z}_1) x_1 + B_2 x_2 = b_2(\tilde{z}_1), x_2 \geq 0 \right\}, \\
&\quad \text{where } A_2(\tilde{z}_1) \in \mathbb{R}^{p_2 \times d_1}, B_2 \in \mathbb{R}^{p_2} \times \mathbb{R}^{d_2}, b_2(\tilde{z}_1) \in \mathbb{R}^{p_2}, \text{ and} \\
\mathcal{X}_3 &= \left\{ x_3 \in \mathbb{R}^{d_3} : A_3(\tilde{z}_1, \tilde{z}_2) x_1 + B_3 x_2 + C_3 x_3 = b_3(\tilde{z}_1, \tilde{z}_2), x_3 \geq 0 \right\},
\end{aligned}$$

where $A_3(\tilde{z}_1, \tilde{z}_2) \in \mathbb{R}^{p_3} \times \mathbb{R}^{d_1}, B_3 \in \mathbb{R}^{p_3} \times \mathbb{R}^{d_2}, C_3 \in \mathbb{R}^{p_3} \times \mathbb{R}^{d_3}, b_3(\tilde{z}_1, \tilde{z}_2) \in \mathbb{R}^{p_3}$.

We assume that problem (1.1) has a solution and is of relatively complete recourse, namely $\mathcal{X}_1 \neq \emptyset, \mathcal{X}_2 \neq \emptyset$ for any $x_1 \in \mathcal{X}_1$ and $\tilde{z}_1$, and $\mathcal{X}_3 \neq \emptyset$ for any $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$ and $\tilde{z}_1, \tilde{z}_2$. A major difficulty in applying (1.1) in practice is that the model requires full information on the distribution of the random variables, which is often unavailable in practice. In order to circumvent this difficulty, a focal point of recent research is to utilize the tools developed in robust optimization to convert TSSLP to a conic optimization problem, which is computationally tractable (more exactly, solvable in polynomial time of the problem size). The key idea is as follows. Consider the distributionally robust TSSLP (DR-TSSLP) model

$$\min_{x_1 \in \mathcal{X}_1} \left\{ c_1^\top x_1 + \sup_{\mathbb{P}_1 \in \mathcal{P}_1} \mathbb{E}_{\mathbb{P}_1} \min_{x_2 \in \mathcal{X}_2} \left[c_2^\top x_2 + \sup_{\mathbb{P}_2 \in \mathcal{P}_2} \mathbb{E}_{\mathbb{P}_2} \left(\min_{x_3 \in \mathcal{X}_3} c_3^\top x_3 \right) \right] \right\}, \tag{1.2}$$

and $\mathcal{P}_1$ and $\mathcal{P}_2$ are certain sets of probability distributions of $\tilde{z}_1$ and $\tilde{z}_2$, respectively. In essence, model (1.2) assumes that we do not know the exact distribution $\tilde{z}_k$, but we know that the distribution of $\tilde{z}_k$ belongs to an "ambiguity set" $\mathcal{P}_k$. We then use the worst-case expectation over the ambiguity sets as the decision criteria. These worst-case expectations are indeed corresponding to the so-called coherent risk measures in risk theory [3] and have many desirable properties. The interested reader may refer to [11] for details of the fundamental theory and [1] for most recent development on this representation of risk measures.

The model (1.2), although looking more complicated due to the worst-case functions, turns out to be much easier for computation. The key point is that we replace the computation of expectation by the solution of an optimization problem, which happens to be "more tractable" in terms of numerical computation. In fact, the major purpose of this paper is to show that the problem (1.2) can be converted to a conic optimization problem of size polynomial in terms of the input data under suitable conditions. Therefore, (1.2) can be solved efficiently.

It should be noted that the format of the set $\mathcal{P}_k$ that we will choose is highly expressive as demonstrated in Wiesemann et al [22], therefore the theoretical result derived in this paper is widely applicable. In particular, a spectrum of statistics could be utilized in "designing" the set $\mathcal{P}_k$ and thus to create different risk measures. These characteristics reinforce our confidence in viability of using risk measures in the modeling of stochastic optimization problems.

The contribution of this paper is to provide a tractable reformulation to DR-TSSLP. Comparing with the traditional TSSLP, DR-TSSLP does not require the full knowledge of the distribution information. Hence, it is more general and easier for real world applications.

The rest of the paper is organized as follows. The structure of the ambiguity set is defined in Section 2. Then, the three-stage stochastic linear program is refomulated as a conic optimization problem in Section 3. Numreical expriments are carried out in Section 4 to show the effectiveness of th proposed method.

2 Structural Assumptions on Set $\mathcal{P}_k$ and Problem Data

2.1 Notations

We denote a random quantity, say $\tilde{z}$, with the tilde sign. Sets, matrices and vectors are usually represented as script, upper case, and lower case letters, respectively. We use subscript k, say x_k, to indicate a vector or a matrix arising in stage k, whose components are denoted by $x_{k1}, x_{k2}, \ldots$ respectively. If M is an $m \times n$ real matrix, we write $M \in \mathbb{R}^{m\times n}$. Given a regular (i.e. pointed, closed, convex, and with nonempty interior) cone $\mathcal{K}$ in a finite-dimensional Euclidean space, such as the second-order cone or the semidefinite cone, for any two vectors x, y, the notation $x \preceq_{\mathcal{K}} y$ or $y \succeq_{\mathcal{K}} x$ means $y - x \in \mathcal{K}$. The dual cone of $\mathcal{K}$ is denoted by

$$\mathcal{K}^* := \{y : \langle y, x\rangle \geq 0, \forall x \in \mathcal{K}\}.$$

For simplicity of notations, unless otherwise specified, we will always use $x^\top y$, rather than $\langle x, y\rangle$, to represent inner products although it may need more subtle interpretations in some specific cases such as $x, y \in \mathbb{S}^n$, where $\langle x, y\rangle = \mathbf{vec}\, x^\top \mathbf{vec}\, y$ and $\mathbf{vec}\, x$ and $\mathbf{vec}\, y$ are the vectors made from stacking all elements of x and y respectively.

Let $\tilde{z}$ and $\tilde{u}$ be two random vectors in $\mathbb{R}^M$ and $\mathbb{R}^T$, respectively. The set $\mathcal{P}_0(\mathbb{R}^M)$ represents the space of probability distributions on $\mathbb{R}^M$ and $\mathcal{P}_0(\mathbb{R}^M \times \mathbb{R}^T)$ represents the space of probability distributions on $\mathbb{R}^M \times \mathbb{R}^T$, respectively.

2.2 Structure of $\mathcal{P}_k$

We adopt approach of Wiesemann, Kuhn and Sim [22] (WKS format for short) to define the ambiguity sets $\mathcal{P}_k$. It is always convenient from the application point of view that we introduce an auxiliary random vector $\tilde{u}_k \in \mathbb{R}^{t_k}$ at stage k and think of the set $\mathcal{P}_k$ is defined by an expectation constraint and by a support constraint, both in conic form. This scheme does not complicates our analysis in this paper; however, it opens a fertile field of imposing constraints involving high order moments and absolute deviations of $\tilde{z}$ through a lifting procedure with $\tilde{u}$, see [22] for details.

We start from the support sets of $(\tilde{z}_1, \tilde{u}_1)$ and $(\tilde{z}_2, \tilde{u}_2)$. We specify them as

$$\Omega_1 = \left\{(z_1, u_1) \in \mathbb{R}^{r_1} \times \mathbb{R}^{t_1} \,:\, G_1 z_1 + H_1 u_1 \succeq_{\mathcal{K}_1} h_1\right\}, \tag{2.1}$$

and

$$\Omega_2 = \left\{(z_2, u_2) \in \mathbb{R}^{r_2} \times \mathbb{R}^{t_2} \,:\, G_2 z_2 + H_2 u_2 \succeq_{\mathcal{K}_2} h_2\right\}, \tag{2.2}$$

where $G_k \in \mathbb{R}^{L_k \times r_k}$, $H_k \in \mathbb{R}^{L_k \times t_k}$, and $\mathcal{K}_k$ is a regular cone for $k = 1, 2$. Note that the specification of Ω_2 means that the support of $(\tilde{z}_2, \tilde{u}_2)$ does not depend on $(\tilde{z}_1, \tilde{u}_1)$. It is easy to see that the usual box support is a special case of Ω_k. For ease of analysis, we moreover assume that both Ω_1 and Ω_2 are compact although the boundedness assumption on them can be removed in more subtle analysis. For the applications we are concerned, this assumption is natural.

We next define two ambiguity sets, $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively for distribution of $(\tilde{z}_1, \tilde{u}_1)$ and distribution of $(\tilde{z}_2, \tilde{u}_2)$. We assume $\mathcal{P}_1$ is represented as

$$\mathcal{P}_1 = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{r_1} \times \mathbb{R}^{t_1}) : \begin{array}{l} \mathbb{E}_{\mathbb{P}}\left[E_1 \tilde{z}_1 + F_1 \tilde{u}_1\right] = g_1, \\ \mathbb{P}\left[(\tilde{z}_1, \tilde{u}_1) \in \Omega_1\right] = 1 \end{array} \right\}. \tag{2.3}$$

where E_1, F_1 and g_1 are matrices defined with the proper dimension. Similarly, we define $\mathbb{P}_2$ as

$$\mathcal{P}_2 = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{r_2} \times \mathbb{R}^{t_2}) : \begin{array}{l} \mathbb{E}_{\mathbb{P}}\left[E_2 \tilde{z}_2 + F_2 \tilde{u}_2\right] = g_2, \\ \mathbb{P}\left[(\tilde{z}_2, \tilde{u}_2) \in \Omega_2\right] = 1 \end{array} \right\}. \tag{2.4}$$

where E_2, F_2 and g_2 are matrices defined with the proper dimension. The two ambiguity sets are closely connected with the notion of "risk envelope" in the theory of risk measure [1, 11, 19].

If $\tilde{u}_k$ does not arise in a specific application, then we simply set the corresponding F_k, H_k $(k = 1, 2)$ and F_3 to be zero matrices. The use of the auxiliary variable $\tilde{u}_k$ helps to cover many important applications. For instance, it is shown in [22] that the ambiguity set with a second-order moment constraint

$$\mathcal{P}' = \left\{ \mathbb{P}' : \mathbb{E}_{\mathbb{P}'}\left[\tilde{z}\right] = \mu,\ \mathbb{E}_{\mathbb{P}'}\left[(\tilde{z} - \tilde{\mu})(\tilde{z} - \tilde{\mu})^\top\right] \preceq \mathbf{\Sigma} \;\middle|\; \mu \in \mathbb{R}^m,\ \mathbf{\Sigma} \in \mathbb{S}^m_+ \right\}.$$

is the projection onto $\tilde{z}$-space of the following ambiguity set in the format of (2.3), if an auxiliary random matrix $\tilde{\mathbf{U}}$ is introduced.

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^m \times \mathbb{R}^{m \times m}\right) : \begin{array}{l} \mathbb{E}_{\mathbb{P}}\left(\tilde{z}, \tilde{\mathbf{U}}\right) = (\mu, \mathbf{\Sigma}), \\ \mathbb{P}\left(\begin{bmatrix} 1 & (\tilde{z} - \mu)^\top \\ (\tilde{z} - \mu) & \tilde{\mathbf{U}} \end{bmatrix} \succeq 0\right) = 1 \end{array} \right\}.$$

Therefore, with the help of the auxiliary variables, the first-order moment constraint $\mathbb{E}(G\tilde{z} + G\tilde{u}) = g$ can indeed include second-order moment constraint for $\tilde{z}$ as a special case. See [8, 22] for more details.

$\boxed{2.3}$ Related duality theorems

Since we are going to use duality extensively in our analysis, either in finite-dimensional Hilbert spaces or in infinite-dimensional spaces, it would be convenient to list the related duality theorems below.

We first consider the infinite-dimensional duality developed by Rockafellar in [18]. Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{U}$ be three linear spaces. Let $F : \mathcal{X} \times \mathcal{U} \to [-\infty, +\infty]$ be a convex function such that $f(x) = F(x, 0)$ and consider the convexly parameterized family of optimization problems:

$$\min F(x, u) \quad \text{s.t. } x \in \mathcal{X}, \tag{2.5}$$

and let

$$\phi(u) := \inf_{x \in \mathcal{X}} F(x, u).$$

Define the Lagrangian function $K : \mathcal{X} \times \mathcal{Y} \to [-\infty, +\infty]$ as

$$K(x, y) = \inf_{u \in \mathcal{U}} [F(x, u) + \langle u, y \rangle]. \tag{2.6}$$

then one has

$$f(x) = \sup_{y \in \mathcal{Y}} K(x, y), \tag{2.7}$$

and

$$g(y) = \inf_{x \in \mathcal{X}} K(x, y). \tag{2.8}$$

Define the primal problem as

$$\min f(x) \quad \text{s.t. } x \in \mathcal{X}, \tag{2.9}$$

and the dual problem as

$$\max g(y) \quad \text{s.t. } y \in \mathcal{Y}. \tag{2.10}$$

Lemma 2.1. (Strong duality in infinite-dimensional spaces [18, Theorem 15]) Assume $F(x, u)$ is closed convex in u, the following conditions are equivalent.

(a) $\inf (2.9) = \sup (2.10)$;

(b) $\phi(0) = \operatorname{cl} \operatorname{conv} \phi(0)$;

(c) The saddle-value of the Lagrangian exists.

In particular, for semi-infinite optimization [18, Example 4]

$$\min_{x \in C} f(x) \quad \text{s.t. } h(x, z) \le 0 \quad \forall z \in \mathcal{Z}, \tag{2.11}$$

where

$$F(x, u) = \begin{cases} f(x), & \text{if } x \in C \text{ and } h(x, z) \le u(z) \ \forall z \in \mathcal{Z}, \\ +\infty, & \text{otherwise,} \end{cases}$$

with $u : \mathcal{Z} \to \mathbb{R}$, we have

Lemma 2.2. (Strong duality theorem for semi-infinite optimization [18, Theorem 15(a) and Example 4]) A sufficient condition for Lemma 2.1(a) to hold for problem (2.11) is the general Slater condition, i.e., there exists $\bar{x} \in \operatorname{ri} C$ such that $h(\bar{x}, z) < u(z) \ \forall z \in \mathcal{Z}$.

In addition, it is shown [18, Theorem 15(a) and Example 4] that the $\sup (2.10)$ is attained in this case.

We next consider the finite-dimensional conic case. Let E be a finite-dimensional Euclidean space with inner product $\langle \cdot, \cdot \rangle$ and let $\mathcal{K} \subset E$ be a regular cone. Consider a conic problem

$$\min_x \ \langle c, x \rangle \quad \text{s.t. } Ax \succeq_{\mathcal{K}} b, \tag{2.12}$$

along with its conic dual

$$\max_y \ \langle b, y \rangle \quad \text{s.t. } A^* y = c, \ y \succeq_{\mathcal{K}^*} 0, \tag{2.13}$$

where A^* is the adjoint operator of A.

Lemma 2.3. (Strong conic duality in finite-dimensional spaces [4, Theorem 1.4.2]) For Problem (2.12) and its dual (2.13) there hold

(1) The duality is symmetric: the dual problem is conic, and the problem dual to dual is the primal.

(2) The duality gap $\langle c, x\rangle - \langle b, y\rangle$ is nonnegative at every primal-dual feasible pair (x, y).

(3a) If the primal (2.12) is bounded below and strictly feasible (i.e. $Ax \succ_{\mathcal{K}} b$ for some x), then the dual probelm (2.13) is solvable and the optimal values in the problems are equal to each other.

(3b) If the dual (2.13) is bounded above and strictly feasible (i.e., exists $y \succ_{\mathcal{K}^*} 0$ such that $A^* y = c$), then the primal problem (2.12) is solvable and $\min(2.12) = \max(2.13)$.

$\boxed{2.4}$ Assumptions on $A_2(\tilde{z}_1), A_3(\tilde{z}_1,\tilde{z}_2)$, $b_2(\tilde{z}_1)$, $b_3(\tilde{z}_1,\tilde{z}_2)$, $x_2(\tilde{z}_1)$ and $x_3(\tilde{z}_1,\tilde{z}_2)$

We assume that $A_2(\tilde{z}_1), A_3(\tilde{z}_1,\tilde{z}_2)$, $b_2(\tilde{z}_1)$, $b_3(\tilde{z}_1,\tilde{z}_2)$ are affinely dependent on $\tilde{z}_k$, namely there exist A_{ij}, b_{ij} $(i = 2, 3)$, such that

$$A_2(\tilde{z}_1) = \sum_{j=1}^{r_1} A_{2j}\tilde{z}_{1j} + A_{20},\ \ b_2(\tilde{z}_1) = \sum_{j=1}^{r_1} b_{2j}\tilde{z}_{1j} + b_{20}, \tag{2.14}$$

and

$$\begin{aligned} A_3(\tilde{z}_1,\tilde{z}_2) = &\ \sum_{j=1}^{r_1} A_{3j}\tilde{z}_{1j} + \sum_{j=1}^{r_2} \bar{A}_{3j}\tilde{z}_{2j} + A_{30},\\ b_3(\tilde{z}_1,\tilde{z}_2) = &\ \sum_{j=1}^{r_1} b_{3j}\tilde{z}_{1j} + \sum_{j=1}^{r_2} \bar{b}_{3j}\tilde{z}_{2j} + b_{20}. \end{aligned} \tag{2.15}$$

The dependence of x_2 on $\tilde{z}_1$ and x_3 on $(\tilde{z}_1,\tilde{z}_2)$ is more subtle and implicit. As a first-order approximation, we assume that both x_2 and x_3 are affinely dependent on the respective random vectors, i.e.,

$$x_2 = \sum_{j=1}^{r_1} x_{2j}\tilde{z}_{1j} + x_{20},\ \ x_3 = \sum_{j=1}^{r_1} x_{3j}\tilde{z}_{1j} + \sum_{j=1}^{r_2} \bar{x}_{3j}\tilde{z}_{2j} + x_{30}. \tag{2.16}$$

Thus, the problem turns to finding optimal x_1, x_{2j}, x_{3j} and $\bar{x}_{3j}$ for all j.

The affine dependence assumption above has been used first by Ben-Tal and Nemirovski [5] and subsequently used in many literatures, e.g., [2, 6, 9, 10, 22] as a standard assumption. An extensive study on this assumption has appeared in the literature such as [8, 13, 16], which indicates that this assumption generally performs well in practice and can be made less restrictive by introducing an auxiliary random vector $\tilde{u}$ and assuming affine dependence on both $\tilde{z}$ and $\tilde{u}$. Since the analysis with $(\tilde{z},\tilde{u})$ is similar to that of $\tilde{z}$, to simplify our notations, we keep using (2.14), (2.15), and (2.16) in the sequel.

Under affine dependence and the assumption $\mathrm{int}\,(\Omega_\mathrm{k}) \neq \emptyset, \mathrm{k} = 1, 2$, the linear constraints defining $\mathcal{X}_2$ and $\mathcal{X}_3$ can be decomposed as follows.

$$\begin{cases} A_2(\tilde{z}_1)x_1 + B_2 x_2 = b_2(\tilde{z}_1),\\ \quad \forall(\tilde{z}_1,\tilde{u}_1) \in \Omega_1, \end{cases} \iff \begin{cases} A_{2j}x_{1j} + B_2 x_{2j} = b_{2j},\\ j = 0, 1, ..., r_1, \end{cases} \tag{2.17}$$

and

$$\begin{cases} A_3(\tilde{z}_1,\tilde{z}_2)x_1 + B_3x_2 + C_3x_3 = b_3(\tilde{z}_1,\tilde{z}_2) \\ \quad \forall(\tilde{z}_1,\tilde{u}_1)\in\Omega_1, (\tilde{z}_2,\tilde{u}_2)\in\Omega_2 \end{cases} \iff \tag{2.18}$$
$$\begin{cases} A_{3j}x_1 + B_3x_{2j} + C_3x_{3j} = b_{3j}, & j=0,1,\ldots,r_1, \\ \bar{A}_{3j}x_1 + C_3\bar{x}_{3j} = \bar{b}_{3j}, & j=1,\ldots,r_2. \end{cases}$$

The inequality constraints $x_2(\tilde{z}_1) \geq 0$ is equivalent to a set of linear constraints on $x_{2j}, j = 0, 1, ..., r_1$. To see this fact, let us introduce a new notation. Let X_2 be the matrix defined as

$$X_2 := [x_{21}, x_{22}, ..., x_{2r_1}] \in \mathbb{R}^{d_2\times r_1},$$

and let x_2^q be the qth column of $X_2^\top$. Similarly, define a block matrix

$$[X_3, \bar{X}_3] := [x_{31}, x_{32}, ..., x_{3r_1}; \bar{x}_{31}, ..., \bar{x}_{3r_2}] \in \mathbb{R}^{d_3\times(r_1+r_2)}$$

and let $\begin{pmatrix} x_3^q \\ \bar{x}_3^q \end{pmatrix}$ be the qth column of $[X_3, \bar{X}_3]^\top$, in which x_3^q corresponds to the X_3-block and $\bar{x}_3^q$ corresponds to the $\bar{X}_3$-block, respectively, $q = 1, ..., d_3$. Let x_{k0}^q be the qth component of x_{k0}, $k = 1, 2$. Then

$$x_2(\tilde{z}_1) \geq 0 \iff \min\{x_{20}^q + \langle x_2^q, z_1\rangle\} \geq 0,\ \forall z_1 \in \Omega_1,\ q = 1,\ldots,d_2. \tag{2.19}$$

By Lemma 2.3, the dual problem of $\min\{x_{20}^q + \langle x_2^q, z_1\rangle : (z_1, u_1) \in \Omega_1\}$ is

$$\max_{s^q\in\mathcal{K}_1^*}\ x_{20}^q + \langle h_1, s^q\rangle \quad \text{s.t. } G^\top s^q = x_2^q,\ H^\top s^q = 0, \tag{2.20}$$

where s^q is the dual vector. Strong duality holds because int $(\Omega_1) \neq \emptyset$. Therefore min (2.19) = max (2.20) and (2.19) is equivalent to the feasibility of the system

$$x_{20}^q + h_1^\top s^q \geq 0,\ G^\top s^q = x_2^q,\ H^\top s^q = 0,\ s^q \in \mathcal{K}_1^*,\ \forall\ q = 1,\ldots,d_2. \tag{2.21}$$

In a similar manner, we can deduce that the requirement of $x_3(\tilde{z}_1, \tilde{z}_2) \geq 0$ is equivalent to the feasibility of the following system:

$$\begin{aligned} & x_{30}^q + h_1^\top t_1^q + h_2^\top t_2^q \geq 0, \\ & G_1^\top t_1^q = x_3^q,\ G_2^\top t_2^q = \bar{x}_3^q,\ H_1^\top t_1^q = 0,\ H_2^\top t_2^q = 0, \\ & t_1^q \in \mathcal{K}_1^*,\ t_2^q \in \mathcal{K}_2^*,\ \forall\ q = 1,\ldots,d_3. \end{aligned} \tag{2.22}$$

To simplify our notation in the subsequent analysis, we aggregate all decision variables so far into a single vector, namely we define a vector w as

$$w^\top := \left(x_1^\top, {x_{20}^q}^\top, {x_2^q}^\top, {s^q}^\top(q = 1, ..., d_2), {x_{30}^q}^\top, {t_1^q}^\top, {t_2^q}^\top(q = 1, ..., d_3)\right),$$

and define the feasible set specified by (2.17), (2.18), (2.21), and (2.22) as $\mathcal{W}$. Hence we can write all constraints imposed by the affine dependence as $w \in \mathcal{W}$. Clearly, conic $\mathcal{W}$ is a polyhedron and we further assume that $\mathcal{W}$ is nonempty for otherwise the optimal value of (DR-TSSLP) is trivially $+\infty$. It would be also useful to note that the constraint $x_k \in \mathcal{X}_k$, $k = 1, 2, 3$, is the projection of $\mathcal{W}$ onto the space $\mathbb{R}^{d_k}$.

$\boxed{3}$ Reformulation of DR-TSSLP as a Conic Optimization Problem

We start from the third stage recourse function

$$\sup_{\mathbb{P}_2 \in \mathcal{P}_2} \mathbb{E}_{\mathbb{P}_2} \left[\min_{x_3 \in \mathcal{X}_3} c_3^\top x_3 \right]. \tag{3.1}$$

Given $(\tilde{z}_2, \tilde{u}_2) = (z_1, u_1)$, we designate

$$\psi_2(x_1, x_2, z_1, u_1, \tilde{z}_2, \tilde{u}_2) := \min_{x_3 \in \mathcal{X}_3} c_3^\top x_3.$$

Note that (3.1) is indeed the optimal value of the following optimization problem

$$\begin{aligned} \max_{\mathbb{P}_2} \quad & \mathbb{E}_{\mathbb{P}_2} \left[\psi_2(x_1, x_2, z_1, u_1, \tilde{z}_2, \tilde{u}_2) \right] \\ \text{s.t.} \quad & \mathbb{E}_{\mathbb{P}_2}(E_2 \tilde{z}_2 + F_2 \tilde{u}_2) = g_2, \\ & \mathbb{P}_2(G_2 \tilde{z}_2 + H_2 \tilde{u}_2 \succeq_{\mathcal{K}_2} h_2) = 1. \end{aligned} \tag{3.2}$$

According to the theory of semi-infinite programming [15], the dual of (3.2) is a semi-infinite program as follows

$$\begin{aligned} \min_{\xi_2, \eta_2} \quad & g_2^\top \xi_2 + \eta_2 \\ \text{s.t.} \quad & (E_2 z_2 + F u_2)^\top \xi_2 + \eta_2 \geq \psi(x_1, x_2, z_1, u_1, z_2, u_2), \ \forall \ (z_2, u_2) \in \Omega_2, \end{aligned} \tag{3.3}$$

where $(\xi_2, \eta_2) \in \mathbb{R}^{L_2} \times \mathbb{R}$ are the dual variables.

Lemma 3.1. Strong duality holds between (3.2) and (3.3) in the sense that (3.2) is solvable and max (3.2) = min (3.3).

Proof. Observe that for any fixed x_1, x_2, z_1, u_1, due to continuity of ψ_2 and the compactness of Ω_2, $\psi_2(x_1, x_2, z_1, u_1, z_2, u_2)$ is a bounded quantity over $(z_2, u_2) \in \Omega_2$, say

$$|\psi_2(x_1, x_2, z_1, u_1, z_2, u_2)| \leq \ell,$$

where ℓ may depend on x_1, x_2, z_1, u_1 but not on z_2 and u_2. Thus, the point $\xi_2 = 0$ and $\eta_2 = \ell + 1$ is a generalized Slater's point for the dual problem. Applying Lemma 2.2, strong duality holds in the specified sense. $\square$

Theorem 3.1. Under the affine dependent assumption, the problem (DR-TSSLP) is equivalent to the following stochastic program

$$\begin{aligned} \min_{x_1 \in \mathcal{X}_1} \quad & c_1^\top x_1 + \sup_{\mathbb{P}_1 \in \mathcal{P}_1} \mathbb{E}_{\mathbb{P}_1} \min_{x_2 \in \mathcal{X}_2} \left[g_2^\top \xi_2 + \eta_2 + c_2^\top x_2 \right] \\ \text{s.t.} \quad & h_1^\top \beta_1 + h_2^\top \alpha_2 - c_3^\top x_{30} + \eta_2 \geq 0, \\ & G_1^\top \beta_1 + X_3^\top c_3 = 0, \ H_1^\top \beta_1 = 0, \ \beta_1 \in \mathcal{K}_1^*, \\ & G_2^\top \alpha_2 = E_2^\top \xi_2 - \bar{X}_3^\top c_3, \ H_2^\top \alpha_2 = F_2^\top \xi_2 \\ & \alpha_2 \in \mathcal{K}_2^*, \ w \in \mathcal{W}. \end{aligned} \tag{3.4}$$

Proof. Consider the constraint in (3.3), namely

$$(E_2 z_2 + F_2 u_2)^\top \xi_2 + \eta_2 \geq \psi_2(x_1, x_2, z_1, u_1, z_2, u_2), \ \forall \ (z_2, u_2) \in \Omega_2, \tag{3.5}$$

which is equivalent to

$$\forall\ (z_2,u_2)\in\Omega_2,\ \exists\ x_3\in\mathcal{X}_3:(E_2z_2+F_2u_2)^\top\xi_2+\eta_2-c_3^\top x_3\geq 0,$$

or equivalently

$$\min_{(z_2,u_2)\in\Omega_2}\max_{x_3\in\mathcal{X}_3}\left[(E_2z_2+F_2u_2)^\top\xi_2+\eta_2-c_3^\top x_3\right]\geq 0.$$

The function $(E_2z_2+F_2u_2)^\top\xi_2+\eta_2-c_3^\top x_3$ is convex in (z_2,u_2) and concave in x_3 and both sets, Ω_2 and $\mathcal{X}_3$, are closed and convex. By Sion's minimax theorem [20], as Ω_2 is bounded, we have

$$\begin{aligned}0\leq&\min_{(z_2,u_2)\in\Omega_2}\max_{x_3\in\mathcal{X}_3}\left[(E_2z_2+F_2u_2)^\top\xi_2+\eta_2-c_3^\top x_3\right]\\=&\max_{x_3\in\mathcal{X}_3}\min_{(z_2,u_2)\in\Omega_2}\left[(E_2z_2+F_2u_2)^\top\xi_2+\eta_2-c_3^\top x_3\right].\end{aligned}$$

The constraint (3.5) is therefore equivalent to

$$\exists\ x_3\in\mathcal{X}_3,\ \forall\ (z_2,u_2)\in\Omega_2:(E_2z_2+F_2u_2)^\top\xi_2+\eta_2-c_3^\top x_3\geq 0,\tag{3.6}$$

which says that constraint (3.5) can be re-written as

$$\exists\ x_3\in\mathcal{X}_3,\ (E_2z_2+Fu_2)^\top\xi_2+\eta_2\geq c_3^\top x_3,\ \forall\ (z_2,u_2)\in\Omega_2.\tag{3.7}$$

Note that

$$c_3^\top x_3=c_3^\top x_{30}+c_3^\top X_3\tilde{z}_1+c_3^\top\bar{X}\tilde{z}_2.$$

It turns out that constraint (3.7) means that $\exists\ x_3\in\mathcal{X}_3$ such that

$$\begin{aligned}0\leq\min\big\{(\xi_2^\top E_2-c_3^\top\bar{X}_3)z_2+\xi_2^\top F_2u_2+\eta_2-c_3^\top x_{30}-c_3^\top X_3z_1:\\G_2z_2+H_2u_2\succeq_{\mathcal{K}_2}h_2\big\}.\end{aligned}\tag{3.8}$$

By Lemma 2.3, since int $(\Omega_2)\neq\emptyset$, strong duality holds. Thus, (3.8) is equivalent to $\exists\ \alpha_2\in\mathcal{K}_2^*$ such that

$$\begin{aligned}0\leq\max\big\{h_2^\top\alpha_2-c_3^\top x_{30}-c_3^\top X_3z_1+\eta_2:G_2^\top\alpha_2=E_2^\top\xi_2-\bar{X}_3^\top c_3,\\H_2^\top\alpha_2=F_2^\top\xi_2,\alpha_2\in\mathcal{K}_2^*\big\},\end{aligned}$$

therefore, constraint (3.7) can be equivalently replaced by the following system

$$\begin{cases}h_2^\top\alpha_2-c_3^\top x_{30}-c_3^\top X_3z_1+\eta_2\geq 0,\ \forall\ (z_1,u_1)\in\Omega_1,\\G_2^\top\alpha_2=E_2^\top\xi_2-\bar{X}_3^\top c_3,\ H_2^\top\alpha_2=F_2^\top\xi_2,\ \alpha_2\in\mathcal{K}_2^*\ .\end{cases}\tag{3.9}$$

The first constraint in (3.9) is equivalent to

$$\min\{h_2^\top\alpha_2-c_3^\top x_{30}-c_3^\top X_3z_1+\eta_2:E_1z_1+F_1u_1\succeq_{\mathcal{K}_1}h_1\}\geq 0,$$

which, by Lemma 2.3 can be equivalently replaced by that $\exists\ \beta_1\in\mathcal{K}_1^*$ such that

$$\begin{cases}h_1^\top\beta_1+h_2^\top\alpha_2-c_3^\top x_{30}+\eta_2\geq 0,\\G_1^\top\beta_1+X_3^\top c_3=0,\ H_1^\top\beta_1=0,\end{cases}$$

which completes the proof. $\square$

Define

$$\psi_1(x_1, z_1, u_1) := \min_{x_2 \in \mathcal{X}_2} \{g_2^\top \xi_2 + \eta_2 + c_2^\top x_2\},$$

and repeat the analysis from Lemma 3.1 to Theorem 3.1 for ψ_1 and problem (3.4), we may come up with the following main result of this paper. For brevity, we omit the proof.

Theorem 3.2. Suppose that problem (DR-TSSLP) is feasible. Then, under the affine dependence assumption, the problem (DR-TSSLP) is equivalent to the following conic program, hence is solvable in polynomial time with respect to $(d_k, L_k, M_k, p_k, r_k, t_k), k = 1, 2$.

$$\begin{array}{ll}
\min & c_1^\top x_1 + g_1^\top \xi_1 + \eta_1 \\
\text{s.t.} & h_1^\top \alpha_1 + \eta_1 - \eta_2 - g_2^\top \xi_2 - c_2^\top x_{20} \geq 0, \\
& h_2^\top \alpha_2 + g_2^\top \xi_2 + \eta_2 \geq 0, \\
& h_1^\top \beta_1 + h_2^\top \alpha_2 - c_3^\top x_{30} + \eta_2 \geq 0, \\
& G_1^\top \alpha_1 = E_1^\top \xi_1 - X_2^\top c_2, \ H_1^\top \alpha_1 = F_1^\top \xi_1, \\
& G_2^\top \alpha_2 = E_2^\top \xi_2 - \bar{X}_3^\top c_3, \ H_2^\top \alpha_2 = F_2^\top \xi_2, \\
& G_1^\top \beta_1 + X_3^\top c_3 = 0, \ H_1^\top \beta_1 = 0, \\
& \alpha_1 \in \mathcal{K}_1^*, \ \alpha_2 \in \mathcal{K}_2^*, \ \beta_1 \in \mathcal{K}_1^*, \\
& x_1 \geq 0, \ A_1 x_1 = b_1, \\
& A_{2j} x_{1j} + B_2 x_{2j} = b_{2j}, \ j = 0, 1, \ldots, r_1, \\
& A_{3j} x_1 + B_3 x_{2j} + C_3 x_{3j} = b_{3j}, \ j = 0, 1, \ldots, r_1, \\
& \bar{A}_{3j} \bar{x}_1 + C_3 \bar{x}_{3j} = \bar{b}_{3j}, \ j = 1, \ldots, r_2, \\
& x_{20}^q + h_1^\top s^q \geq 0, \ G_1^\top s^q = x_2^q, \ H_1^\top s^q = 0, \ s^q \in \mathcal{K}_1^*, \ q = 1, \ldots, d_1, \\
& x_{30}^q + h_1^\top t_1^q + h_2^\top t_2^q \geq 0, \ G_1^\top t_1^q = x_3^q, \ G_2^\top t_2^q = \bar{x}_3^q, \ q = 1, \ldots, d_2, \\
& H_1^\top t_1^q = 0, \ H_2^\top t_2^q = 0, \ t_1^q \in \mathcal{K}_1^*, \ t_2^q \in \mathcal{K}_2^*, \ q = 1, \ldots, d_2.
\end{array} \tag{3.10}$$

4 Numerical Results

4.1 A Classroom Example

Example.[‡] A company manager is considering the amount of steel to purchase (at \$58/lb) for producing wrenches and pliers in next two months. The manufacturing process involves moulding the tools on a moulding machine and then assembling the tools on an assembly machine. Here are the technical data required for making the tools.

There are uncertainties that will influence his decision. 1. The total available moulding hours of next month ($\tilde{z}_{11}$) could be between 21,000 or 25,000 with mean of 23,000. 2. The total available assembly hours ($\tilde{z}_{12}$)of next month could be between 8,000 and 10,000 with mean of 9,000. 3. The total available moulding hours of next next month ($\tilde{z}_{21}$) could be between 23,000 and 27,000 with mean of 25,000, and the total available assembly hours of next next month ($\tilde{z}_{22}$) could be between 9,000 or 12,000 with mean of 10,500, respectively. 4. $\tilde{z}_1 = (\tilde{z}_{11}, \tilde{z}_{12})$ and $\tilde{z}_2 = (\tilde{z}_{21}, \tilde{z}_{22})$ are mutually independent random vectors. The manager would like to plan the production of wrenches and pliers of next two months so as to maximize the worst-case expected net revenue of the next two months.

[‡]The prototype of this example is Example 7.3 in the book of Bertsimas and Freund [7] and it was used in [2]. We use it again for comparison purpose.

	Wrench	Plier
Steel (lbs.)	1.5	1
Moulding Machine (hours)	1	1
Contribution to Earnings ($/1000 units)	1300	1000

Table 4.1: Cost and earnings for the products

Scenario	Moulding	Assembly	Probability
1	25000	10000	.25
2	25000	8000	.25
3	21000	10000	.25
4	21000	8000	.25

Table 4.2: Scenarios in Stage 2

For easy comparison, we also construct another three-stage model where the probability of each scenario is exactly known. In other words, the information on the distribution is fully known. For fair comparison purpose, the mean of moulding time and the mean of assembly time in each stage are the same as that in the above example. Particularly, the second-stage information is the same as that in [2]. The details each scenario in the second-stage and the third-stage are listed in Table 4.2 and Table 4.3, respectively. We refer to this model as stochastic model in this paper. This problem can be formulated as a linear optimization problem and can be solved by CVX [14]. We omit the formulation for brevity.

$\boxed{4.2}$ DR-TSSLP formulation and its conic reformulation with first-order moment information

In this section, we will formulate the steel purchase problem as a DR-TSSLP. The DR-TSSLP will, then, be reformulated to a conic optimization problem by applying Theorem 3.2. In our paper, all the cone optimization problems are numerically solved by the well-known optimization software package CVX [14].

We set up our decision variables as follows: y_1 is the amount of steel to purchase in stage 1; w_1 and p_1 are the number of wrenches and pilers to produce in stage 2; y_2 is the amount of steel to purchase in stage 2; w_3 and p_3 are the number of wrenches and pilers to produce in stage 3. Here, the unit for moulding and assembly hours is 1000 hours.

In order to formulate the example into a standard DR-TSSLP as that in (1.2), in stage

Scenario	Moulding	Assembly	Probability
1 (1), 5 (2), 9 (3), 13 (4)	27000	12000	.1250
2 (1), 6 (2), 10 (3), 14 (4)	27000	9000	.0625
3 (1), 7 (2), 11 (3), 15 (4)	23000	12000	.0417
4 (1), 8 (2), 12 (3), 16 (4)	23000	9000	.0208

Table 4.3: Scenarios in Stage 3 (The number in the bracket shows the scenario it is branching from in Stage 2)

2, we introduce 3 slack variables $\tau_{1i},\ i=1,2,3$ and 2 random variables $z_{1i},\ i=1,2$ for the mould constraint, the assembly constraint and the steel constraint, which yields the following equality constraints. In fact, τ_{13} is the steel left in stage 2. The steel left can be reused in the third-stage for production. However, we consider the cost cs associated with the stock of the left steel in the second-stage

$$\begin{aligned} w_1+p_1+\tau_{11} &= \tilde{z}_{11}, \\ .3w_1+.5p_1+\tau_{12} &= \tilde{z}_{12}, \\ -y_1+1.5w_1+p_1+\tau_{13} &= 0. \end{aligned} \tag{4.1}$$

Similarly, in stage 3, we introduce 3 slack variables $\tau_{2i},\ i=1,2,3$ and 2 random variables $z_{2i},\ i=1,2$ for the mould constraint, the assembly constraint and the steel constraint, which yields the following equality constraints. In fact, τ_{23} is the steel left in stage 3,

$$\begin{aligned} w_2+p_2+\tau_{21} &= \tilde{z}_{21}, \\ .3w_2+.5p_2+\tau_{22} &= \tilde{z}_{22}, \\ -y_2-\tau_{13}+1.5w_2+p_2+\tau_{23} &= 0. \end{aligned} \tag{4.2}$$

By defining

$$x_1=y_1,\ x_2=\begin{bmatrix} w_1 & p_1 & \tau_{11} & \tau_{12} & \tau_{13} & y_2\end{bmatrix}^\top,\ x_3=\begin{bmatrix} w_2 & p_2 & \tau_{21} & \tau_{22} & \tau_{23}\end{bmatrix}^\top,$$

we can formulate this example into the form of (1.2) with corresponding coefficient matrices and vectors chosen as

$$A_1=0,\ b_1=0,\ A_2=\begin{bmatrix}0\\0\\-1\end{bmatrix},\ B_2=\begin{bmatrix}1&1&1&0&0&0\\ .3&.5&0&1&0&0\\ 1.5&1&0&0&1&0\end{bmatrix},$$

$$b_2(\tilde{z}_1)=\begin{bmatrix}\tilde{z}_{11}\\ \tilde{z}_{12}\\ 0\end{bmatrix}, A_3=\begin{bmatrix}0\\0\\0\end{bmatrix},\ B_3=\begin{bmatrix}0&0&0&0&0&0\\0&0&0&0&0&0\\0&0&0&0&-1&-1\end{bmatrix},$$

$$C_3=\begin{bmatrix}1&1&1&0&0\\ .3&.5&0&1&0\\ 1.5&1&0&0&1\end{bmatrix}, b_3(\tilde{z}_1,\tilde{z}_2)=\begin{bmatrix}\tilde{z}_{21}\\ \tilde{z}_{22}\\ 0\end{bmatrix},\ c_1=58,$$

$$c_2=\begin{bmatrix}-130 & -100 & 0 & 0 & cs & 58\end{bmatrix}^\top,\ c_3=\begin{bmatrix}-130 & -100 & 0 & 0 & 0\end{bmatrix}^\top.$$

With respect to the linear decision rule, we have

$$A_{20}=\begin{bmatrix}0\\0\\-1\end{bmatrix},\ A_{21}=A_{22}=\begin{bmatrix}0\\0\\0\end{bmatrix}, b_{20}=\begin{bmatrix}0\\0\\0\end{bmatrix},\ b_{21}=\begin{bmatrix}1\\0\\0\end{bmatrix},\ b_{22}=\begin{bmatrix}0\\1\\0\end{bmatrix},$$

$$A_{30}=A_{31}=A_{32}=\bar{A}_{31}=\bar{A}_{32}=\begin{bmatrix}0\\0\\0\end{bmatrix},\ b_{30}=b_{31}=b_{32}=\begin{bmatrix}0\\0\\0\end{bmatrix},$$

$$\bar{b}_{31}=\begin{bmatrix}1\\0\\0\end{bmatrix},\ \bar{b}_{32}=\begin{bmatrix}0\\1\\0\end{bmatrix}.$$

Then, we construct the ambiguity sets as defined in (2.3). Based on the first-order information (mean) on the uncertainties of moulding hours and assembly hours, as mentioned in Subsection 4.1, we known that

$$\mathbb{E}(\tilde{z}_{11}) = 23,\ \mathbb{E}(\tilde{z}_{12}) = 9,\ \mathbb{E}(\tilde{z}_{21}) = 25,\ \mathbb{E}(\tilde{z}_{22}) = 10.5,$$

$$\mathbb{E}(\tilde{z}_{11}^2) \leq 533,\ \mathbb{E}(\tilde{z}_{12}^2) \leq 82,\ \mathbb{E}(\tilde{z}_{21}^2) \leq 629,\ \mathbb{E}(\tilde{z}_{22}^2) \leq 112.5.$$

$$21 \leq z_{11} \leq 25,\ 8 \leq z_{12} \leq 10,\ 23 \leq z_{21} \leq 27,\ 9 \leq z_{22} \leq 12.$$

To formulate the uncertainties with the second-order moment information into the WKS-type ambiguity set, we need the following result, which is a special case of Theorem 5 in [22].

Lemma 4.1 (Lifting Theorem)**.** Let $f \in \mathbb{R}^T$ and let $f : \mathbb{R}^m \to \mathbb{R}^T$ be a function with a conic representable $\mathcal{K}$-epigraph. Consider the ambiguity set

$$\mathcal{P}' = \{\mathbb{P}' \in \mathcal{P}_0(\mathbb{R}^m) : \mathbb{E}_{\mathbb{P}'}[g(\tilde{z})] \preceq_{\mathcal{K}} f\} \tag{4.3}$$

and the lifted ambiguity set

$$\mathcal{P} = \left\{\mathbb{P} \in \mathcal{P}_0\left(\mathbb{R}^m \times \mathbb{R}^T\right) :\ \mathbb{E}_{\mathbb{P}}[\tilde{u}] = f, \mathbb{P}[g(\tilde{z}) \preceq_{\mathcal{K}} \tilde{u}] = 1\right\},$$

which involves the auxiliary random vector $\tilde{u} \in \mathbb{R}^T$. Then it follows that (i) $\mathcal{P}' = \prod_{\tilde{z}} \mathcal{P}$; and (ii) $\mathcal{P}$ is an instance of the standardized ambiguity set (2.3) and (2.4).

The lifting theorem has a great modeling power and it provides a significant flexibility to convert various ambiguity sets into the WKS-form as shown in [22]. In the following, we shall show how to use the Lemma 4.1 to formulate the uncertainties in our problem into the form of (2.3) and (2.4).

To begin, we use the available information to construct two ambiguity sets $\mathcal{P}_1'$ and $\mathcal{P}_{2|1}'$ as (4.3) in Lemma 4.1, which can be done straightforwardly.

$$\begin{aligned}\mathcal{P}_1' = \{\mathbb{P}' \in \mathcal{P}_0(\mathbb{R}^m) : \mathbb{P}'[\Omega_1'] = 1,\ \mathbb{E}_{\mathbb{P}'}(\tilde{z}_{11}) \leq 23,&\\ \mathbb{E}_{\mathbb{P}'}(\tilde{z}_{12}) \leq 9,\ \mathbb{E}_{\mathbb{P}'}\left(\tilde{z}_{11}^2\right) \leq 533,\ \mathbb{E}_{\mathbb{P}'}\left(\tilde{z}_{12}^2\right) \leq 82\right\},&\end{aligned}$$

where

$$\Omega_1' = \left\{\begin{pmatrix} z_{11} \\ z_{12} \end{pmatrix} : \begin{matrix} 21 \leq z_{11} \leq 25, \\ 8 \leq z_{12} \leq 10 \end{matrix}\right\},$$

$$\begin{aligned}\mathcal{P}_2' = \{\mathbb{P}' \in \mathcal{P}_0(\mathbb{R}^m) : \mathbb{P}'[\Omega_2'] = 1,\ \mathbb{E}_{\mathbb{P}'}(\tilde{z}_{21}) \leq 25,&\\ \mathbb{E}_{\mathbb{P}'}(\tilde{z}_{22}) \leq 10.5,\ \mathbb{E}_{\mathbb{P}'}\left(\tilde{z}_{21}^2\right) \leq 629,\ \mathbb{E}_{\mathbb{P}'}\left(\tilde{z}_{22}^2\right) \leq 112.5\right\},&\end{aligned}$$

where

$$\Omega_2' = \left\{\begin{pmatrix} z_{21} \\ z_{22} \end{pmatrix} : \begin{matrix} 23 \leq z_{21} \leq 27, \\ 9 \leq z_{22} \leq 12 \end{matrix}\right\}.$$

By applying Lemma 4.1 to $\mathcal{P}_1'$ and $\mathcal{P}_2'$, we obtain the following two lifted ambiguity sets $\mathcal{P}_1$ and $\mathcal{P}_2$.

$$\mathcal{P}_1 = \left\{\mathbb{P} \in \mathcal{P}_0(\mathbb{R}^m) :\ \mathbb{E}_{\mathbb{P}}[E_1\tilde{z}_1 + F_1\tilde{u}_1] = g_1,\ \mathbb{P}\left[\bar{\Omega}_1\right] = 1\right\},$$

where

$$E_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, F_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, g_1 = \begin{bmatrix} 23 \\ 9 \\ 533 \\ 82 \end{bmatrix},$$

and

$$\bar{\Omega}_1 = \left\{ (z_1, u_1) : \begin{array}{ll} 21 \le z_{11} \le 25, & u_{11} \ge z_{11}^2 \\ 8 \le z_{12} \le 10, & u_{12} \ge z_{12}^2 \end{array} \right\},$$

$$\mathcal{P}_2 = \left\{ \mathbb{P} \in \mathcal{P}_0 (\mathbb{R}^m) : \ \mathbb{E}_{\mathbb{P}} [E_2 \tilde{z}_2 + F_2 \tilde{u}_2] = g_2, \ \mathbb{P} \left[\bar{\Omega}_2 \right] = 1 \right\},$$

where

$$E_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, F_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, g_2 = \begin{bmatrix} 25 \\ 10.5 \\ 629 \\ 112.5 \end{bmatrix}.$$

$E_3 = F_3$ are both 4×2 zero matrices and

$$\bar{\Omega}_2 = \left\{ (z_2, u_2) : \begin{array}{ll} 23 \le z_{21} \le 27, & u_{21} \ge z_{21}^2 \\ 9 \le z_{22} \le 12, & u_{22} \ge z_{22}^2 \end{array} \right\}.$$

Noting that

$$\begin{aligned} \left\{ (z_{11}, u_{11}) : z_{11}^2 \le u_{11} \right\} &= \left\{ (z_{11}, u_{11}) : \left\| \begin{matrix} z_{11} \\ \frac{u_{11}-1}{2} \end{matrix} \right\|_2 \le \frac{u_{11}+1}{2} \right\} \\ = \quad & \left\{ (z_{11}, u_{11}) : \begin{bmatrix} z_{11} \\ \frac{u_{11}-1}{2} \\ \frac{u_{11}+1}{2} \end{bmatrix} \in \mathbb{L}^3 \right\} \\ = \quad & \left\{ (z_{11}, u_{11}) : \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} z + \begin{bmatrix} 0 & 0 \\ 1/2 & 0 \\ 1/2 & 0 \end{bmatrix} u \succeq_{\mathbb{L}^3} \begin{bmatrix} 0 \\ 1/2 \\ -1/2 \end{bmatrix} \right\}, \end{aligned}$$

where $\mathbb{L}^3$ is the 3-dimensional Lorenz cone. Then, by defining the above set as $\mathcal{L}_1^3$ and letting $\mathcal{K}_1 = \mathbb{R}^4 \times \mathbb{L}^3 \times \mathbb{L}^3$, we obtain an equivalent set of $\bar{\Omega}_1$ as follows

$$\Omega_1 = \{ (z_1, u_1) : G_1 z_1 + H_1 u_1 \succeq_{\mathcal{K}_1} h_1 \},$$

where

$$G_1 = \begin{bmatrix} G_{11} \\ G_{12} \\ G_{13} \end{bmatrix}, H_1 = \begin{bmatrix} H_{11} \\ H_{12} \\ H_{13} \end{bmatrix}, h_1 = \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \end{bmatrix}, G_{11} = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix},$$

$$H_{11} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, h_{11} = \begin{bmatrix} 21 \\ -25 \\ 8 \\ -10 \end{bmatrix}, G_{12} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, H_{12} = \begin{bmatrix} 0 & 0 \\ 1/2 & 0 \\ 1/2 & 0 \end{bmatrix},$$

$$h_{12} = \begin{bmatrix} 0 \\ 1/2 \\ -1/2 \end{bmatrix}, G_{13} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, H_{13} = \begin{bmatrix} 0 & 0 \\ 0 & 1/2 \\ 0 & 1/2 \end{bmatrix}, h_{13} = \begin{bmatrix} 0 \\ 1/2 \\ -1/2 \end{bmatrix}.$$

In a similar manner, we can obtain an equivalent set of $\bar{\Omega}_2$ as follows.

$$\Omega_2 = \{ (z_2, u_2) : G_2 z_2 + H_2 u_2 \succeq_{\mathcal{K}_2} h_2 \},$$

where $\mathcal{K}_2 = \mathcal{K}_{21} \times \mathcal{K}_{22} \times \mathcal{K}_{23}$, $\mathcal{K}_{21} = \mathbb{R}^4$, $\mathcal{K}_{22} = \mathcal{K}_{23} = \mathbb{L}^3$,

$$G_2 = \begin{bmatrix} G_{21} \\ G_{22} \\ G_{23} \end{bmatrix},\ H_2 = \begin{bmatrix} H_{11} \\ H_{22} \\ H_{23} \end{bmatrix},\ h_2 = \begin{bmatrix} h_{21} \\ h_{22} \\ h_{23} \end{bmatrix},\ G_{21} = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix},$$

$$H_{21} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},\ h_{21} = \begin{bmatrix} 23 \\ -27 \\ 9 \\ -12 \end{bmatrix},\ G_{22} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},\ H_{22} = \begin{bmatrix} 0 & 0 \\ 1/2 & 0 \\ 1/2 & 0 \end{bmatrix},$$

$$h_{22} = \begin{bmatrix} 0 \\ 1/2 \\ -1/2 \end{bmatrix},\ G_{23} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},\ H_{23} = \begin{bmatrix} 0 & 0 \\ 0 & 1/2 \\ 0 & 1/2 \end{bmatrix},\ h_{23} = \begin{bmatrix} 0 \\ 1/2 \\ -1/2 \end{bmatrix}.$$

Now the problem is formulated in the standard form stated in Section 1 and 2 with the same notations. Applying Theorem 3.2 to the formulated problem, we obtain the following second order cone optimization problem.

$$\begin{array}{ll}
\min & c_1 x_1 + g_1^\top \xi_1 + \eta_1 \\
\text{s.t.} & h_1^\top \alpha_1 + \eta_1 - \eta_2 - g_2^\top \xi_2 - c_2^\top x_{20} \ge 0, \\
& h_1^\top \beta_1 + h_2^\top \alpha_2 - c_3^\top x_{30} + \eta_2 \ge 0, \\
& G_1^\top \alpha_1 = E_1^\top \xi_1 - X_2^\top c_2;\ H_j^\top \alpha_j = F_j^\top \xi_j,\ j = 1, 2, \\
& G_2^\top \alpha_2 = E_2^\top \xi_2 - \bar{X}_3^\top c_3;\ G_1^\top \beta_1 + X_3^\top c_3 = 0, \\
& H_1^\top \beta_1 = 0;\ A_{20} x_1 + B_2 x_{20} = 0, \\
& B_2 x_{2j} = b_{2j},\ j = 1, 2;\ B_3 x_{2j} + C_3 x_{3j} = 0,\ j = 0, 1, 2, \\
& C_3 \bar{x}_{3j} = \bar{b}_{3j},\ j = 1, 2;\ x_{20}^q + h_1^\top s^q \ge 0,\ q = 1, 2, \ldots, 6, \\
& G_1^\top s^q = x_2^q,\ q = 1, 2, \ldots, 6;\ H_1^\top s^q = 0,\ q = 1, 2, \ldots, 6, \\
& x_{30}^q + h_1^\top t_1^q + h_2^\top t_2^q \ge 0,\ q = 1, 2, \ldots, 5, \\
& G_1^\top t_1^q = x_3^q,\ q = 1, 2, \ldots, 5, \\
& G_2^\top t_2^q = \bar{x}_3^q,\ q = 1, 2, \ldots, 5, \\
& H_j^\top t_j^q = 0,\ j = 1, 2,\ q = 1, 2, \ldots, 5, \\
& \alpha_{12},\ \alpha_{13},\ \alpha_{22},\ \alpha_{23},\ \beta_{12},\ \beta_{13} \in \mathbb{L}^3, \\
& s_2^q,\ s_3^q,\ t_{12}^q,\ t_{13}^q,\ t_{22}^q,\ t_{23}^q \in \mathbb{L}^3,\ q = 1, 2, \ldots, 5, \\
& \alpha_{11} \ge 0,\ \alpha_{21} \ge 0,\ \beta_{11} \ge 0,\ x_1 \ge 0,\ s_1 \ge 0,\ t_{11} \ge 0,\ t_{21} \ge 0.
\end{array} \tag{4.4}$$

Problem (4.4) is a second order cone programming (SOCP) problem, which can be solved efficiently by using which can be solved by using [14].

4.3 Comparisons and discussions

The major difference between the three-stage model and the second-stage model [2] is that the steel left in the second-stage τ_{13} can be reused in the third-stage. Therefore, we set the stock cost cs of τ_{13} with different values and then solve (4.4). Then, we compare our results with that in the stochastic model, and we also compare it with the stochastic model and the proposed model in [2]. The details of the results are summarized in Table 4.4 and Table 4.5.

	(4.4)	Stochastic	Stochastic [2]	[2]
Optimal x_1^* (1000 lb)	37.5	37.5	31.5	30.5
Expected Profits ($)	2021.67	2078.33	961.89	929.88

Table 4.4: Results Comparison, $cs = 1$

	(4.4)	Stochastic	Stochastic [2]	[2]
Optimal x_1^* (1000 lb)	31.5	31.5	31.5	30.5
Expected Profits ($)	1976.44	2054.22	961.89	929.88

Table 4.5: Results Comparison, $cs = 50$

From Table 4.4 and Table 4.5, we can see as the cost on stocking the left steel increases the solution becomes more conservative and it reduces to the second-stage decision when the cost is high enough. In addition, as expected, the expected profits for three-stage distributionally robust model is less than that of the three-stage stochastic model.

References

[1] M. Ang, J. Sun and Q. Yao, On dual representation of coherent risk measures, *Ann. Oper. Res.* 262 (2018) 29–46.

[2] J. Ang, F. Meng and J. Sun, Two-stage stochastic linear programs with incomplete information on uncertainty, *Eur. J. Oper. Res.* 233 (2014) 16–22.

[3] P. Artzner, F. Delbaen, J.M. Eber and D. Heath, Coherent measures of risk, *Math. Finan.* 9 (1999) 203–227.

[4] A. Bental and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, MOS-SIAM Series on Optimization, 2001.

[5] A. Bental and A. Nemirovski, Robust optimization-methodology and applications, *Math. Program.* 92 (2002) 453–480.

[6] D. Bertsimas, X.V. Duan, K. Natarajan and C.P. Teo, Model for minimax stochastic linear optimization problems with risk aversion, *Math. Oper. Res.* 35 (2010) 580–602.

[7] D. Bertsimas and R. Freund, *Data, Models, and Decisions: The Fundamentals of Management Science*, South-Western College Publishing, Cincinnati, 2000.

[8] D. Bertsimas, M. Sim and M. Zhang. A practicable framework for distributionally robust linear optimization problems, appeared at http://www.optimization-online.org/DB FILE/2013/07/3954.pdf

[9] W. Chen, M. Sim, J. Sun and C.P. Teo, From CVaR to uncertainty set: implications in joint chance constrained optimization, *Oper. Res.*. 58 (2010) 470–485.

[10] X. Chen, M. Sim, P. Sun and J. Zhang, A linear-decision based approximation approach to stochastic programming, *Oper. Res.* 56 (2008) 344–357.

[11] H. Föllmer and A. Schied, *Stochastic Finance*, Walter de Gruyter, Berlin, 2002.

[12] S. Gao, L. Kong and J. Sun, Two-stage stochastic linear programs with moment information on uncertainty, *Optimization* 63 (2014) 829–837.

[13] J. Goh and M. Sim, Distributionally robust optimization and its tractable approximations, *Oper. Res.* 58 (2010) 902–917.

[14] M. Grant and S. Boyd, CVX: Matlab software for disciplined convex programming, Apr. 2011, http://cvxr.com/cvx.

[15] R. Hettich and K.O. Kortanek, Semi-infinite programming: theory, methods, and applications, *SIAM Rev.* 35 (1994) 380–429.

[16] M.J. Hadjiyiannis, P.J. Goulart and D. Kuhn, An efficient method to estimate the suboptimality of affine controllers,*IEEE Trans. Auto. Control.* 56 (2011) 2841–2853.

[17] A. Ling, J. Sun and X. Yang, Robust tracking error portfolio selection with worst-case downside risk measures, *J. Econam. Dynam. Control.* 39 (2014) 178–207.

[18] R.T. Rockafellar, *Conjugate Duality and Optimization*, AMS-SIAM Publication, Philadelphia, 1974.

[19] R.T. Rockafellar,*Coherent Approaches to Risk in Optimization under Uncertainty*, NFORMS Tutorials Oper. Res., Hanover, 2007.

[20] M. Sion, On general minimax theorems, *Pacific J. Math.* 8 (1958) 171–176.

[21] J. Sun,K. Tsai and L. Qi, A simplex method for network programs with convex separable piecewise linear costs and its application to stochastic transshipment problems, in: *Network Optimization Problems: Algorithms, Applications and Complexity*, D.Z. Du and P.M. Pardalos (eds), World Scientific Publishing Co, London, 1993, pp. 281–300.

[22] W. Wiesemman, D. Kuhn and M. Sim, Distributionally robust convex optimization, *Oper. Res.* 62 (2014) 1358–1376.

Manuscript received 3 December 2018
revised 12 December 2018
accepted for publication 14 December 2018

Bin Li
College of Electrical Engineering and Information Technology
Sichuan University, Chengdu, 610065, China
E-mail address: bin.li@scu.edu.cn

Jie Sun
School of EECMS, Curtin University
Perth, 6845, Australia
E-mail address: jie.sun@curtin.edu.au

Kok Lay Teo
School of EECMS, Curtin University
Perth, 6845, Australia
E-mail address: k.l.teo@curtin.edu.au

Changjun Yu
College of Sciences, Shanghai University
Shanghai, 200444, China
E-mail address: yuchangjun@126.com

Min Zhang
Xinjiang Institute of Ecology and Geography
Chinese Academy of Sciences, Urumqi, 830011, China
E-mail address: min.zhang2@curtin.edu.au