Optimization Letters manuscript No.
(will be inserted by the editor)

A Semi-infinite Programming Approach to Two-stage Stochastic Linear Programs with High-order Moment Constraints

Sarah Yini Gao · Jie Sun · Soon-Yi Wu

Received: date / Accepted: date

Abstract We consider distributionally robust two-stage stochastic linear optimization problems with higher-order (say $p \geq 3$ and even possibly irrational) moment constraints in their ambiguity sets. We suggest to solve the dual form of the problem by a semi-infinite programming approach, which deals with a much simpler reformulation than the conic optimization approach. Some preliminary numerical results are reported.

Keywords Semi-infinite optimization · Stochastic programming

Mathematics Subject Classification (2010) 90C15 · 90C47

1 Introduction

Many decision-making problems that involve uncertainty are modeled as stochastic programs. In general, stochastic optimization models require detailed information on the probability distribution of the random variables. Under such assumptions, the decision makers seek to minimize the aggregated expected cost over the multi-stage planning horizon. In order to solve the stochastic optimization problems, one often resorts to Monte Carlo sampling approximation approaches, which can be very challenging in practice. Motivated by recent development in robust optimization, a new model of two-stage stochastic programming is proposed, in which the second-stage objective is a worst-case

S. Gao
Department of Decision Sciences, National University of Singapore E-mail: yini.gao@nus.edu.sg

J. Sun
School of Science, Curtin University, Australia and School of Business, National University of Singapore E-mail: jie.sun@curtin.edu.au

S.Y. Wu
Department of Mathematics, National Cheng Kung University E-mail: soony@math.ncku.edu.tw

recourse function subject to constraints in first and second-order moments of the underlying random variables. Typically, these constraints specify a range for the moments rather than specify a distribution for the random variables, thus greatly reducing the computational load. In fact, the worst-case models can be often reduced to second-order cone or semidefinite programs. However, due to nonconvexity, this "distributionally robust approach" fails to work when third or higher order moment constraints are introduced. There are rich literatures in distributionally robust models. For early literatures on distributionally robust models see Scarf (1958), Landau (1987), Dupacova (1987), Kall and Wallace (1994), while Bertsimas et. al. (2013), Delage and Ye (2010), Ang et. al. (2014), and Wiesemann et. al. (2014) provide more recent development.

The purpose of this paper is three fold. First, we propose to use a semi-infinite programming (SIP) method to solve distributionally robust two-stage stochastic linear programs with moment constraints of arbitrary pth order ($p \geq 1$) in the definition of their ambiguity sets. Second, we present a simple analysis to convert the stochastic program to an SIP problem. Third, we test our method with a numerical example and compare the solutions to the case without higher-order moment constraints to see whether higher-order moment information can significantly improve the quality of solutions.

It should be noted that the considered distributionally robust problem can be reformulated as conic optimization problem in a recent study of Ang et. al. (2014) and Bertsimas et al. (2013). However, they both assume up to second-order moments of random variables. When dealing with higher-order moment constraints, the reformulation involves a complex progressive decomposition procedure that can only handle moment constraints of rational order (i.e. $p = r/s$) (Chapter 2 of Ben-Tal and Nemirovski 2001). Our paper extends their results in a way that the SIP algorithm proposed can deal with higher-order moment information. The SIP conversion is straightforward for any $p \geq 1$, and there are plenty of choices for possible packages of SIP solvers.

The rest of this paper is organized as follows. In Section 2, we establish the optimization model of the distributionally robust two-stage stochastic programming problem with moment information and show its equivalence to SIP problem under the so-call linear decision rule. In addition, we show that in the simple case where the second-stage objective parameters are the only uncertain terms, the problem can be also cast to SIP without additional assumptions. Section 3 contains numerical results with certain interesting observations.

Notations. We denote a random vector, say $\tilde{z}$, with the tilde sign. Matrices and vectors are usually represented as upper and lower case letters, respectively. If x is a vector, we use the notation x_i to denote the ith component of the vector. A random vector is associated with its support Ω and a probability distribution $\mathbb{P}$ on a σ-algebra Σ of events. We use $\mathbb{E}_{\mathbb{P}}(\tilde{z}_i^p)$ to denote the pth-order moments of $\tilde{z}_i$ under $\mathbb{P}$, where $\tilde{z}_j$ are continuous random variables.

2 The Distributionally Robust Two-stage Stochastic Linear Program

We are concerned with the following two-stage stochastic programming problem with fixed recourse:

$$\min_{x\in X}\left\{c'x+\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}[Q(x,\tilde z)]\right\} \tag{1}$$

where the apostrophe ($'$) stands for the transpose and

$$\begin{aligned} Q(x,\tilde z)=\min\ \ & d'y(\tilde z)\\ \text{s. t.}\ \ & A(\tilde z)x+Dy(\tilde z)=b(\tilde z),\\ & y(\tilde z)\ge 0, \end{aligned} \tag{2}$$

where $Q(x,\tilde z)$ is the recourse cost, $x\in\mathbb{R}^n$ is the vector of first-stage decision variables in a feasible polyhedron X, $d\in\mathbb{R}^k$, $b(\tilde z)\in\mathbb{R}^l$, $A(\tilde z)\in\mathbb{R}^{l\times n}$ are second-stage data, and $D\in\mathbb{R}^{l\times k}$ represents the fixed recourse matrix. Moreover, $\tilde z$ is a random vector with support $\Omega\subset\mathbb{R}^m$ and $\mathbb{P}$ is the probability distribution of $\tilde z$ subject to an ambiguity set $\mathcal{F}$ defined by certain moment constraints.

A basic condition is imposed on problem (2). We assume for all $x\in X$ the feasible set of (2) is not empty for otherwise the optimal value of (1) is trivially infinite. This condition can be guaranteed by certain assumptions on the dual problem of (2), say, the dual feasible set is bounded. In practice, this assumption is naturally valid, see Section 3 for an example.

The advantage of the distributionally robust approach, compared with the traditional approach is that it does not require that $\mathbb{P}$ is exactly known and computationally, it is more tractable.

2.1 Assumptions on Distributions of $\tilde z$

Since the information on moments of $|\tilde z_i|$ are relatively easy to estimate from statistical data, we assume that the p-th order moment of $|\tilde z_i|$ exists. In particular, let $\mathcal{F}$ denote the family of probability distributions of $\tilde z$ defined as

$$\mathcal{F}:=\{\mathbb{P}:\mathbb{P}(\tilde z\in\Omega)=1,\ \mathbb{E}_{\mathbb{P}}(|\tilde z_j|^p)\le\mu_{pj},\ p=p_{j1},...,p_{jP}\ \ j=1,\dots,m\}, \tag{3}$$

where μ_{pj}s are given constants. Note that, for simplicity of analysis, without loss of generality we assume that P does not depend on j, i.e., each $|\tilde z_j|$ has the same number of moment constraints, which can be achieved by adding redundant constraints like $\mathbb{E}_{\mathbb{P}}(|\tilde z_j|^p)\le K$ for sufficiently large K. It will be seen that our analysis is independent of the concrete values of p as long as $p\ge 1$.

The absolute value $|z_j|$ is necessary because it makes the constraints convex. In addition, we assume that Ω is a nonnegative box and $\operatorname{int}\Omega\neq\emptyset$.

2.2 Reformulation under the Linear Decision Rule

We assume the uncertain data $b(\tilde{z})$ and $A(\tilde{z})$, together with the vector $y(\tilde{z})$, in (1) are affinely dependent on the random vector $\tilde{z}$, namely

$$y(\tilde{z}) = y^0 + \sum_{j=1}^{m} \tilde{z}_j y^j, \quad b(\tilde{z}) = b^0 + \sum_{j=1}^{m} \tilde{z}_j b^j, \quad \text{and} \quad A(\tilde{z}) = A_0 + \sum_{j=1}^{m} \tilde{z}_j A_j, \tag{4}$$

where, $b^j \in \mathbb{R}^l$, and $A_j \in \mathbb{R}^{l \times n}$, $j = 0, 1, \ldots, m$, are deterministic values given in advance.

The above affine-dependence assumption, also called *the linear decision rule*, is often adopted in dealing with the uncertainties in robust optimization models. See, e.g., Ben-Tal et al.(2004). Chen et al. (2008) used it in the context of robust stochastic programming. Chen et al. (2010) adopted it in dealing with joint chance constraints. It is easy to see that if int $\Omega \neq \emptyset$, then the following equivalence is valid.

$$A(\tilde{z})x + Dy(\tilde{z}) = b(\tilde{z}), \quad \forall \tilde{z} \in \Omega \iff A_j x + D y^j = b^j, \quad j = 0, 1, ..., m. \tag{5}$$

Note that, instead of x and $y(\tilde{z})$, the new decision variables are x and $y^0, y^1, ..., y^m$. By strong duality of linear programming, we obtain the following equivalence.

$$\begin{gathered} y(\tilde{z}) \geq 0,\ \forall \tilde{z} \in \Omega \iff \min \left[y_i^0 + \textstyle\sum_{j=1}^{m} z_j y_i^j \right] \geq 0\ \forall\ -\ell \leq z \leq h, \forall i \iff \\ \exists s_q, t_q \in \mathbb{R}_+^m \text{ such that } \ y_q^0 - \ell' s_q - h' t_q \geq 0 \text{ and } s_q - t_q = y_q, \quad \forall q = 1, ..., k, \end{gathered}$$

where y_q is the qth row of the matrix $[y^1, \cdots, y^m]$ and y_q^0 is the qth component of y^0.

Therefore, we have

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[Q(x, \tilde{z})] = \mathbb{E}_{\mathbb{P}}[\min_{y,s,t} \ & d' y^0 + \sum_{j=1}^{m} d' y^j \tilde{z}_j] \\ \text{s. t.} \ & A_j x + D y^j = b^j,\ j = 0, 1, ..., m, \\ & y_q^0 - \ell' s_q - h' t_q \geq 0,\ q = 1, ..., k, \\ & s_q - t_q = y_q,\ q = 1, ..., k, \\ & s_q, t_q \geq 0,\ q = 1, \ldots, k. \end{aligned} \tag{6}$$

In view of (6), the term $\sup_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[Q(x,\tilde{z})]$ is the optimal value of the following maximization problem

$$\begin{aligned}
\max_{\mathbb{P}} \;\; & \mathbb{E}_{\mathbb{P}}\left[\min_{y,s,t}\left(d'y^0+\sum_{j=1}^{m} d'y^j\tilde{z}_j\right)\right] \\
\text{s. t.} \;\; & A_jx+Dy^j=b^j,\;\; j=0,1,...,m, \\
& y_q^0-\ell' s_q-h't_q\geq 0,\;\; q=1,...,k, \\
& s_q-t_q=y_q\;\; q=1,...,k, \\
& \mathbb{E}_{\mathbb{P}}(|\tilde{z}_j|^p)\leq \mu_j^p,\;\; p=1,...,P\; j=1,\dots,m, \\
& \mathbb{P}\{\tilde{z}\in\Omega\}=1, \\
& s_q,t_q\geq 0,\;\; q=1,\dots,k.
\end{aligned} \tag{7}$$

Using the duality theory of linear optimization in probability spaces (see Rockafellar (1974), see also Vandenberghe et al. (2007) for some examples), the dual problem of (7) is

$$\begin{aligned}
\min_{y,s,t,v_0,v^1,...,v^P} \;\; & v_0+\sum_{r=1}^{k}(\mu_p)'(v_p) \\
& \text{where } \mu_p=(\mu_{p1},...,\mu_{pm})',\;\; v_p=(v_{p1},...,v_{pm})', p=1,...,P \\
\text{s. t.} \;\; & v_0+\sum_{p=1}^{P}(v^p)'(|z^p|)\geq \min_{y,s,t}\left(d'y^0+\sum_{j=1}^{m} d'y^j z_j\right),\;\; \forall\, z\in\Omega, \\
& \text{where } z^p:=(z_1^p,...,z_m^p)', |z^p|:=(|z_1^p|,...,|z_m^p|)', \\
& A_jx+Dy^j=b^j,\;\; j=0,1,...,m, \\
& y_q^0-\ell' s_q-h't_q\geq 0,\;\; q=1,...,k, \\
& s_q-t_q=y_q,\;\; q=1,...,k, \\
& s_q,t_q,v^p\geq 0,\;\; p=1,...,P, q=1,...,k,
\end{aligned} \tag{8}$$

where $v_0\in\mathbb{R}$ and $\forall p=1,...,P\; v_p=(v_{p1},...,v_{pm})'\in\mathbb{R}^m$ are the dual variables.

Theorem 1 *Under linear decision rule, the two-stage problem (1) can be written as*

$$\begin{aligned}
\min_{x,y,s,t,v_0,...,v^P} \;\; & c'x+v_0+\sum_{p=1}^{P}(\mu_p)'(v_p) \\
\text{s. t.} \;\; & v_0+\sum_{p=1}^{P}(v_p)'|z|^p\geq \min_{y,s,t}\left(d'y^0+\sum_{j=1}^{m} d'y^j z_j\right) \forall z\in\Omega, \\
& A_jx+Dy^j=b^j,\;\; j=0,1,...,m, \\
& y_q^0-\ell' s_q-h't_q\geq 0,\;\; q=1,...,k, \\
& s_q-t_q=y_q,\;\; q=1,...,k, \\
& s_q,t_q\geq 0,\;\; q=1,...,k,\;\; x\in X.
\end{aligned} \tag{9}$$

Proof. We prove this result by applying a generalized Slater condition. Specifically, strong duality holds between (7) and (8) due to the special structure of problem (8). To see this point, note that there exists an upper bound L, which does not depend on z, such that

$$L \geq \min_{y,s,t} \left(d'y^0 + \sum_{j=1}^{m} d'y^j z_j \right) \forall z \in \Omega$$

due to the compactness of Ω and the basic assumption that the recourse problem is feasible for any $x \in X$. Thus, by setting $v_0 = L + 1, v_p = 0$ and y_q^0 sufficiently large, the convex program (8) has a generalized Slater's point. Since the generalized Slater's condition is always valid for the dual problem (8). Then Theorem 18 of Rockafellar (1974) implies that

$$\inf(8) = \sup(7)$$

and the solution to (**??**) exists. Hence, problem (1) is equivalent to problem (9). □

Corollary 1 *Let*

$\Pi := \{(y, s, t) : \exists x \in X$ *such that the last four constraints in (9) are satisfied.*$\}$

Then Problem (9) is equivalent to

$$\left.\begin{array}{ll} \min_{x,y,s,t,v_0,\ldots,v_P} & c'x + v_0 + \sum_{p=1}^{P} (\mu_p)'(v_p) \\ \text{s. t.} & v_0 + \sum_{p=1}^{P} (v_p)'|z|^p \geq d'y^0 + \sum_{j=1}^{m} d'y^j z_j \forall z \in \Omega, \\ & A_j x + Dy^j = b^j,\ j = 0, 1, ..., m, \\ & y_q^0 - \ell' s_q - h' t_q \geq 0,\ q = 1, ..., k, \\ & s_q - t_q = y_q,\ q = 1, ..., k, \\ & s_q, t_q \geq 0,\ q = 1, ..., k,\ x \in X. \end{array}\right\} \quad (10)$$

if Π is not empty.

Proof. The first constraint of (9) can be written as follows.

$$\forall z \in \Omega, \quad \exists (y, s, t) \in \Pi : v_0 + \sum_{p=1}^{P} (v_p)'|z^p| - \left(d'y^0 + \sum_{j=1}^{m} d'y^j z_j \right) \geq 0,$$

or equivalently

$$\min_{z \in \Omega} \max_{(y,s,t) \in \Pi} \left\{ v_0 + \sum_{p=1}^{P} (v_p)'|z^p| - \left(d'y^0 + \sum_{j=1}^{m} d'y^j z_j \right) \right\} \geq 0.$$

The above function is convex in $\tilde{z}$ and is concave in (y,s,t) and both sets, Ω and Π, are closed and convex. By Sion's minimax theorem [17], as long as Ω or Π is compact, we have

$$\min_{z\in\Omega}\max_{(y,s,t)\in\Pi}\left\{v_0+\sum_{p=1}^{P}(v_p)'|z^p|-\left(d'y^0+\sum_{j=1}^{m}d'y^jz_j\right)\right\}$$

$$=\max_{(y,s,t)\in\Pi}\min_{z\in\Omega}\left\{v_0+\sum_{p=1}^{P}(v_p)'|z^p|-\left(d'y^0+\sum_{j=1}^{m}d'y^jz_j\right)\right\}.$$

The first constraint of (9) is therefore equivalent to

$$\exists(y,s,t)\in\Pi,\quad \forall z\in\Omega: v_0+\sum_{p=1}^{P}(v_p)'|z^p|-\left(d'y^0+\sum_{j=1}^{m}d'y^jz_j\right)\geq 0,$$

which proves the corollary. $\square$

Problem (10) is a linear SIP problem, in which the relationship $z\in\Omega$ defines an SIP constraints.

2.3 Reformulation without the Linear Decision Rule

In the case where only $d(\tilde{z})$ is uncertain, problem (1) reduces to

$$\min_{x\in X}\left\{c'x+\mathbb{E}_{\mathbb{P}}[Q(x,\tilde{z})]\right\}\text{ and }\quad \begin{aligned}Q(x,\tilde{z})=\min\ & \tilde{z}'y\\ \text{s t.}\ & Ax+Dy=b,\\ & y\geq 0,\end{aligned}\qquad (11)$$

where we directly write $d(\tilde{z})$ as $\tilde{z}$ without loss of generality. Since x is independent of $\mathbb{P}$, Problem (11) can be re-written as follows.

$$\begin{aligned}\min_{x\in X}\ & \left\{c'x+\max_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}[Q(x,\tilde{z})]\right\}\\ \text{s.t.}\ & Ax+Dy=b, y\geq 0\\ & \mathbb{E}_{\mathbb{P}}(|\tilde{z}_j|^p)\leq\mu_{pj},\ p=1,...,P,\ j=1,\ldots,m,\\ & \mathbb{P}\{\tilde{z}\in\Omega\}=1.\end{aligned}$$

By the strong duality theory of semi-infinite programming (Example 4 of Rockafellar 1974) and our assumption on Ω, problem (11) is equivalent to

$$\min_{v_0, v^1, \ldots, v^P, x, y} \quad c'x + v_0 + \sum_{p=1}^{P} (\mu_p)'(v_p) \tag{12}$$
$$\text{s. t.} \quad v_0 + \sum_{p=1}^{P} (v_p)'(|z^p|) \geq \min_y (z'y), \quad \forall\, z \in \Omega,$$
$$\text{where} \quad z^p = (z_1^p, ..., z_m^p)', \; |z^p| = (|z_1^p|, ..., |z_m^p|)',$$
$$Ax + Dy = b,$$
$$y \geq 0, v_p \geq 0, p = 1, ..., P.$$

where $v_0 \in \mathbb{R}$ and $v_p = (v_{p1}, ..., v_{pm})' \in \mathbb{R}^m$ are the dual variables.

Let

$$\Pi := \{(x, y) : Ax + Dy = b\}$$ [1]

Note that the set Ω is compact and int $\Omega \neq \emptyset$. Then Problem (12) is equivalent to the following SIP problem, in which the relationship $z \in \Omega$ defines an infinite number of constraints.

Theorem 2 *Problem (11) can be reformulated as a linear semi-infinite program as follows.*

$$\left.\begin{array}{ll} \min_{x,y,v_0,v^1,\ldots,v^P} & c'x + v_0 + \sum_{p=1}^{P} (\mu^p)'(v^p) \\ \text{s. t.} & v_0 + \sum_{p=1}^{P} (v^p)'|z^p| \geq (z'y)\, \forall z \in \Omega, \\ & Ax + Dy = b, \\ & y \geq 0, v^p \geq 0, p = 1, ..., P. \end{array}\right\} \tag{13}$$

Proof. Based on the analysis above, we only need to show that (11) is equivalent to (13). The proof is exactly the same as that of Theorem 1 and Corollary 1. □

3 Numerical Experiment – A Production Planning Example

This numerical example aims to demonstrate how the SIP model works on the classical two-stage stochastic programming problem. In particular, we examine how to incorporate higher-order moments information into decision making and further study the value of the additional information. The SIP algorithm we used is the exchange algorithm developed in Wu et al.(2005), which guarantee finite termination of the SIP reformulation of (1).

Example.[2] A company manager is considering the amount of steel to purchase (at \$58/1000lb) for producing wrenches and pliers in next month. The

[1] The set Π here is a simplified version of the set Π in Corollary 1.

[2] This is a slightly different version of Example 7.3 in the book of Bertsimas and Freund (2000). It is also the same production planning example appears in Ang et. al. (2014).

manufacturing process involves molding the tools on a molding machine and then assembling the tools on an assembly machine. Here is the technical data.

	Wrench	Plier
Steel (lbs.)	1.5	1
Molding Machine (hours)	1	1
Assembly Machine (hours)	.3	.5
Contribution to Earnings ($/1000 units)	130	100

There are uncertainties (continuously distributed) that will influence his decision. 1. The total available assembly hours (in thousand) of next month could be between 8,000 and 10,000, with expectation 9, second moments 82 and third moment 756. 2. The total available molding hours of next month could be between 21 and 25, with expectation 23, second moments 533, and third moment 12,443. The manager would like to plan, in addition to the amount of steel to purchase, for the production of wrenches and pliers of next month so as to maximize the worst-case expected net revenue (i.e. The worst-case expected earnings minus the cost of purchasing steel) of this company.

3.1 The SIP Formulation under the Linear Decision Rule

Now, we assume that the moments of uncertain molding hours and assembly hours are respectively known up to the third-order moment, namely

$$
\begin{aligned}
&\mathbb{E}_{\mathbb{P}}[|\tilde{z}_1|] = 23; \mathbb{E}_{\mathbb{P}}[|\tilde{z}_2|] = 9;\\
&\mathbb{E}_{\mathbb{P}}[|\tilde{z}_1|^2] = 533; \mathbb{E}_{\mathbb{P}}[|\tilde{z}_2|^2] = 82;\\
&\mathbb{E}_{\mathbb{P}}[|\tilde{z}_1|^3] = 12443; \mathbb{E}_{\mathbb{P}}[|\tilde{z}_2|^3] = 756;\\
&\Omega = \{z \in \mathbb{R}^2 : 21 \le z_1 \le 25,\ 8 \le z_2 \le 10\}.
\end{aligned}
$$

The SIP formulation of this problem is as follows:

$$\min_{x,y,s,t,v_0,v_1,\ldots,v_P} \quad 58x + v_0 + 23v_{11} + 9v_{12} + 533v_{21} + 82v_{22} + 12443v_{31} + 756v_{32}$$

$$\text{s.t.} \qquad -v_0 - \sum_{i=1}^{2} |z_i| v_{1i} - \sum_{i=1}^{2} |z_i|^2 v_{2i} - \sum_{i=1}^{2} |z_i|^3 v_{3i}$$
$$-(130y_1^0 + 100y_2^0) - \sum_{i=1}^{2}(130y_1^i + 100y_2^i)|z_i| \le 0, \quad \forall\, z \in \Omega$$

$$y_1^0 + y_2^0 + y_3^0 = 0, 0.3y_1^0 + 0.5y_2^0 + y_4^0 = 0, -x + 1.5y_1^0 + y_2^0 = 0$$
$$y_1^1 + y_2^1 + y_3^1 = 1, 0.3y_1^1 + 0.5y_2^1 + y_4^1 = 0, 1.5y_1^1 + y_2^1 = 0$$
$$y_1^2 + y_2^2 + y_3^2 = 0, 0.3y_1^2 + 0.5y_2^2 + y_4^2 = 1, 1.5y_1^2 + y_2^2 = 0$$
$$y_1^3 + y_2^3 + y_3^3 = 0, 0.3y_1^3 + 0.5y_2^3 + y_4^1 = 0, 1.5y_1^3 + y_2^3 = 1$$
$$-21s_1^1 - 8s_2^1 + s_3^1 + 25t_1^1 + 10t_2^1 + t_3^1 - y_1^0 \le 0$$
$$-21s_1^2 - 8s_2^2 + s_3^2 + 25t_1^2 + 10t_2^2 + t_3^2 - y_2^0 \le 0$$
$$-21s_1^3 - 8s_2^3 + s_3^3 + 25t_1^3 + 10t_2^3 + t_3^3 - y_3^0 \le 0$$
$$-21s_1^4 - 8s_2^4 + s_3^4 + 25t_1^4 + 10t_2^4 + t_3^4 - y_4^0 \le 0$$
$$s_1^1 - t_1^1 - y_1^1 = 0, s_2^1 - t_2^1 - y_1^2 = 0, s_3^1 - t_3^1 - y_1^3 = 0$$
$$s_1^2 - t_1^2 - y_2^1 = 0, s_2^2 - t_2^2 - y_2^2 = 0, s_3^2 - t_3^2 - y_2^3 = 0$$
$$s_1^3 - t_1^3 - y_3^1 = 0, s_2^3 - t_2^3 - y_3^2 = 0, s_3^3 - t_3^3 - y_3^3 = 0$$
$$s_1^4 - t_1^4 - y_4^1 = 0, s_2^4 - t_2^4 - y_4^2 = 0, s_3^4 - t_3^4 - y_4^3 = 0$$
$$x \ge 0, s_k, t_k \ge 0, \;\; k = 1, ..., 4.$$

The numerical results are as follows: $x = 30,500$, minimal cost = -929.89. The algorithm converges after 15 iterations. CPU time is 7.74 sec.

3.2 Value of Moment Information

One of the advantages of the SIP formulation is the easiness of incorporating higher-order moment information. This gives us an edge to explore the "value of information" by looking at marginal value added from additional information, especially, higher-order moments. The production planning example above is used for constructing comparisons by assuming various levels of knowledge we have about moments. Different steel purchasing levels are calculated respectively and the comparison is shown in the table below.

Orders of moments known	Steel purchased
=1	30500
≤2	30500
≤3	30500
≤4	30500
≤5	27861
≤6	17876
≤7	17876

From the results above, it is interestingly noted although the general trend is knowing more information enabling a more aggressive result (buying less steel), adding one extra moment information may have no value in the sense of making distributionally robust decision (e.g., the case when we increase our knowledge from 1 moment to 4 moments). When the cost of evaluating uncertainty moments is high or the confidence level of obtaining correct information is low, dealing with low level of knowledge can be sufficient for distributionally robust decision making, at least from our example.

4 Concluding Remarks

We reformulated the distributionally robust two-stage stochastic linear program with separable moment constraints of arbitrary order as a semi-infinite optimization problem under certain conditions such as the linear decision rule. A numerical example in production planning is tested. The computational results appear to show that the improvement margin tends to fade as more and higher order moment information is provided. Possible future directions of research may include nonlinear cases of the problem and nonseparable moment constraints (see e.g., Ling et. al. (2014) and Mehrotra and Zhang (2013)).

Acknowledgment. The authors are in debt to five anonymous referees for their comments that help greatly in improving an earlier version of the manuscript. The research is partially supported by the Provost Chair Fund at National University of Singapore and by the Australian Research Council Grant DP160102819.

References

1. Ang, J., F. Meng, J. Sun. 2014. Two-stage Stochastic Linear Programs with Incomplete Information on Uncertainty. *European Journal of Operational Research* 233, 16-22.
2. Ben-Tal, A., A. Goryashko, E. Guslitzer, and A. Nemirovski. 2004. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming* **99**(2), 351-376.
3. Ben-Tal A., A. Nemirovski. 2001. Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications. *MOS-SIAM Series on Optimization* , Philadelphia.
4. Bertsimas, D., X. Duan, K. Natarajan, C.-P. Teo. 2010. Model for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research* **35**(3) 580-602.
5. Bertsimas, D., R. Freund. 2000. *Data, Models, and Decisions: The Fundamentals of Management Science*. South-Western College Publishing, Cincinnati.
6. Bertsimas, D., M. Sim, M. Zhang. A practicable framework for distributionally robust linear optimization problems. Appeared at $http : //www.optimization-online.org/DB_FILE/$2013/07/3954.*pdf*
7. Chen, W., M. Sim, J. Sun, C.-P. Teo. 2010. From CVaR to uncertainty set: implications in joint chance constrained optimization. *Operations Research* **58**(2) 470-485.
8. Chen, X., M. Sim, P. Sun, J. Zhang. 2008. A linear-decision based approximation approach to stochastic programming. *Operations Research* **56** 344-357.
9. Delage, E., Y. Ye. 2010. Distributionally robust optimization under moment uncertainty with application to data-driven problems. Oper. Res. **58**(3) 596-612.

10. Dupacova, J. 1987. The minimax approach to stochastic programming and an illustrative application. *Stochastics* **20**(1) 73-87.
11. Kall, P. and S.W. Wallace. 1994. *Stochastic Programming.* John Wiley and Sons.
12. Landau, H. J. 1987. *Moments in Mathematics: Lecture Notes Prepared for the AMS Short Course*. American Mathematical Society, San Antonio, Texas, USA.
13. Ling, A., J. Sun, X. Yang, 2014. "Robust Tracking Error Portfolio Selection with Worst-case Downside Risk Measures", **Journal of Economic Dynamics and Controll 39** 178-207.
14. Mehrotra, S. and H. Zhang 2013. Models and Algorithms for Distributionally Robust Least Squares Problems, *Math. Program.*, DOI 10.1007/s10107-013-0681-9.
15. Rockafellar R. T. 1974. *Conjugate Duality and Optimization.* AMS-SIAM Publication.
16. Scarf H. 1958. A min-max solution of an inventory problem. K.J. Arrow, S. Karlin, H. E. Scarf, eds. *Studies in The Mathematical Theory of Inventory and Production.* Stanford University Press, 201-209.
17. Sion, M. 1958. On general minimax theorems. *Pacific Journal of Mathematics* **8** 171-176.
18. Vandenberghe, L., Boyd, S., Comanor, K. 2007. Generalized Chebyshev bounds via semidefinite programming.*SIAM review*, **49**(1) 52-64.
19. Wiesemann, W., D. Kuhn, and M. Sim. 2014. Distributionally Robust Convex Optimization. *Operations Research.* **62**(6), 1358-1376.
20. Wu, S.Y, D. H. Li, L. Qi, and G. Zhou. 2005. An Iterative Method for Solving KKT System of the Semi-Infinite Programming, *Optimization Methods and Software* **20**, 629-643.