Noname manuscript No.
(will be inserted by the editor)

Quadratic Two-Stage Stochastic Optimization with Coherent Measures of Risk

Jie Sun · Li-Zhi Liao · Brian Rodrigues

Received: date / Accepted: date

Abstract A new scheme to cope with two-stage stochastic optimization problems uses a risk measure as the objective function of the recourse action, where the risk measure is defined as the worst-case expected values over a set of constrained distributions. This paper develops an approach to deal with the case where both the first and second stage objective functions are convex linear-quadratic. It is shown that under a standard set of regularity assumptions, this two-stage quadratic stochastic optimization problem with measures of risk is equivalent to a conic optimization problem that can be solved in polynomial time.

Keywords Conic duality · quadratic programs · risk measures · stochastic optimization

Mathematics Subject Classification (2010) 90C15 · 90C25 · 90C34

1 Introduction

In the two-stage recourse model of stochastic optimization, a vector $x \in \mathbb{R}^n$ must be selected optimally with respect to the first (current) stage costs and

This research was partially supported by Australian Research Council under Grant DP160102819.

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

L.-Z. Liao
Department of Mathematics, Hong Kong Baptist University, Hong Kong E-mail: liliao@hkbu.edu.hk

B. Rodrigues
Lee Kong Chian School of Business, Singapore Management University, Singapore E-mail: brianr@smu.edu.sg

constraints as well as certain expected costs and induced constraints associated with corrective actions available in the second (future) stage. The second stage costs and constraints depend on the choice of x as well as a random vector $\tilde{z} := \tilde{z}(\omega) \in \mathscr{L}_p^M(\Omega, \mathcal{F}, \mathbb{P})$ that is not yet realized at stage one. It is convenient to denote the first and second stage cost functions by $f_1(x)$ and $f_2(x, \tilde{z})$, respectively, and formulate the two-stage stochastic optimization problem as

$$\text{(2SSO)} \qquad \min \; f_1(x) + \mathbb{E}_{\mathbb{P}}[f_2(x, \tilde{z})],$$

where $\mathbb{E}$ stands for expectation, $\mathbb{P}$ is the joint probability distribution of $\tilde{z}$. Implicitly, we assume here that for each feasible solution $x \in X = \operatorname{dom} f_1$, the random variable $f_2(x, \tilde{z})$ is measurable.

There is no need to assume that $\tilde{z}$ is continuously distributed or discretely distributed at this juncture although the mathematical tools of treating these two types of problems might be different. However, in classical numerical stochastic optimization it is always assumed that the distribution of $\mathbb{P}$ is given, either in the form of a distribution function or as a complete scenario tree, for otherwise the value of $\mathbb{E}_{\mathbb{P}}[f_2(x, \tilde{z})]$ is not computable. This requirement is seriously restrictive since usually only partial statistical information, such as certain order of moments and the range of support of $\tilde{z}$, is available in practice.

Yet, another disadvantage of the (2SSO) model is that the expectation $\mathbb{E}_{\mathbb{P}}(\cdot)$ may not be a good measure for the "risk" of the second stage recourse action. In many applications, a "coherent" risk measure is much preferred. Here by "risk measure" we mean a functional that maps a random variable to a real number that satisfies certain "coherency" requirements. Let $\mathscr{L}_p$ be the space of a certain random variable and $\mathcal{R}$ be the risk measure. The specific choice of $\mathscr{L}_p$ depends on the applications, it could be $\mathscr{L}_\infty$ [23] or $\mathscr{L}_2$ [3] for examples. Then we have $\mathcal{R} : \mathscr{L}_p \to (-\infty, +\infty]$. For detailed discussion about "convex" or "coherent" risk measures and their impact on optimization, see [17,19,25]. We will provide more details of $\mathcal{R}$ in Section 2.1. Nevertheless, a more reasonable model than (2SSO) is

$$\text{(RM-2SSO)} \qquad \min \; f_1(x) + \mathcal{R}(f_2(x, \tilde{z})),$$

where $\mathcal{R}(\cdot)$ is a coherent risk measure.

Much of the recent work, for instance, [2,9,11,12,15,16], on computational methods for solving (RM-2SSO) focus on the linear case although the ultimate importance of quadratic stochastic programming has been clear in the literature [21,22,27]. In this work, we aim to develop a new solution scheme for the case where both f_1 and f_2 are convex and quadratic (or where one of them is linear as a special case). We assume that $\tilde{z}$ is a continuously distributed random vector without a known distribution, except that certain information on its expectation and support is given. We will make these assumptions clear in Section 2.

The basic problem we would like to address is

$$\text{(P)} \qquad \min \; \frac{1}{2} x^T C x + c^T x + \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}[\psi(x, \tilde{z})], \; \text{ over all } x \in X \subset \mathbb{R}^n,$$

where $C \in \mathbb{S}^n_+, c \in \mathbb{R}^n, X$ is a convex polyhedron, and $\psi(x, \tilde{z})$ is the cost of the second stage recourse problem that depends on $(x, \tilde{z})$. Here as usual, T stands for the transpose, $\mathbb{S}^n$ stands for the space of all symmetric $n \times n$ matrices and $\mathbb{S}^n_+$ is the cone of positive semidefinite symmetric matrices. Moreover, we suppose a representation

$$\psi(x, z) = \sup_{w \in \mathcal{W}(z)} \left\{ w^T [h(z) - T(z)x] - \frac{1}{2} w^T H(z) w \right\}, \tag{1}$$

where $w, h(z) \in \mathbb{R}^W, T(z) \in \mathbb{R}^W \times \mathbb{R}^n, H(z) \in \mathbb{S}^W_+$, and $\mathcal{W}(z)$ is a convex polyhedra. This "quadratic conjugate" format of ψ is well known to be able to cover a wide class of constrained recourse problems including the case where the second stage is a convex quadratic programming problem [21,22]. As explained in detail in [22], $\psi(x, \tilde{z})$ could also be thought of as a penalty for the violation of the constraints $h(\tilde{z}) - T(\tilde{z})x = 0$, while $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}[\cdot]$ is a coherent risk measure for such a penalty.

A basic condition is imposed on the given data. We assume $\mathcal{W}(z), h(z), T(z)$ and $H(z)$ are such that for every $x \in X$ the set

$$\operatorname*{argmax}_{w \in \mathcal{W}(z)} \left\{ w^T [h(z) - T(z)x] - \frac{1}{2} w^T H(z) w \right\} \tag{2}$$

is nonempty and bounded. This assumption will make the optimal recourse action to exist in the second stage in response to any feasible first stage decision, which can be made true by a certain "pre-processing" procedure as described in [20].

The major result in this paper shows that, under a standard set of assumptions on the sets $\mathcal{P}$ and $\mathcal{W}(z)$ and on the functions $h(z), T(z)$ and $H(z)$, the problem (P) is equivalent to a conic optimization problem that can be solved in polynomial time. Indeed, this is quite surprising given that the function $\psi(x, z)$ is in general piecewise quadratic in x (and therefore not generally differentiable). It also completely eliminates the "curse of dimensionality" that arises in the traditional algorithms used in stochastic programming. On the other hand, since the format of the set $\mathcal{P}$ that we choose is highly expressive as demonstrated in [28], the theoretical result is widely applicable. In particular, a spectrum of statistics could be utilized in "designing" the set $\mathcal{P}$ 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.

2 Structural assumptions on set $\mathcal{P}$ and function $\psi(x, z)$

2.1 Notations and coherent risk measures

We denote a random quantity, 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. Given a regular (i.e. pointed, proper, and with nonempty interior) cone $\mathcal{K}$, 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

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

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. If $\mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^M \times \mathbb{R}^T)$ is a joint probability distribution of two random vectors $\tilde{z} \in \mathbb{R}^M$ and $\tilde{u} \in \mathbb{R}^T$, then $\prod_z \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^M)$ denotes the marginal distribution of $\tilde{z}$ under $\mathbb{Q}$. We extend this definition to any sets $\mathcal{Q} \subseteq \mathcal{P}_0(\mathbb{R}^M \times \mathbb{R}^T)$ by setting $\prod_z \mathcal{Q} = \bigcup_{\mathbb{Q}\in\mathcal{Q}}\{\prod_z \mathbb{Q}\}$. Note that there is no assumption on the dependence among $\tilde{z}_i$s and $\tilde{u}_j$s – they could be dependent if they are so in practice.

A risk measure $\mathcal{R} : \mathscr{L}_p \to (-\infty, +\infty]$ is coherent if it satisfies the following axioms.

(A1) $\mathcal{R}(C) = C$ for all constant C,
(A2) $\mathcal{R}((1-\lambda)X + \lambda X') \leq (1-\lambda)\mathcal{R}(X) + \lambda\mathcal{R}(X')$ for $\lambda \in [0, 1]$ ("convexity"),
(A3) $\mathcal{R}(X) \leq \mathcal{R}(X')$ if $X \leq X'$ almost surely ("monotonicity"),
(A4) $\mathcal{R}(\lambda X) = \lambda\mathcal{R}(X)$ for $\lambda > 0$ ("positive homogeneity").

In early literature on coherency [1] it was required to have $\mathcal{R}(X+C) = \mathcal{R}(X) + C$. It can be shown that this follows automatically by (**A1**) and (**A2**) [25]. A coherent risk measure is necessarily representable by a support function, as shown in Proposition 2.91 of Föllmer and Schied [13] and Rockafellar, Uryasev and Zabarankin [24]. Therefore, the term $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\psi(x, \tilde{z})$ that appears in problem (P) is a coherent risk measure of $\psi(x, \tilde{z})$.

2.2 Assumptions on the set $\mathcal{P}$

Here we adopt the "distributionally robust" approach of Wiesemann, Kuhn and Sim [28] (WKS format for short) to define the coherent risk measure $\mathcal{R}$. That is,

$$\mathcal{R}(\psi(x, \tilde{z})) := \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}(\psi(x, \tilde{z})).$$

It is always convenient from the application point of view that we introduce an auxiliary random vector $\tilde{u}$ and think of the set $\mathcal{P}$ used above is the projection of a set $\mathcal{Q}$ in $\mathcal{P}_0(\mathbb{R}^M \times \mathbb{R}^T)$ onto $\mathcal{P}_0(\mathbb{R}^M)$. 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, see [28] for details.

The key point of the WKS format is the description of $\mathcal{P}$, which is

$$\mathcal{P} = \prod_z \mathcal{Q}, \text{ and}$$
$$\mathcal{Q} = \left\{ \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^M \times \mathbb{R}^T) \;:\; \begin{array}{l} \mathbb{E}_{\mathbb{Q}}\left[A\tilde{z} + B\tilde{u}\right] = b, \\ \mathbb{Q}\left[(\tilde{z}, \tilde{u}) \in \Omega\right] = 1 \end{array} \right\},$$

where $\mathbb{Q}$ represents a joint probability distribution of the random vector $(\tilde{z}, \tilde{u})$. We assume that $A \in \mathbb{R}^{K\times N}$, $B \in \mathbb{R}^{K\times T}$, $b \in \mathbb{R}^K$. Ω is the support set of $\mathbb{Q}$ and is defined as

$$\Omega = \left\{(z, u) \in \mathbb{R}^N \times \mathbb{R}^T \;:\; Ez + Fu \preceq_{\mathcal{K}} d\right\}, \tag{3}$$

where $E \in \mathbb{R}^{L\times N}$, $F \in \mathbb{R}^{L\times T}$, and $\mathcal{K}$ is a proper cone. We moreover assume that the set Ω has a non-empty interior and is bounded. A more general definition of $\mathcal{P}$ was first appeared in [28] and therefore we call the above set $\mathcal{Q}$ the *ambiguity set,* since this phrase was used in [28]. This set is closely connected with the notion of "risk envelope" in the theory of risk measure [3, 13, 19].

2.3 Assumptions on $h(\tilde{z}), T(\tilde{z})$, $H(\tilde{z})$ and $\mathcal{W}(\tilde{z})$

We specify $\mathcal{W}(\tilde{z}) := \{w : D(\tilde{z})w \le p(\tilde{z})\}$ with $D(\tilde{z}) \in \mathbb{R}^M \times \mathbb{R}^W,\ p(\tilde{z}) \in \mathbb{R}^M$. Let us consider the functions $h(\tilde{z}), T(\tilde{z})$, $H(\tilde{z})$, $D(\tilde{z})$ and $p(\tilde{z})$ appearing in the definition of ψ. We assume that their dependence on $\tilde{z}$ is affine. In other words, we suppose that there exist $H_m \in \mathbb{S}^W, T_m \in \mathbb{R}^W \times \mathbb{R}^n$, $h_m \in \mathbb{R}^W$, $D_m \in \mathbb{R}^M \times \mathbb{R}^W$ and $p_m \in \mathbb{R}^M$ for $m = 0, 1, ..., N$ such that

$$\begin{cases} H(\tilde{z}) = \sum_{m=1}^N H_m\tilde{z}_m + H_0, \\ T(\tilde{z}) = \sum_{m=1}^N T_m\tilde{z}_m + T_0, \\ h(\tilde{z}) = \sum_{m=1}^N h_m\tilde{z}_m + h_0; \end{cases} \text{ and } \begin{cases} D(\tilde{z}) = \sum_{m=1}^N D_m\tilde{z}_m + D_0, \\ \\ p(\tilde{z}) = \sum_{m=1}^N p_m\tilde{z}_m + p_0. \end{cases}$$

In addition we need assume $H(\tilde{z}) \succeq 0$ a.e. although H_m may not be positive semidefinite for some of the indices m. Overall, these assumptions are called the *affine decision rule,* which has been used first by Ben-Tal and Nemirovski [4] and subsequently used in many literatures, e.g., [2, 6–8, 28] as a standard assumption. It could be thought of as a first order approximation of other (nonlinear) relationships among $\tilde{z}$, $h(\tilde{z}), T(\tilde{z})$, $H(\tilde{z}), D(\tilde{z})$ and $p(\tilde{z})$.

2.4 Duality of quadratic programming

Consider the following convex quadratic conic programming

$$\begin{array}{l} \min\ \frac{1}{2}\langle x, Qx\rangle + \langle q, x\rangle \\ \text{s.t. } \mathcal{A}x - b \in \mathcal{C}_1, x \in \mathcal{C}_2, \end{array} \tag{4}$$

where Q is a self-adjoint positive semidefinite linear operator from $\mathcal{X}$ to $\mathcal{X}$, $\mathcal{A} : \mathcal{X} \to \mathcal{Y}$ is a linear map, $c \in \mathcal{X}$ and $\in \mathcal{Y}$ are given data, $\mathcal{C}_1 \subset \mathcal{Y}$ and $\mathcal{C}_2 \subset \mathcal{X}$ are two closed convex cones, $\mathcal{X}$ and $\mathcal{Y}$ are two real finite dimensional Hilbert spaces. The Lagrangian function associated with problem (4) is given by

$$L(x; y, t) = \frac{1}{2} \langle x, Qx \rangle + \langle q, x \rangle + \langle y, b - \mathcal{A}x \rangle - \langle t, x \rangle .$$

Then, the dual of problem (4) is given by

$$\begin{array}{ll} \max & -\frac{1}{2} \langle v, Qv \rangle + \langle b, y \rangle \\ \text{s.t.} & t - Qv + \mathcal{A}^* y = q, \\ & t \in \mathcal{C}_2^*, y \in \mathcal{C}_1^*, v \in \mathcal{V}, \end{array} \tag{5}$$

where $\mathcal{V} \subset \mathcal{X}$ is any subspace such that $\mathrm{Range}\,(Q) \subset \mathcal{V}$, $\mathcal{C}_1^*$ and $\mathcal{C}_2^*$ are the dual cones of $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively.

Lemma 1 (Strong duality of quadratic conic optimization [5, Theorem 1.7.1]) *If the primal problem (4) is bounded below and strictly feasible (i.e. $Ax \succeq_{C_1} b, x \in \mathrm{int}\; \mathcal{C}_2$ for some x), then the dual problem (5) is solvable and the optimal values in the problems are equal to each other.*

If the dual (5) is bounded above and strictly feasible (i.e., exists $t \in \mathrm{int}\; \mathcal{C}_1^*$ and $y \in \mathrm{int}\; \mathcal{C}_2^*$ such that $t - Qv + \mathcal{A}^* x = c$), then the primal (4) is solvable and the optimal values in the problems are equal to each other.

The conditions to guarantee strong duality can be relaxed, as described in the following lemma, if the conic programming (4) is in fact a classical quadratic programming in the sense that $Q \in \mathbb{S}^n_+, \mathcal{C}_1 = \{0\}$ and $\mathcal{C}_2 = \mathbb{R}^n_+$, whose proof can be found in Dorn [10].

Lemma 2 *If (4) and (5) are in fact a pair of classical quadratic programming problems and (4) and (5) both have feasible solutions, or either (4) or (5) has finite optimal value, then both have optimal solutions and*

$$\min(4) = \max(5).$$

This occurs if and only if the Lagrangian has a saddle point $(\bar{x}; \bar{y}, \bar{t})$, in which case the saddle value $L(\bar{x}; \bar{y}, \bar{t})$ coincides with the common optimal value in (4) and (5), and the saddle points are the pairs such that $\bar{x}$ is an optimal solution to (4) and there exists $\bar{v}$ such that $(\bar{y}, \bar{v}, \bar{t})$ is an optimal solution to (5).

3 Conversion of problem (P) to a conic optimization problem

The last term in the objective function of problem (P) is indeed the optimal value of the following optimization problem

$$\begin{array}{ll} \max\limits_{\mathbb{Q} \in \mathcal{Q}} & \mathbb{E}_{\mathbb{P}} \psi(x, \tilde{z}) \Big(= \mathbb{E}_{\mathbb{Q}} \psi(x, \tilde{z}) \Big) \\ \text{s.t.} & \mathbb{E}_{\mathbb{Q}}(A\tilde{z} + B\tilde{u}) = b \\ & \mathbb{Q}(E\tilde{z} + F\tilde{u} \preceq_{\mathcal{K}} d) = 1, \end{array}$$

where the last constraint means that the probability of $(\tilde{z}, \tilde{u}) \in \Omega$ is one. We may write the problem explicitly as

$$\begin{aligned} \max_{\mathbb{Q}} \quad & \int_{\Omega} \psi(x,z) d\mathbb{Q} \\ \text{s. t.} \quad & \int_{\Omega} (Az + Bu)\, \mathrm{d}\mathbb{Q} = b, \\ & \int_{\Omega} \mathbf{1}_{[(\Omega]} \mathrm{d}\mathbb{Q} = 1, \end{aligned} \tag{6}$$

where $1_{[\Omega]}$ is the characteristic function of set Ω. According to the theory of semi-infinite programming [14], the dual of (6) is a semi-infinite program as follows

$$\begin{array}{ll} \min_{\beta,\eta} & b^T\beta + \eta \\ \text{s.t.} & (Az+Bu)^T\beta + \eta \geq \psi(x,z) \quad \forall (z,u) \in \Omega, \end{array} \tag{7}$$

where $(\beta, \eta) \in \mathbb{R}^M \times \mathbb{R}$ are the dual variables.

Lemma 3 *Strong duality holds between (6) and (7) in the sense that (6) is solvable and* $\max(6) = \min(7)$.

Proof. Observe that for any fixed x, due to our assumptions on $\psi(x,z)$ and the compactness of Ω, $\psi(x,z)$ is a bounded quantity over $(z,u) \in \Omega$, say $|\psi(x,z)| \leq \ell$, where ℓ may depend on x but not on z. Thus, the point $\beta = 0$ and $\eta = \ell + 1$ is a generalized Slater's point for the dual problem. Applying Theorem 18 of Rockafellar [18], strong duality holds in the specified sense. □

Lemma 3 leads to the following result.

Theorem 1 *Under the affine decision rule, problem (P) is equivalent to the following semi-infinite program.*

$$\begin{aligned} \min_{x,\beta,\eta,y,v,t} \quad & \frac{1}{2}x^TCx + c^Tx + b^T\beta + \eta \\ \text{s.t.} \quad & (Az+Bu)^T\beta + \eta \geq \min_{y,v,t} \left\{ \frac{1}{2} v^T H(z) v - p(z)^T y \right\} \ \forall (z,u) \in \Omega \\ & -H_i v + D_i^T y = T_i x - h_i \ \ i = 0,1,...,N \\ & x \in X, v \in \mathcal{V}, y \geq 0 \end{aligned} \tag{8}$$

where $\mathcal{V}$ is any subspace such that $\mathrm{Range}\,(Q) \subset \mathcal{V}$.

Proof. In view of Lemma 3, problem (P) can be written as

$$\begin{aligned} \min_{x\in X,\beta,\eta} \quad & \frac{1}{2}x^TCx + c^Tx + b^T\beta + \eta \\ \text{s.t.} \quad & (Az+Bu)^T + \eta \geq \psi(x,z) \quad \forall (z,u) \in \Omega. \end{aligned}$$

By the assumption on (2), $\psi(x,z)$ is finite for every $x \in X$. Then by Lemma 2, one has

$$\begin{aligned}\psi(x,z) = & \sup_{w\in\mathcal{W}(z)} \{w^T[h(z) - T(z)x] - \frac{1}{2}w^T H(z)w\} \\ = & -\min_{w\in\mathcal{W}(z)} \{w^T[T(z)x - h(z)] + \frac{1}{2}w^T H(z)w\} \\ = & -\max_{y,v,t} \left\{ p(z)^T y - \frac{1}{2}v^T H(z)v : \begin{array}{l} -H(z)v + D(z)^T y = T(z)x - h(z) \\ v \in \mathcal{V}, y \geq 0 \quad \forall (z,u) \in \Omega \end{array} \right\}.\end{aligned}$$

Since int $\Omega \neq \emptyset$, it follows from the affine decision rule that the relation

$$-H(z)v + D(z)^T y = T(z)x - h(z) \quad \forall (z,u) \in \Omega$$

is valid iff

$$-H_i v + D_i^T y = T_i x - h_i \quad \forall i = 0,1,...,N,$$

which completes the proof. □

Corollary 1 *Problem (8) is equivalent to*

$$\begin{aligned} \min_{x,y,v,t,\beta,\eta} \ & \frac{1}{2}x^T Cx + c^T x + b^T \beta + \eta \\ s.t. \ & (Az + Bu)^T \beta + \eta \geq \frac{1}{2}v^T H(z)v - p(z)^T y \quad \forall (z,u) \in \Omega \\ & -H_i v + D_i^T y = T_i x - h_i, \quad i = 0,1,...,N \\ & x \in X, v \in \mathcal{V}, y \geq 0. \end{aligned} \tag{9}$$

Proof. Let

$$F := \left\{ (x,y,v,t) : x \in X, v \in \mathcal{V}, y \geq 0, t - H_i v + D_i^T y = T_i x - h_i, \ i = 0,1,...,N \right\}.$$

Clearly, F is a closed convex set. Its projection onto the (y,v,t)-space is defined as

$$\prod_{yvt} F := \{(y,v,t) : \exists x \text{ such that } (x,y,v,t) \in F\}.$$

The first constraint in (8) can be written as

$$\forall (z,u) \in \Omega, \quad \exists (y,v,t) \in \prod_{yvt} F : (Az+Bu)^T \beta + \eta - \left[\frac{1}{2}v^T H(z)v - p(z)^T y\right] \geq 0,$$

or equivalently

$$\min_{(z,u)\in\Omega} \max_{(y,v,t)\in\prod_{yvt} F} \left\{ (Az + Bu)^T \beta + \eta - \left[\frac{1}{2}v^T H(z)v - p(z)^T y\right] \right\} \geq 0.$$

The function is convex in (z,u) and concave in (y,v,t) and both sets, Ω and $\prod_{yvt} F$, are closed and convex. In addition, Ω is bounded. By Sion's minimax theorem [26], we have

$$\min_{(z,u)\in\Omega} \max_{(y,v,t)\in\prod_{yvt} F} \left\{ (Az+Bu)^T\beta + \eta - \left[\frac{1}{2}v^T H(z)v - p(z)^T y\right]\right\}$$

$$= \max_{(y,v,t)\in\prod_{yvt} F} \min_{(z,u)\in\Omega} \left\{ (Az+Bu)^T\beta + \eta - \left[\frac{1}{2}v^T H(z)v - p(z)^T y\right]\right\}.$$

The first constraint in (8) is therefore equivalent to

$$\exists (y,v,t) \in \prod_{yvt} F, \quad \forall (z,u) \in \Omega : (Az+Bu)^\top \beta + \eta - \left[\frac{1}{2}v^T H(z)v - p(z)^T y\right] \geq 0,$$

which proves the corollary. □

For simplicity of notation, let $P := [p_1, ..., p_N]$. Then $p(z)^T y = y^T P z + p_0^T y$. Note that here P is a matrix and each $p_i, i = 1, ..., N$ is a vector.

Lemma 4 *Under strong duality, or under the condition that one of $H_1, ..., H_N$ is positive definite, the semi-infinite constraint in (9)*

$$(Az+Bu)^\top \beta + \eta - \left[\frac{1}{2}v^T H(z)v - p(z)^T y\right] \geq 0 \quad \forall (z,u) \in \Omega \tag{10}$$

is equivalent to the following set of constraints: $\exists \lambda \in \mathcal{K}^*, \Lambda \in \mathbb{S}^W$ such that

$$\begin{aligned}
& d^T\lambda + \langle H+0, \Lambda\rangle \geq \eta + p_0^T y \\
& E^T\lambda + \begin{bmatrix} \langle H_1, \Lambda\rangle \\ \vdots \\ \langle H_N, \Lambda\rangle \end{bmatrix} = A^T\beta + P^T y \\
& F^T\lambda = B^T\beta \\
& \begin{pmatrix} 2 & v^T \\ v & \Lambda \end{pmatrix} \succeq 0.
\end{aligned} \tag{11}$$

Proof. Since $H(z) \succeq 0$ over Ω, the constraint (10) means that the optimal value of the semidefinite program

$$\begin{aligned}
\max\nolimits_{H,z,u} \quad & \frac{1}{2}v^T H v - (Az+Bu)^T\beta - y^T P z \\
\text{s.t.} \quad & H - \sum_{m=1}^{N} H_m z_m = H_0 \\
& Ez + Fu \preceq_{\mathcal{K}} d \\
& H \succeq 0
\end{aligned} \tag{12}$$

is less than or equal to $\eta + p_0^T y$.

If strong duality holds, then the optimal value of (12) is less than or equal to $\eta + p^T y$ if and only if its dual optimal value is so, namely,

$$\begin{aligned} \min\nolimits_{\lambda,\Lambda} \; & \left(d^T\lambda + \langle H_0, \Lambda\rangle\right) \le \eta + p_0^T y \\ \text{s.t.} \quad & E^T\lambda + \begin{bmatrix} \langle H_1, \Lambda\rangle \\ \vdots \\ \langle H_N, \Lambda\rangle \end{bmatrix} = A^T\beta + P^T y \\ & F^T\lambda = B^T\beta, \lambda \in \mathcal{K}^* \\ & \Lambda \succeq \frac{1}{2} v v^T. \end{aligned} \tag{13}$$

Observe that

$$\min_{\lambda,\Lambda} \left(d^T\lambda + \langle H_0, \Lambda\rangle\right) \le \eta + p_0^T y \iff \exists(\lambda, \Lambda) : d^T\lambda + \langle H_0, \Lambda\rangle \le \eta + p_0^T y$$

and

$$\Lambda \succeq \frac{1}{2} v v^T \iff \begin{pmatrix} 2 & v^T \\ v & \Lambda \end{pmatrix} \succeq 0.$$

Hence, if strong duality holds between (12) and (13), then the conclusion of this lemma is true.

It remains to show that if one of $H_1, ..., H_N$ is positive definite, then strong duality holds for problem (12) and (13). Since $H(z) \succeq 0$ over Ω and $\text{int}\,\Omega \ne \emptyset$, there exists $(z^0, u^0) \in \text{int}\,\Omega$ that satisfy

$$Ez^0 + Fu^0 \succ_{\mathcal{K}} d \text{ and } H = \sum_{m=1}^{N} H_m z_m^0 + H_0 \succeq 0.$$

Suppose without loss of generality that $H_1 \succ 0$. Then for small $\varepsilon > 0$, the point

$$(z^1, u^1) := (z_1^0 + \varepsilon, z_2^0, ..., z_N, u_1^0, ..., u_T^0)^T \in \text{int}\,\Omega$$

and it satisfies

$$Ez^1 + Fu^1 \succ_{\mathcal{K}} d \text{ and } \bar{H} := \sum_{m=1}^{N} H_m z_m^1 + H_0 = H + \varepsilon H_1 \succ 0.$$

Therefore $(z^1, u^1, \bar{H})$ is a strictly feasible point of (12), and (12) is bounded above by $\eta + p_0^T y$. Hence, by Lemma 1, strong duality holds between (12) and (13). The proof is thus completed. □

Since the semi-infinite constraint in (8) can be converted to a set of conic constraints, whose dimension is polynomial in the given data, we come up to the following theorem, whose proof is evident.

Theorem 2 *Under the affine decision rule the quadratic risk measure two-stage stochastic optimization problem (P) with WKS format of ambiguity sets can be solved in polynomial time as a conic optimization problem.*

References

1. P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. Math. Finance **9** (1999) 203-227.
2. J. Ang, F. Meng, and J. Sun. Two-stage stochastic linear programs with incomplete information on uncertainty. European J. Oper. Res. **233** (2014) 16-22.
3. M. Ang, J. Sun, and Q. Yao. On dual depresentation of coherent risk measures. Manuscript, School of Science, Curtin University, Australia (2015) to appear in Ann. Oper. Res.
4. A. Ben-Tal and A. Nemirovski. Robust Convex Optimization. Math. Oper. Res., **23** (1998) 769-805.
5. A. Ben-Tal and A. Nemirovski. Lectures on Modern Convex Optimization: Analysis, Algorithms, Engineering Applications, MPS-SIAM Series on Optimization, SIAM, Philadelphia (2001).
6. D. Bertsimas, X. 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. X. Chen, M. Sim, P. Sun, and J. Zhang. A linear-decision based approximation approach to stochastic programming. *Oper. Res.* **56** (2008) 344-357.
8. 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.
9. E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. Oper. Res. **58** (2010) 595-612.
10. W. S. Dorn. Duality in quadratic programming. Quart. Appl. Math. **18** (1960) 155-162.
11. S. Gao, L. Kong, and J. Sun. Two-stage stochastic linear programs with moment information on uncertainty, Optimization **63** (2014) 829-837.
12. S. Gao, J. Sun, and S.-Y. Wu. Solving robust two-stage stochastic linear programs by a semi-infinite programming approach, Manuscript, School of Science, Curtin University, Australia (2015).
13. H. Föllmer and A. Schied. Stochastic Finance. Walter de Gruyter, Berlin (2002).
14. R. Hettich, K. O. Kortanek, Semi-Infinite programming: theory, methods, and applications, SIAM Rev. **35(3)** (1994) 380-429.
15. B. Li, J. Sun, K.-L. Teo and C. Yu. Distributionally robust two-stage stochastic linear programming with a WKS-type of ambiguity set. Manuscript School of Science, Curtin University, Australia (2015).
16. A. Ling, J. Sun and X. Yang. Robust tracking error portfolio selection with worst-case downside risk measures, J. Econom. Dynam. Control **39** (2014) 178-207.
17. H-J. Lüthi and J. Doege. Convex risk measures for portfolio optimization and concepts of flexibility. Math. Program. **104** (2005) 541-559.
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. INFORMS Tutorials in Operations Research (2007) 38-61.
20. R. T. Rockafellar and R. J-B. Wets. Stochastic convex programming: relatively complete recourse and induced feasibility. SIAM J. Control. Optim. **14** (1976) 574-589.
21. R. T. Rockafellar and R. J-B. Wets. Linear-quadratic programming problems with stochastic penalties: the finite generation algorithm. in: Stochastic Optimization, V.I. Arkin, A. Shiraev, R. J-B. Wets eds. Springer-Verlag Lecture Series in Control and Information Sciences **81** (1986) 545-560.
22. R. T. Rockafellar and R. J-B. Wets. A Lagrangian finite generation technique for solving linear-quadratic problems in stochastic programming. Mathematical Programming Study **28** (1986) 63-93.
23. R. T. Rockafellar and R. J-B. Wets. Stochastic variational inequalities:single-stage to multistage. Manuscript, Department of Mathematics, University of Washington, (2015).
24. R. T. Rockafellar, S. P. Uryasev, and M. Zabarankin. Generalized deviations in risk analysis. Finance Stoch. **10** (2006) 51-74.
25. A. Shapiro, D. Dentcheva, and A. Ruszczyski. Lectures on Stochastic Programming: Modeling and Theory. SIAM, Philadelphia (2009).
26. M. Sion. On general minimax theorems. Pacific J. Math. **8** (1958) 171-176.

27. R. J-B. Wets. Stochastic programming: solution techniques and approximation schemes. in: A. Bachem *et al.*, eds. Mathematical Programming: The State-of-the-Art, Springer, Berlin (1983) 566-603.
28. W. Wiesemann, D. Kuhn and M. Sim. Distributionally robust convex optimization. Oper. Res. **62** (2014) 1358-1376.