**Mathematical Programming, Series B 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 constraints as well as certain expected costs and constraints associated with

* This paper is dedicated to Terry Rockafellar in celebration of his 80th birthday.

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

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 $\tilde{z} := \tilde{z}(\omega) \in \mathscr{L}_p^Z(\Omega, \mathcal{F}, \mathbb{P})$ that is not yet realized at stage one, where $\mathscr{L}_p$ is the $p$-integrable Lebesgue space and $Z$ is the dimension of $\tilde{z}$. The specific choice of $\mathscr{L}_p$ depends on the applications, it could be $\mathscr{L}_\infty$ [26] or $\mathscr{L}_2$ [3] for example. 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 and $\mathbb{P}$ 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 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 suitable measure for the "risk" of the second stage recourse action. In many applications, a more general "coherent" risk measure is much preferred. Here by "risk measure" we mean a functional $\mathcal{R} : \mathscr{L}_p \to (-\infty, +\infty]$ that maps a random variable to a real number or $+\infty$ and satisfies certain "coherency" requirements. For detailed discussion about "convex" or "coherent" risk measures and their impact on optimization, see [18, 21, 27]. We will provide more details of $\mathcal{R}$ in Section 2.1. Nevertheless, a more flexible 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, including the expectation as a special case.

Much of the recent work, for instance, [2, 10, 14, 15, 7, 17], on methods and applications of (RM-2SSO) focus on the linear case although the ultimate importance of quadratic stochastic programming has been clear in the literature [24, 25, 29]. A good new example could be the model of two-stage stochastic games, in which the Nash equilibrium reduces to solving a two-stage stochastic linear complementarity problem that turns out to be equivalent to a quadratic (2SSO), where $f_1$ is convex quadratic and $f_2$ is the optimal value of a convex quadratic program parameterized by $(x, \tilde{z})$. Note that the stochastic two-stage complementarity problem is a special case of the multistage stochastic variational inequalities recently studied by Rockafellar and Wets [26]. Therefore the study on quadratic models could lead us go beyond the area of stochastic optimization to reach the area of stochastic equilibrium, which is not yet explored by the current literature on linear models of (RM-2SSO).

In this paper we aim to develop a new solution scheme for (RM-2SSO). We assume that $\tilde{z}$ is a continuously distributed random vector with an unknown distribution, except that certain information on its expectation and support is given. We will make these assumptions clear in Section 2. The basic model we would like to address is

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

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 the apostrophe $'$ 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

$$\psi(x,z) = \sup_{w\in\mathcal{W}(z)} \left\{ w'[h(z) - T(z)x] - \frac{1}{2}w'H(z)w \right\}, \tag{1}$$

where $w \in \mathbb{R}^W$ is the decision vector of the second stage problem, and $h(z) \in \mathbb{R}^W, T(z) \in \mathbb{R}^W \times \mathbb{R}^n, H(z) \in \mathbb{S}^W_+$, and $\mathcal{W}(z)$ is a convex polyhedron for each realization $z$ of $\tilde{z}$. This "quadratic conjugate" format of $\psi$ 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 [24, 25]. As explained in detail in [25], $\psi(x,\tilde{z})$ 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

$$\operatorname*{argmax}_{w\in\mathcal{W}(z)} \left\{ w'[h(z) - T(z)x] - \frac{1}{2}w'H(z)w \right\} \tag{2}$$

is nonempty. This assumption will ensure the optimal recourse action exists 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 [23]. For ease of reference, this condition is henceforth called the existence of recourse assumption.

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 surprising given that the function $\psi(x,z)$ is in generally nonlinear and nonsmooth in $(x,z)$ and the distribution of $\tilde{z}$ is unknown. It helps to avoid the "curse of dimensionality" that arises in some existing algorithms used in stochastic programming.

## 2 Coherent risk measures and structural assumptions

### 2.1 Notations and notational conventions

In this paper by $\mathbb{R}^k$ we mean a finite $k$-dimensional real Hilbert space equipped with inner product $\langle\cdot,\cdot\rangle$ . In particular, a special case of it is the space $\mathbb{S}^n$ with Frobenius norm. 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. However, when there is no confusion, upper case letters are also used to represent natural numbers. Script letters are used for sets or mappings (operators). As usual, if $\mathcal{B}(u)$ is a linear mapping of $u$, then we simply write it as $\mathcal{B}u$. We denote the adjoint operator of $\mathcal{B}$ by $\mathcal{B}^*$. If $x$ is a vector, we use the notation $x_i$ to denote the $i$th component of the vector. Given a regular (i.e., convex, pointed, closed, and having nonempty interior) cone $\mathcal{K}$ in $\mathbb{R}^k$, such as the second-order cone or the semidefinite cone, for any two vectors $x, y$, the notation $x \preceq_{\mathcal{K}} y$ (respectively, $x \prec_{\mathcal{K}} y$) or $y \succeq_{\mathcal{K}} x$ (respectively, $y \succ_{\mathcal{K}} x$) means $y - x \in \mathcal{K}$ (respectively, $y - x \in \operatorname{int} \mathcal{K}$, where "int" means "the interior of"). Given a closed convex cone $\mathcal{C}$, the dual cone of $\mathcal{C}$ is denoted by

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

Let $Z$ and $U$ be the dimensions of random vectors $\tilde{z}$ and $\tilde{u}$, respectively. The set $\mathcal{P}_0(\mathbb{R}^Z)$ represents the space of probability distributions on $\mathbb{R}^Z$ and $\mathcal{P}_0(\mathbb{R}^M \times \mathbb{R}^U)$ represents the space of probability distributions on $\mathbb{R}^Z \times \mathbb{R}^U$, respectively. If $\mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^Z \times \mathbb{R}^U)$ is a joint probability distribution of two random vectors $\tilde{z} \in \mathbb{R}^Z$ and $\tilde{u} \in \mathbb{R}^U$, then $\prod_z \mathbb{Q} \in \mathcal{P}_0(\mathbb{R}^Z)$ denotes the marginal distribution of $\tilde{z}$ under $\mathbb{Q}$. We extend this definition to any set $\mathcal{Q} \subseteq \mathcal{P}_0(\mathbb{R}^Z \times \mathbb{R}^U)$ 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.

### 2.2 Coherent risk measures

A random variable in our discussion is regarded as an element in the space $\mathscr{L}_p$ $(1 \leq p \leq \infty)$, which in particular means its 0th,..., $p$th order moments are well-defined and finite. A risk measure $\mathcal{R} : \mathscr{L}_p \to (-\infty, +\infty]$ is coherent if it satisfies the following axioms in which $\tilde{X}$ and $\tilde{Y}$ are two arbitrary random variables.

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

In early literature on coherency [1] it was required to have $\mathcal{R}(\tilde{X} + C) = \mathcal{R}(\tilde{X}) + C$. It can be shown that this follows automatically from (**A1**) and (**A2**) [22].

It can be verified that the term $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\psi(x,\tilde{z})$ that appears in problem (P) satisfies **(A1)-(A4)** with respect to $\psi(x,\tilde{z})$, therefore it is a coherent risk measure of $\psi(x,\tilde{z})$.

### 2.3 Assumptions on the set $\mathcal{P}$

It should be noted that a specific risk measure of the form $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}(\cdot)$ depends on the specific form of $\mathcal{P}$. Here we adopt the "distributionally robust" approach of Wiesemann, Kuhn and Sim [30] (WKS format for short) to define the set $\mathcal{P}$.

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 $\mathcal{P}_0(\mathbb{R}^Z \times \mathbb{R}^U)$ onto $\mathcal{P}_0(\mathbb{R}^Z)$. This scheme does not complicate 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 Example 1 below and [30] for details.

The key of the WKS format is the following description of $\mathcal{P}$ and the support set $\Omega$ of $(\tilde{z},\tilde{u})$, in which $\mathcal{A},\mathcal{B},\mathcal{E}$ and $\mathcal{F}$ are linear mappings.

$$\mathcal{P} = \prod_z \mathcal{Q} \text{ and}$$

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

where $\mathcal{A} : \mathbb{R}^Z \to \mathbb{R}^K$, $\mathcal{B} : \mathbb{R}^U \to \mathbb{R}^K$, and $b \in \mathbb{R}^K$. We assume that $\Omega$ is defined as

$$\Omega = \left\{ (z,u) \in \mathbb{R}^Z \times \mathbb{R}^U \ : \ \mathcal{E}z + \mathcal{F}u \succeq_{\mathcal{K}} d \right\}, \tag{3}$$

where $\mathcal{E} : \mathbb{R}^Z \to \mathbb{R}^L$, $\mathcal{F} : \mathbb{R}^U \to \mathbb{R}^L$, $d \in \mathbb{R}^L$ and $\mathcal{K}$ is a regular cone in $\mathbb{R}^L$. We moreover assume that the set $\Omega$ has a non-empty interior and is bounded.

A more general definition of $\mathcal{P}$ first appeared in [30] and therefore we call the above set $\mathcal{Q}$ the " ambiguity set" since this phrase was used in [30]. This set is closely connected with the notion of "risk envelope" in the theory of risk measures [3,12,21].

**Example 1** *From past statistics, we have an empirical bound for the second moment of* $\tilde{z}$ *in the sense of*

$$\mathbb{E}(zz') \preceq_{\mathcal{K}} \bar{\Sigma} \ \textit{with} \ \mathcal{K} = \mathbb{S}^Z_+.$$

*Therefore we want to define the set* $\mathcal{P}$ *as*

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^Z) \ : \ \mathbb{E}_{\mathbb{P}}(zz') \preceq_{\mathcal{K}} \bar{\Sigma} \right\}. \tag{4}$$

*However, (4) is not in the WKS format. We then introduce an auxiliary vector* $\tilde{u} \in \mathbb{S}^Z$ *(so* $\tilde{u}$ *is actually a matrix) and define* $\mathcal{Q}$ *as*

$$\mathcal{Q} := \left\{ \mathbb{Q} \in \mathcal{P}_0\left(\mathbb{R}^Z \times \mathbb{S}^Z\right) : \begin{array}{l} \mathbb{E}_{\mathbb{Q}}\left[\tilde{u}\right] = \bar{\Sigma}, \\ \mathbb{Q}\left( \begin{bmatrix} 1 & \tilde{z}' \\ \tilde{z} & \tilde{u} \end{bmatrix} \succeq_{\mathcal{K}} 0 \right) = 1 \end{array} \right\}. \tag{5}$$

*Then (5) is in the WKS format and it can be readily shown that* $\mathcal{P} = \prod_{\tilde{z}} \mathcal{Q}$.

*If we want to impose another condition, say,* $\mathbb{E}_{\mathbb{P}}(z) = \mu$ *in the definition of* $\mathcal{P}$*, then, in (5), we simply change*

$$\mathbb{E}_{\mathbb{Q}}\left[\tilde{u}\right] = \bar{\Sigma} \quad \text{to} \quad \mathbb{E}_{\mathbb{Q}}\begin{pmatrix} \tilde{z} \\ \tilde{u} \end{pmatrix} = \begin{pmatrix} \mu \\ \bar{\Sigma} \end{pmatrix},$$

*and this new constraint is still in the WKS format. In both cases, the corresponding* $\mathcal{A}, \mathcal{B}, b, \mathcal{E}, \mathcal{F}$*, and* $d$ *can be straightforwardly determined.*

Since the format of the set $\mathcal{Q}$ that we choose is highly expressive as demonstrated in [30], the theoretical result of this paper is expected to be useful in a spectrum of applications. In particular, a number of statistics such as moments of positive rational order, the mean absolute deviation, and the marginal median could be cast into the form $\mathcal{Q}$ above and thus create different risk measures. These characteristics reinforce our confidence in using risk measures in the modeling of stochastic optimization problems.

### 2.4 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 $D(\tilde{z}) \in \mathbb{R}^V \times \mathbb{R}^W,\ p(\tilde{z}) \in \mathbb{R}^V$. Let us consider the functions $h(\tilde{z}), T(\tilde{z})$, $H(\tilde{z})$, $D(\tilde{z})$ and $p(\tilde{z})$ that appeared in the definition of $\psi$. We assume that their dependence on $\tilde{z}$ is affine. In other words, we suppose that there exist $H_m \in \mathbb{S}^W, T_m \in \mathbb{R}^W \times \mathbb{R}^n$, $h_m \in \mathbb{R}^W$, $D_m \in \mathbb{R}^V \times \mathbb{R}^W$ and $p_m \in \mathbb{R}^V$ for $m = 0, 1, ..., Z$ such that

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

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 papers, e.g., [2,6,8,9,30] 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 $p(\tilde{z})$.

## 2.5 Duality of conic quadratic programming

Consider the following convex conic quadratic programming

$$\begin{array}{ll} \min & \frac{1}{2}\langle x, Qx\rangle + \langle q, x\rangle \\ \text{s. t.} & \mathcal{S}x - b \in \mathcal{C}_1, x \in \mathcal{C}_2, \end{array} \tag{6}$$

where $Q$ is a self-adjoint positive semidefinite linear operator from $\mathcal{X}$ to $\mathcal{X}$, $\mathcal{S} : \mathcal{X} \to \mathcal{Y}$ is a linear operator, whose adjoint operator is denoted by $\mathcal{S}^*$, $c \in \mathcal{X}$ and $b \in \mathcal{Y}$ 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 finite dimensional real Hilbert spaces. The Lagrangian function associated with problem (6) is given by

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

Then, the dual of problem (6) is given by

$$\begin{array}{ll} \max & -\frac{1}{2}\langle v, Qv\rangle + \langle b, y\rangle \\ \text{s. t.} & t - Qv + \mathcal{S}^* y = q, \\ & t \in \mathcal{C}_2^*, y \in \mathcal{C}_1^*, v \in \mathcal{V}, \end{array} \tag{7}$$

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

**Lemma 1 (Strong duality of conic quadratic optimization [5, Theorem 1.7.1])** *If the primal problem (6) is bounded below and strictly feasible (i.e. $\mathcal{S}x \succ_{C_1} b, x \in \text{int }\mathcal{C}_2$ for some $x$), then the dual problem (7) is solvable and the optimal values in the problems are equal to each other.*

*If the dual (7) is bounded above and strictly feasible (i.e., $\exists t \in \text{int }\mathcal{C}_1^*$ and $y \in \text{int }\mathcal{C}_2^*$ such that $t - Qv + \mathcal{S}^* x = c$), then the primal (6) 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 (6) is in fact a standard quadratic program in the sense that either $Q \in \mathbb{S}^n_+, \mathcal{C}_1 = \{0\}$ and $\mathcal{C}_2 = \mathbb{R}^n_+$ or $Q \in \mathbb{S}^n_+, \mathcal{C}_1 = \mathbb{R}^m_+$ and $\mathcal{C}_2 = \mathbb{R}^n$, whose proof can be found in Dorn [11].

**Lemma 2** *If (6) and (7) are in fact a pair of standard quadratic programming problems and (6) and (7) both have feasible solutions, or either (6) or (7) has finite optimal value, then both have optimal solutions and*

$$\min\ (6) = \max\ (7).$$

*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 (6) and (7), and the saddle points are the pairs such that $\bar{x}$ is an optimal solution to (6) and there exists $\bar{v}$ such that $(\bar{y}, \bar{v}, \bar{t})$ is an optimal solution to (7).*

Later in this paper, Lemma 1 will be applied in the analysis of Theorem 2. Lemma 2 will be used to prove Lemma 4.

2.6 An assumption on the dual affine space

Applying Lemma 2 to the recourse function in (RM-2SSO), we obtain

$$\begin{aligned}\psi(x,z) &= \sup_{w\in\mathcal{W}(z)}\left\{w'[h(z)-T(z)x]-\frac{1}{2}w'H(z)w\right\}\\ &= -\min_{w\in\mathcal{W}(z)}\left\{w'[T(z)x-h(z)]+\frac{1}{2}w'H(z)w\right\}\\ &= -\max_{y,v}\left\{p(z)'y-\frac{1}{2}v'H(z)v:\begin{array}{l}-H(z)v+D(z)'y=T(z)x-h(z)\\ y\geq 0, v\in\mathcal{V}\end{array}\right\}.\\ &\quad\text{(by Lemma 2 since } \psi(x,z)<\infty)\end{aligned}$$

Our last assumption requires the equivalence between the system

$$-H(z)v+D(z)'y=T(z)x-h(z) \tag{8}$$

and the system

$$-H_iv+D_i'y=T_ix-h_i \quad \forall i=0,1,...,Z. \tag{9}$$

From the affine decision rule, this equivalence is valid if the system (8) has a common solution $(v,y)$ for $Z$ linearly independent realizations of $\tilde{z}$ in $\Omega$.

We will call the equivalence of (8) and (9) the strong dual affine feasibility assumption.

## 3 Reformulation of problem (P) into a conic optimization problem

The last term in the objective function of problem (P), $\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}[\psi(x,\tilde{z})]$, is indeed the optimal value of the following optimization problem due to the definition of $\mathcal{Q}$.

$$\begin{aligned}\max_{\mathbb{Q}}\ & \mathbb{E}_{\mathbb{Q}}\psi(x,\tilde{z}) \quad \left(=\mathbb{E}_{\mathbb{P}}\psi(x,\tilde{z})\right)\\ \text{s. t.}\ & \mathbb{E}_{\mathbb{Q}}(\mathcal{A}\tilde{z}+\mathcal{B}\tilde{u})=b,\\ & \mathbb{Q}(\mathcal{E}\tilde{z}+\mathcal{F}\tilde{u}\succeq_{\mathcal{K}} d)=1,\end{aligned}$$

where the constraints mean that $\mathbb{Q}$ is a probability measure of $(\tilde{z},\tilde{u})$ satisfying $\mathbb{Q}\in\mathcal{Q}$. We may write the problem explicitly as

$$\begin{aligned}\max_{\mathbb{Q}}\ & \int_{\Omega}\psi(x,z)d\mathbb{Q}(z,u)\\ \text{s. t.}\ & \int_{\Omega}(Az+Bu)\,\mathrm{d}\mathbb{Q}(z,u)=b,\\ & \int_{\Omega}\mathbb{I}_{[(\Omega]}\mathrm{d}\mathbb{Q}(z,u)=1,\end{aligned} \tag{10}$$

where $\mathbb{I}_{[\Omega]}$ is the characteristic function of set $\Omega$. According to the theory of semi-infinite programming [16], the dual of (10) is a semi-infinite program as follows

$$\begin{array}{ll} \min_{\beta,\eta} \; \langle b, \beta\rangle + \eta & \\ \text{s. t.} \quad \langle \mathcal{A}z + \mathcal{B}u, \beta\rangle + \eta \geq \psi(x,z) \quad \forall (z,u) \in \Omega, \end{array} \tag{11}$$

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

**Lemma 3** *Strong duality holds between (10) and (11) in the sense that (10) is solvable and* $\max(10) = \min(11)$.

**Proof.** Observe that for any fixed $x$, the function $\psi(x,z)$ is convex in $z$ because it is a pointwise maximum of affine functions of $z$ due to the affine decision rule. Our assumption on the nonemptyness of

$$\operatorname*{argmax}_{w \in \mathcal{W}(z)} \left\{ w'[h(z) - T(z)x] - \frac{1}{2} w' H(z) w \right\}$$

then implies that $\psi(x,\cdot)$ is finite everywhere for fixed $x \in \mathcal{X}$. Thus, the function $\psi(x,\cdot)$ is continuous since a convex function is necessarily continuous in the relative interior of its domain (In this case the domain of $\psi(x,\cdot)$ is $\mathbb{R}^Z$) [19, Theorem 10.1].

The continuity of $\psi(x,\cdot)$ and the compactness of $\Omega$ guarantee that $\psi(x,z)$ is a bounded quantity over $(z,u) \in \Omega$, say $|\psi(x,z)| \leq \ell$, where $x \in \mathcal{X}$ and $\ell$ may depend on $x$ but not on $z$. Thus, the point $(\beta,\eta) = (0, \ell+1)$ is a generalized Slater's point for the dual problem (11). Applying Theorem 18 of Rockafellar [20], strong duality holds in the specified sense. □

Lemma 3 leads to the following result.

**Lemma 4** *Under the existence of recourse assumption, the affine decision rule, and the strong dual affine feasibility, problem (P) is equivalent to the following semi-infinite program.*

$$\begin{array}{rl} \displaystyle\min_{x,\beta,\eta} & \frac{1}{2} x' C x + c' x + \langle b, \beta \rangle + \eta \\ \text{s.t.} & \langle \mathcal{A}z + \mathcal{B}u, \beta \rangle + \eta \geq \phi(x,z) \quad \forall (z,u) \in \Omega, \\ & x \in X, \end{array} \tag{12}$$

where

$$\phi(x,z) = \min_{y,v} \left\{ \frac{1}{2} v' H(z) v - p(z)' y : \begin{array}{l} -H_i v + D_i' y = T_i x - h_i \;\; i = 0,1,...,Z \\ y \geq 0, v \in \mathcal{V}, \\ \mathcal{V} \textit{ is any subspace such that Range } (Q) \subset \mathcal{V} \end{array} \right\}.$$

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

$$\begin{array}{rl} \displaystyle\min_{x \in X, \beta, \eta} & \frac{1}{2} x' C x + c' x + \langle b, \beta \rangle + \eta \\ \text{s. t.} & \langle \mathcal{A}z + \mathcal{B}u, \beta \rangle + \eta \geq \psi(x,z) \quad \forall (z,u) \in \Omega. \end{array}$$

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

$$\begin{aligned}\psi(x,z) &= \sup_{w\in\mathcal{W}(z)} \{w'[h(z)-T(z)x] - \frac{1}{2}w'H(z)w\} \\ &= -\min_{w\in\mathcal{W}(z)} \{w'[T(z)x-h(z)] + \frac{1}{2}w'H(z)w\} \\ &= -\max_{y,v} \left\{ p(z)'y - \frac{1}{2}v'H(z)v : \begin{array}{l} -H(z)v + D(z)'y = T(z)x - h(z) \\ y \geq 0, v \in \mathcal{V} \end{array} \right\}.\end{aligned}$$

It follows from the strong dual affine feasibility that the relation

$$-H(z)v + D(z)'y = T(z)x - h(z) \quad \forall (z,u) \in \Omega$$

can be substituted by

$$-H_i v + D_i' y = T_i x - h_i \quad \forall i = 0,1,...,Z,$$

which completes the proof. □

**Theorem 1** *Under the assumptions of Lemma 3, problem (P) is equivalent to*

$$\begin{aligned}\min_{x,y,v,\beta,\eta} \quad & \frac{1}{2}x'Cx + c'x + \langle b, \beta\rangle + \eta \\ \text{s. t.} \quad & \langle \mathcal{A}z + \mathcal{B}u, \beta\rangle + \eta \geq \frac{1}{2}v'H(z)v - p(z)'y \quad \forall (z,u) \in \Omega, \\ & -H_i v + D_i' y = T_i x - h_i, \quad i = 0,1,...,Z, \\ & x \in X, v \in \mathcal{V}, y \geq 0.\end{aligned} \tag{13}$$

**Proof.** By Lemma 4 we only need to prove that problem (12) is equivalent to problem (13). Let

$$F := \{(x,y,v) : x \in X, v \in \mathcal{V}, y \geq 0, -H_i v + D_i' y = T_i x - h_i,\ i = 0,1,...,Z\}.$$

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

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

The first constraint in (12) can be written as

$$\forall (z,u) \in \Omega, \quad \exists (y,v) \in \prod_{yv} F : \langle \mathcal{A}z + \mathcal{B}u, \beta\rangle + \eta - \left[\frac{1}{2}v'H(z)v - p(z)'y\right] \geq 0,$$

or equivalently

$$\min_{(z,u)\in\Omega} \max_{(y,v)\in\prod_{yv} F} \left\{ \langle \mathcal{A}z + \mathcal{B}u, \beta\rangle + \eta - \left[\frac{1}{2}v'H(z)v - p(z)'y\right] \right\} \geq 0.$$

The function in the braces is convex in $(z, u)$ and concave in $(y, v)$ and both sets, $\Omega$ and $\prod_{yv} F$, are closed and convex. In addition, $\Omega$ is bounded. By Sion's minimax theorem [28], we have

$$\min_{(z,u)\in\Omega} \max_{(y,v)\in\prod_{yv} F} \left\{ \langle \mathcal{A}z + \mathcal{B}u, \beta \rangle + \eta - \left[ \frac{1}{2} v' H(z) v - p(z)' y \right] \right\}$$

$$= \max_{(y,v)\in\prod_{yv} F} \min_{(z,u)\in\Omega} \left\{ \langle \mathcal{A}z + \mathcal{B}u, \beta \rangle + \eta - \left[ \frac{1}{2} v' H(z) v - p(z)' y \right] \right\}.$$

The first constraint in (12) is therefore equivalent to

$$\exists (y,v) \in \prod_{yv} F, \quad \forall (z,u) \in \Omega : \langle \mathcal{A}z + \mathcal{B}u, \beta \rangle + \eta - \left[ \frac{1}{2} v' H(z) v - p(z)' y \right] \geq 0,$$

which proves the Theorem. □

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

**Lemma 5** *Under the condition that one of $H_1, ..., H_N$ is positive definite or under the condition that one of the two systems*

$$\begin{aligned} & H - \sum_{m=1}^{N} H_m z_m = H_0, \\ & \mathcal{E}z + \mathcal{F}u \preceq_{\mathcal{K}} d, \\ & H \succeq 0 \end{aligned}$$

*and*

$$\begin{aligned} & \mathcal{E}^* \lambda + \begin{bmatrix} \langle H_1, \Lambda \rangle \\ \vdots \\ \langle H_N, \Lambda \rangle \end{bmatrix} = \mathcal{A}^* \beta + P' y, \\ & \mathcal{F}^* \lambda = \mathcal{B}^* \beta, \quad \lambda \in \mathcal{K}^*, \\ & \Lambda \succeq \frac{1}{2} v v' \end{aligned}$$

*has a Slater's point, the semi-infinite constraint in (13)*

$$\langle \mathcal{A}z + \mathcal{B}u, \beta \rangle + \eta - \left[ \frac{1}{2} v' H(z) v - p(z)' y \right] \geq 0 \quad \forall (z,u) \in \Omega \qquad (14)$$

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

$$\left.\begin{array}{l} \langle d, \lambda\rangle + \langle H_0, \Lambda\rangle \geq \eta + p_0' y, \\ \mathcal{E}^*\lambda + \begin{bmatrix} \langle H_1, \Lambda\rangle \\ \vdots \\ \langle H_N, \Lambda\rangle \end{bmatrix} = A'\beta + P'y, \\ \mathcal{F}^*\lambda = \mathcal{B}^*\beta, \\ \begin{bmatrix} 2 & v' \\ v & \Lambda \end{bmatrix} \succeq 0. \end{array}\right\} \tag{15}$$

**Proof.** Since $H(z) \succeq 0$ over $\Omega$, the constraint (14) means that the optimal value of the semidefinite program

$$\begin{array}{ll} \max_{H,z,u} & \frac{1}{2} v' H v - \langle \mathcal{A}z + \mathcal{B}u, \beta\rangle - y' P z, \\ \text{s. t.} & H - \sum_{m=1}^{N} H_m z_m = H_0, \\ & \mathcal{E}z + \mathcal{F}u \succeq_{\mathcal{K}} d, \\ & H \succeq 0 \end{array} \tag{16}$$

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

The dual problem to (16) is

$$\begin{array}{ll} \min_{\lambda,\Lambda} & \langle d, \lambda\rangle + \langle H_0, \Lambda\rangle \\ \text{s. t.} & \mathcal{E}^*\lambda + \begin{bmatrix} \langle H_1, \Lambda\rangle \\ \vdots \\ \langle H_N, \Lambda\rangle \end{bmatrix} = \mathcal{A}^*\beta + P'y, \\ & \mathcal{F}^*\lambda = \mathcal{B}^*\beta, \quad \lambda \in \mathcal{K}^*, \\ & \Lambda \succeq \frac{1}{2} v v'. \end{array} \tag{17}$$

If either (16) or (17) has a Slater's point, then by Lemma 1, strong duality holds for (16) and (17). Then the optimal value of (16) is less than or equal to $\eta + p' y$ if and only if its dual optimal value is so, namely,

$$\begin{array}{ll} \min_{\lambda,\Lambda} & (\langle d, \lambda\rangle + \langle H_0, \Lambda\rangle) \leq \eta + p_0' y \\ \text{s. t.} & \mathcal{E}^*\lambda + \begin{bmatrix} \langle H_1, \Lambda\rangle \\ \vdots \\ \langle H_N, \Lambda\rangle \end{bmatrix} = \mathcal{A}^*\beta + P'y \\ & \mathcal{F}^*\lambda = \mathcal{B}^*\beta, \lambda \in \mathcal{K}^*, \\ & \Lambda \succeq \frac{1}{2} v v'. \end{array} \tag{18}$$

Observe that

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

and

$$\Lambda \succeq \frac{1}{2}vv' \iff \begin{bmatrix} 2 & v' \\ v & \Lambda \end{bmatrix} \succeq 0.$$

Hence, (15) $\iff$ (14). In conclusion, if strong duality holds between (16) and (17), then the constraint (15) can be replaced by the set of constraints (16) in problem (13), and problem (P) is therefore a conic optimization problem.

It remains to show that if one of $H_1, ..., H_N$ is positive definite, then strong duality holds for problems (16) and (17). Since $H(z) \succeq 0$ over $\Omega$ and int $\Omega \neq \emptyset$, there exists $(z^0, u^0) \in \text{int}\ \Omega = \{(\mathrm{z},\mathrm{u}) : \mathcal{E}\mathrm{z} + \mathcal{F}\mathrm{u} \succ_{\mathcal{K}} \mathrm{d}\}$ that satisfy

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

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

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

and it satisfies

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

Therefore $(z^1, u^1, \bar{H})$ is a strictly feasible point (i.e., Slater's point) of (16), and (16) is always bounded above by $\eta + p_0'y$. Hence, by Lemma 1, strong duality holds between (16) and (17). The proof is completed. □

Since the semi-infinite constraint in (12) can be converted to a set of conic constraints, whose dimension is polynomial in the given data, we come up with the following final result, whose proof is evident.

**Theorem 2** *Under the existence of recourse assumption, the affine decision rule, the strong dual affine feasibility assumption, and the positive definiteness assumption on $H_i$ for some $i$, the quadratic two-stage stochastic optimization problem (P) with risk measures generated by the WKS format of ambiguity sets can be solved in polynomial time as a conic optimization problem.*

## 4 Concluding remarks

This notes sketches a framework of solving two-stage linear-quadratic stochastic optimization problems, where the second stage costs are certain coherent risk measures of the recourse costs. As described in [30], with appropriate construction of the auxiliary variables $u$, the model generalizes the linear two-stage distributionally robust models in, for instances, [2, 7, 14, 15, 17]. Since the model is quadratic, it is possible to apply it to certain stochastic equilibrium problems that the distributionally robust linear models may not be able to cover.

Possible extensions of this work could be problems of three or more stages and more general convex multistage stochastic optimization. On the modeling side, it would be interesting to study a multistage stochastic game problem with risk measures, which, as we mentioned in Section 1, is related to the recent seminal work of Rockafellar and Wets on multistage stochastic variational inequalities [26].

**Acknowledgment.** This research is partially supported by Australian Research Council under Grant DP160102819 and by a grant from General Research Fund (GRF) of Hong Kong.

## 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 representation 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. 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 (2013)
8. X. Chen, M. Sim, P. Sun, and J. Zhang. A linear-decision based approximation approach to stochastic programming. *Oper. Res.* **56** (2008) 344-357.
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. E. Delage and Y. Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. Oper. Res. **58** (2010) 595-612.
11. W. S. Dorn. Duality in quadratic programming. Quart. Appl. Math. **18** (1960) 155-162.
12. H. Föllmer and A. Schied. Stochastic Finance. Walter de Gruyter, Berlin (2002).
13. M. Fukushima and J.-S. Pang. Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games. Comput. Manag. Sci. **2** (2005) 21-56.
14. S. Gao, L. Kong, and J. Sun. Two-stage stochastic linear programs with moment information on uncertainty, Optimization **63** (2014) 829-837.
15. S. Gao, J. Sun, and S.-Y. Wu. Solving robust two-stage stochastic linear programs by a semi-infinite programming approach, Optim. Lett. (2017) DOI 10.1007/s11590-016-1095-4.
16. R. Hettich, K. O. Kortanek, Semi-Infinite programming: theory, methods, and applications, SIAM Rev. **35(3)** (1994) 380-429.
17. 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.

18. H-J. Lüthi and J. Doege. Convex risk measures for portfolio optimization and concepts of flexibility. Math. Program. **104** (2005) 541-559.
19. R. T. Rockafellar. Convex Analysis. Princeton University Press, NJ, (1970).
20. R. T. Rockafellar. Conjugate Duality and Optimization. AMS-SIAM Publication, Philadelphia, (1974).
21. R. T. Rockafellar. Coherent approaches to risk in optimization under uncertainty. INFORMS Tutorials in Operations Research (2007) 38-61.
22. R. T. Rockafellar, S. P. Uryasev, and M. Zabarankin. Generalized deviations in risk analysis. *Finance and Stochastics* **10** (2006) 51-74.
23. 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.
24. 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.
25. 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.
26. R. T. Rockafellar and R. J-B. Wets. Stochastic variational inequalities: single-stage to multistage. To appear in Math. Program. B. (2015).
27. A. Shapiro, D. Dentcheva, and A. Ruszczyski. Lectures on Stochastic Programming: Modeling and Theory. SIAM, Philadelphia (2009).
28. M. Sion. On general minimax theorems. Pacific J. Math. **8** (1958) 171-176.
29. 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.
30. W. Wiesemann, D. Kuhn and M. Sim. Distributionally robust convex optimization. Oper. Res. **62** (2014) 1358-1376.