Shabbir Ahmed

Convexity and decomposition of mean-risk stochastic programs

April 20, 2004. Revised February 8, 2005.

Abstract. Traditional stochastic programming is risk neutral in the sense that it is concerned with the optimization of an expectation criterion. A common approach to addressing risk in decision making problems is to consider a weighted mean-risk objective, where some dispersion statistic is used as a measure of risk. We investigate the computational suitability of various mean-risk objective functions in addressing risk in stochastic programming models. We prove that the classical mean-variance criterion leads to computational intractability even in the simplest stochastic programs. On the other hand, a number of alternative mean-risk functions are shown to be computationally tractable using slight variants of existing stochastic programming decomposition algorithms. We propose decomposition-based parametric cutting plane algorithms to generate mean-risk efficient frontiers for two particular classes of mean-risk objectives.

Key words. Stochastic programming, mean-risk objectives, computational complexity, decomposition, cutting plane algorithms.

---

1. Introduction

This paper is concerned with stochastic programming problems of the form

$$\min\{ \mathbb{E}[f(x,\omega)] : \ x \in X\}, \tag{1}$$

where $x \in \mathbb{R}^n$ is a vector of decision variables; $X \subset \mathbb{R}^n$ is a non-empty set of feasible decisions; $(\Omega, \mathcal{F}, P)$ is a probability space with elements ω; and $f : \mathbb{R}^n \times \Omega \mapsto \mathbb{R}$ is a cost function such that $f(\cdot,\omega)$ is convex for almost every (a.e.) $\omega \in \Omega$, and $f(x,\cdot)$ is $\mathcal{F}$-measurable and P-integrable for all $x \in \mathbb{R}^n$. Of particular interest are instances of (1) corresponding to two-stage stochastic linear programs (cf. [2,13]), where X is a polyhedron and

$$f(x,\omega) = c^T x + Q(x, \xi(\omega)) \tag{2}$$

with

$$Q(x,\xi) = \min\{q^T y : Wy = h + Tx, \ y \geq 0\}. \tag{3}$$

Here $\xi = (q, W, h, T)$ represents a particular realization of the random data $\xi(\omega)$ for the linear program in (3). It is assumed that problem (3) has a finite optimal value for all $x \in X$ and for a.e. $\omega \in \Omega$, i.e., relatively complete and sufficiently expensive recourse.

School of Industrial & Systems Engineering, Georgia Institute of Technology, 765 Ferst Drive, Atlanta, GA 30332. E-mail: sahmed@isye.gatech.edu

Formulation (1) is risk-neutral in the sense that it is concerned with the optimization of an expectation objective. A common approach to addressing risk is to consider a weighted mean-risk criterion, where some dispersion statistic is used as a proxy for risk, i.e.,

$$\min\{\ \mathbb{E}[f(x,\omega)] + \lambda \mathbb{D}[f(x,\omega)] :\ x \in X\}, \tag{4}$$

where $\mathbb{D}$ is a dispersion statistic, and λ is a non-negative weight to trade-off expected cost with risk. The classical Markowitz [8] mean-variance portfolio optimization model is an example of (4) where variance is used as the dispersion statistic.

In this paper we investigate various mean-risk objective functions and their computational suitability in addressing risk in stochastic programming models. We prove that the mean-variance criterion leads to computational intractability even in the simplest stochastic programs. On the other hand, a number of alternative mean-risk functions are shown to be computationally tractable using slight variants of existing stochastic programming decomposition algorithms. We propose parametric algorithms to generate mean-risk efficient frontiers for two specific mean-risk objectives in the context of two-stage stochastic linear programs.

2. Complexity of mean-variance stochastic programming

It is well-known that using the mean-variance criterion in two-stage stochastic linear programming, in general, leads to non-convex formulations [13, 18]. In this section, we prove the computational intractability of such non-convex formulations by showing that, even for very simple two-stage stochastic linear programs, the mean-variance criterion leads to NP-hard optimization problems.

Consider the class of two-stage stochastic linear programs with *simple recourse* (cf. [2]), i.e., problem (1) with

$$f(x,\omega) = \sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{n} Q_j(x_j, \xi_j(\omega)) \tag{5}$$

and

$$Q_j(x_j,\xi_j) = q_j^+(x_j - \xi_j)_+ + q_j^-(\xi_j - x_j)_+, \tag{6}$$

where $(\cdot)_+ = \max\{\cdot, 0\}$, and $q_j^+ + q_j^- \geq 0$ for all $j = 1, \ldots, n$. The following lemma provides a closed-form formula for the variance of the simple recourse function (6) when $\xi_j(\omega)$ has a discrete distribution. The proof follows from algebra and is omitted.

Lemma 1. *Consider the function $Q(x,\xi) = q^+(x-\xi)_+ + q^-(\xi - x)_+$, where $x, \xi \in \mathbb{R}$. Let $\xi(\omega)$ be a random variable with the discrete distribution $\xi(\omega) = \xi_k$*

w.p. p_k *for* $k = 1, \dots, K$, *and* $\xi_1 \le \dots \le \xi_K$. *The variance of* $Q(x, \xi(\omega))$ *is then given by*

$$\mathbb{V}[Q(x,\xi(\omega))] = \begin{cases} a & \text{if } x \le \xi_1, \\ b_k x^2 + c_k x + d_k & \text{if } \xi_{k-1} \le x \le \xi_k \quad k = 2, \dots, K, \\ e & \text{if } x \ge \xi_K, \end{cases}$$

where

$$a = (q^-)^2 \left[\sum_{i=1}^{K} p_i \xi_i^2 - \left(\sum_{i=1}^{K} p_i \xi_i \right)^2 \right],$$

$$b_k = \left(q^+ + q^-\right)^2 \left(\sum_{i=1}^{k-1} p_i \sum_{i=k}^{K} p_i \right),$$

$$c_k = -2 \left[(q^+)^2 \sum_{i=1}^{k-1} p_i \xi_i + (q^-)^2 \sum_{i=k}^{K} p_i \xi_i + \left(q^+ \sum_{i=1}^{k-1} p_i - q^- \sum_{i=k}^{K} p_i \right) \left(q^- \sum_{i=k}^{K} p_i \xi_i - q^+ \sum_{i=1}^{k-1} p_i \xi_i \right) \right],$$

$$d_k = \left[(q^+)^2 \sum_{i=1}^{k-1} p_i \xi_i^2 + (q^-)^2 \sum_{i=k}^{K} p_i \xi_i^2 - \left(q^+ \sum_{i=1}^{k-1} p_i \xi_i - q^- \sum_{i=k}^{K} p_i \xi_i \right)^2 \right],$$

$$e = (q^+)^2 \left[\sum_{i=1}^{K} p_i \xi_i^2 - \left(\sum_{i=1}^{K} p_i \xi_i \right)^2 \right].$$

From the above lemma, it can be seen that the function $\mathbb{V}[Q(x, \xi(\omega))]$ is piecewise convex quadratic (note that $b_k \ge 0$ for all k), but non-convex in general. Consequently, we encounter computational complications in using the mean-variance criterion in stochastic programs with simple recourse.

Theorem 1. *The mean-variance stochastic programming problem*

$$\min\{\ \mathbb{E}[f(x,\xi)] + \lambda \mathbb{V}[f(x,\xi)] :\ x \in X\} \tag{7}$$

corresponding to the simple recourse function (5)-(6) is NP-hard for any $\lambda > 0$.

Proof. Consider the Binary Integer Feasibility problem:

> *Given an integer matrix* $A \in \mathbb{Z}^{m \times n}$, *and integer vector* $b \in \mathbb{Z}^m$,
> *is there a vector* $x \in \{-1, 1\}^n$ such that $Ax \le b$? (8)

The binary integer feasibility problem (8) is known to be NP-complete [3]. We shall show that given any instance of (8) with n variables, we can construct a polynomial (in n) sized instance of the mean-variance stochastic program (7) for any $\lambda > 0$ such that (8) has an answer "yes" if and only if (7) has an optimal objective value of $(3 + \frac{3}{4\lambda} + \frac{3\lambda}{2})n$.

An instance of (8) is given by the data pair (A, b). Given any such instance, we can construct an instance of (7) for any $\lambda > 0$ as follows. Let $X = \{x \in \mathbb{R}^n : Ax \leq b, \ -e \leq x \leq e\}$, e be a n-vector of ones, and $q_j^+ = 1, q_j^- = 1, c_j = 0$ for all $j = 1, \ldots, n$. Let $\xi_j(\omega)$ for $j = 1, \ldots, n$ be i.i.d random variables, with the distribution

$$\xi_j(\omega) = \begin{cases} -3 - \frac{1}{\lambda} & \text{w.p. } \frac{1}{4}, \\ 0 & \text{w.p. } \frac{1}{2}, \\ 3 + \frac{1}{\lambda} & \text{w.p. } \frac{1}{4}. \end{cases}$$

Owing to the independence of $\xi_j(\omega)$, the mean-variance stochastic program (7) corresponding to the above data reduces to

$$\min\{\sum_{j=1}^{n}(\mathbb{E}[Q_j(x_j, \xi_j(\omega))] + \lambda\mathbb{V}[Q_j(x_j, \xi_j(\omega))]) : \ x \in X\}. \tag{9}$$

It follows from Lemma 1 that for all $j = 1, \ldots, n$,

$$\mathbb{E}[Q_j(x_j, \xi_j(\omega))] + \lambda\mathbb{V}[Q_j(x_j, \xi_j(\omega))] =$$

$$\begin{cases} \frac{3}{4\lambda}(4\lambda + 1 + \lambda^2 x_j^2 + 2\lambda^2 x_j + 3\lambda^2) & \text{if} \quad -3 - \frac{1}{\lambda} \leq x_j \leq 0 \\ \frac{3}{4\lambda}(4\lambda + 1 + \lambda^2 x_j^2 - 2\lambda^2 x_j + 3\lambda^2) & \text{if} \quad 0 \leq x_j \leq 3 + \frac{1}{\lambda}. \end{cases}$$

Note that for all $j = 1, \ldots, n$, $\mathbb{E}[Q_j(x_j, \xi_j(\omega))] + \lambda\mathbb{V}[Q_j(x_j, \xi_j(\omega))] \geq (3 + \frac{3}{4\lambda} + \frac{3\lambda}{2})$ for any $x_j \in [-1, 1]$, with equality holding if and only if $x_j \in \{-1, 1\}$. Thus (9) has an optimal objective value of $(3 + \frac{3}{4\lambda} + \frac{3\lambda}{2})n$ if and only if there exists $x \in X$ such that $x \in \{-1, 1\}^n$, i.e., problem (8) has an affirmative answer. □

In the classical setting of portfolio optimization, the function $f(x, \omega) = -r(\omega)^T x$ where $r(\omega) \in \mathbb{R}^n$ is a random vector of returns, and $X \subset \mathbb{R}^n$ is a polyhedral set of feasible weights for the n assets in the portfolio. In this case, $\mathbb{V}[f(x, \omega)] = x^T C x$ where C is the covariance matrix of the random vector $r(\omega)$. Consequently, (7) reduces to a deterministic (convex) quadratic program suitable for very efficient computation. For typical stochastic programs $f(x, \omega)$ is nonlinear (although convex) in x. Furthermore, the variance operator, although convex, is non-monotone. Consequently $\mathbb{E}[f(x, \omega)] + \lambda\mathbb{V}[f(x, \omega)]$ is not guaranteed to be convex in x, leading to the computational complication proven in Theorem 1. In the following section, we investigate mean-risk objectives that preserve convexity, hence computational tractability.

3. Tractable mean-risk objectives

Given a random variable $Y : \Omega \mapsto \mathbb{R}$, representing cost, belonging to the linear space $\mathcal{X}_p = \mathcal{L}_p(\Omega, \mathcal{F}, P)$ for $p \geq 1$, a scalar $\lambda \geq 0$, and an appropriate function $\mathbb{D} : \mathcal{X}_p \mapsto \mathbb{R}$ to measure the risk associated with Y, we define a mean-risk function $g_{\lambda,\mathbb{D}} : \mathcal{X}_p \mapsto \mathbb{R}$ as

$$g_{\lambda,\mathbb{D}}[Y] = \mathbb{E}[Y] + \lambda\mathbb{D}[Y]. \tag{10}$$

Using (10) to address risk in the context of the stochastic program (1), we arrive at the formulation

$$\min\{\phi(x) = g_{\lambda,\mathbb{D}}[f(x,\omega)] : \ x \in X\}. \tag{11}$$

From a computational viewpoint, it is desirable that the objective function $\phi(\cdot)$ in (11) be convex. As discussed in Section 2, even though $f(\cdot,\omega)$ is convex for a.e. $\omega \in \Omega$, the convexity of the composite function $\phi(x) = g_{\lambda,\mathbb{D}}[f(x,\omega)]$ may not be preserved, for example, if variance is used as the measure of risk in $g_{\lambda,\mathbb{D}}$.

3.1. Sufficient conditions for preserving convexity

We shall say that a function $g : \mathcal{X}_p \mapsto \mathbb{R}$ is *convexity-preserving*, if the composite function $\phi(x) = g[f(x,\omega)]$ is convex for any function $f(x,\omega)$ such that $f(\cdot,\omega)$ is convex for a.e. $\omega \in \Omega$ and $f(x,\cdot) \in \mathcal{X}_p$ for all $x \in \mathbb{R}^n$. Recall that a function $g : \mathcal{X}_p \mapsto \mathbb{R}$ is *convex* if $g[\lambda Y_1 + (1-\lambda)Y_2] \leq \lambda g[Y_1] + (1-\lambda)g[Y_2]$, for all $Y_1, Y_2 \in \mathcal{X}_p$ and $\lambda \in [0,1]$; is *non-decreasing* if $g[Y_1] \geq g[Y_2]$ for all $Y_1, Y_2 \in \mathcal{X}_p$ such that $Y_1 \geq Y_2$; and is *positively homogenous* if $g[\lambda Y] = \lambda g[Y]$ for all $Y \in \mathcal{X}_p$ and $\lambda \geq 0$.

The following result is well-known (cf. [14]).

Proposition 1. *If $g : \mathcal{X}_p \mapsto \mathbb{R}$ is convex and non-decreasing, then g is convexity-preserving.*

Lemma 2. *A convex and positively homogenous function $g : \mathcal{X}_p \mapsto \mathbb{R}$ is non-decreasing if and only if it satisfies $g[Y] \leq 0$ for all $Y \leq 0$.*

Proof. Suppose g satisfies $g[Y] \leq 0$ for all $Y \leq 0$. Let $Y_1 \geq Y_2$, i.e., $Y_2 = Y_1 + \Delta$ for some $\Delta \leq 0$. Then

$$\begin{aligned}
\tfrac{1}{2}g[Y_2] &= \tfrac{1}{2}g[Y_1 + \Delta] \\
&= g[\tfrac{1}{2}Y_1 + \tfrac{1}{2}\Delta] \\
&\leq \tfrac{1}{2}g[Y_1] + \tfrac{1}{2}g[\Delta] \\
&\leq \tfrac{1}{2}g[Y_1],
\end{aligned}$$

where the second line follows from positive-homogeneity, the third line follows from convexity, and the fourth line follows from the fact that $g[\Delta] \leq 0$ (since $\Delta \leq 0$). Thus $g[\cdot]$ is non-decreasing. Conversely, if $g[\cdot]$ is non-decreasing, then $g[Y] \leq g[0] = 0$ for all $Y \leq 0$. □

The following result immediately follows from Lemma 2 and Proposition 1.

Proposition 2. *If $g : \mathcal{X}_p \mapsto \mathbb{R}$ is convex, positively homogenous, and satisfies $g[Y] \leq 0$ for all $Y \leq 0$, then g is convexity-preserving.*

3.2. Examples

Here we show that a number of common mean-risk objectives are convexity preserving, and hence suitable for optimization.

Semideviation from a target. For $Y \in \mathcal{X}_p$ and a fixed target $T \in \mathbb{R}$, the p-th Semideviation from T [11] is defined as

$$S_{T,p}[Y] = (\mathbb{E}[(Y-T)_+^p])^{1/p}.$$

Proposition 3. *The mean-risk objective*

$$g_{\lambda,u_{T,p}}[Y] = \mathbb{E}[Y] + \lambda S_{T,p}[Y]$$

is convexity preserving for all $p \geq 1$ and $\lambda \geq 0$.

Proof. Since $S_{T,p}[Y]$ is convex and non-decreasing in Y for all $p \geq 1$, the result follows from Proposition 1. □

Central semideviation. For $Y \in \mathcal{X}_p$, the p-th central semideviation [9] is defined as

$$\delta_p[Y] = \left(\mathbb{E}[(Y-\mathbb{E}Y)_+^p]\right)^{\frac{1}{p}}.$$

Proposition 4. *The mean-semideviation objective*

$$g_{\lambda,\delta_p}[Y] = \mathbb{E}[Y] + \lambda\delta_p[Y]$$

is convexity-preserving for all $p \geq 1$ and $\lambda \in [0,1]$.

Proof. Ogryczak and Ruszczyński [9] have shown that δ_p is convex for all $p \geq 1$. Furthermore, it can be verified that δ_p is positively-homogenous. Thus $g_{\lambda,\delta_p}[Y]$ is convex and positively homogenous for all $\lambda \geq 0$. Moreover, if $Y \leq 0$, then $(Y-\mathbb{E}Y)_+ \leq -\mathbb{E}Y$. Thus $g_{\lambda,\delta_p}[Y] \leq (1-\lambda)\mathbb{E}Y \leq 0$ for all $\lambda \leq 1$. The result then follows from Proposition 2. □

Quantile-deviation and CVaR. Given $\alpha \in (0,1)$, the Quantile-deviation [10] for $Y \in \mathcal{X}_1$ is defined as

$$h_\alpha[Y] = \mathbb{E}\left[(1-\alpha)(\kappa_\alpha[Y]-Y)_+ + \alpha(Y-\kappa_\alpha[Y])_+\right],$$

where κ_α is α-quantile of the distribution of Y, i.e., $\Pr(Y \leq \kappa_\alpha[Y]) \geq \alpha$ and $\Pr(Y \geq \kappa_\alpha[Y]) \geq 1-\alpha$. In [10], it is shown that h_α is equivalent to the following risk measure (defined for $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$),

$$h_{\varepsilon_1,\varepsilon_2}[Y] = \min_{z\in\mathbb{R}}\{\mathbb{E}[\varepsilon_1(z-Y)_+ + \varepsilon_2(Y-z)_+]\},$$

when $\alpha = \varepsilon_2/(\varepsilon_1+\varepsilon_2)$. The α-Conditional value at risk [12] for $Y \in \mathcal{X}_1$ is defined as

$$\mathrm{CVaR}_\alpha[Y] = \min_{z\in\mathbb{R}}\{z + \frac{1}{1-\alpha}\mathbb{E}[(Y-z)_+]\}.$$

By simple algebra, CVaR_α can be expressed in terms of $h_{\varepsilon_1,\varepsilon_2}$ as follows [14]:

$$\mathrm{CVaR}_\alpha[Y] = \mathbb{E}[Y] + \frac{1}{\varepsilon_1}h_{\varepsilon_1,\varepsilon_2}[Y],$$

where $\alpha = \varepsilon_2/(\varepsilon_1+\varepsilon_2)$.

Proposition 5. *Consider $Y \in \mathcal{X}_1$.*

(i) Given $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, the mean-risk objective

$$g_{\lambda,h_{\varepsilon_1,\varepsilon_2}}[Y] = \mathbb{E}[Y] + \lambda h_{\varepsilon_1,\varepsilon_2}[Y]$$

is convexity preserving for all $\lambda \in [0, 1/\varepsilon_1]$.

(ii) Given $\alpha \in (0,1)$, the mean-quantile deviation objective

$$g_{\lambda,h_\alpha}[Y] = \mathbb{E}[Y] + \lambda h_\alpha[Y]$$

is convexity preserving for all $\lambda \in [0, 1/(1-\alpha)]$.

(iii) Given $\alpha \in (0,1)$, the mean-CVaR objective

$$g_{\lambda,\mathrm{CVaR}_\alpha}[Y] = \mathbb{E}[Y] + \lambda \mathrm{CVaR}_\alpha[Y]$$

is convexity preserving for all $\lambda \geq 0$.

Proof. Ogryczak and Ruszczyński [10] have shown that $h_{\varepsilon_1,\varepsilon_2}[Y]$ is convex and positively homogenous. Thus $g_{\lambda,h_{\varepsilon_1,\varepsilon_2}}[Y]$, $g_{\lambda,h_\alpha}[Y]$ and $g_{\lambda,\mathrm{CVaR}_\alpha}[Y]$ are convex and positively homogenous for all $\lambda \geq 0$. Note that $h_{\varepsilon_1,\varepsilon_2}[Y] \leq \varepsilon_1 \mathbb{E}(-Y)_+ + \varepsilon_2 \mathbb{E}(Y)_+$, and if $Y \leq 0$, then $h_{\varepsilon_1,\varepsilon_2}[Y] \leq -\varepsilon_1 \mathbb{E}[Y]$. Thus $g_{\lambda,h_{\varepsilon_1,\varepsilon_2}}[Y] \leq (1 - \lambda\varepsilon_1)\mathbb{E}[Y] \leq 0$ since $\lambda \leq 1/\varepsilon_1$. Statement (i) then follows from Proposition 2. Statement (ii) follows from the equivalence of $h_\alpha[Y]$ and $h_{\varepsilon_1,\varepsilon_2}[Y]$ when $\alpha = \varepsilon_2/(\varepsilon_1 + \varepsilon_2)$. Note that $g_{\lambda,\mathrm{CVaR}_\alpha}[Y] = (1+\lambda)\mathbb{E}[Y] + \lambda/\varepsilon_1 h_{\varepsilon_1,\varepsilon_2}[Y]$. If $Y \leq 0$, $g_{\lambda,\mathrm{CVaR}_\alpha}[Y] \leq (1+\lambda)\mathbb{E}[Y] - \lambda\mathbb{E}[Y] = \mathbb{E}[Y] \leq 0$ for all λ. Statement (iii) then follows. □

Gini mean difference. For $Y \in \mathcal{X}_1$ with distribution function F_Y, the Gini mean difference [19] is defined as

$$\Gamma[Y] = \int \mathbb{E}[(\xi - Y)_+] dF_Y(\xi).$$

Proposition 6. *The mean-Gini mean difference objective*

$$g_{\lambda,\Gamma}[Y] = \mathbb{E}[Y] + \lambda \Gamma[Y]$$

is convexity preserving for all $\lambda \in [0,1]$.

Proof. Ogryczak and Ruszczyński [10] have shown that $\Gamma[\cdot]$ is convex and positively-homogenous. Thus $g_{\lambda,\Gamma}[\cdot]$ is convex and positively-homogenous for all $\lambda \geq 0$. Moreover, if $Y \leq 0$ then

$$\int \mathbb{E}[(\xi - Y)_+] dF_Y(\xi) \leq \int -\mathbb{E}[Y] dF_Y(\xi) = -\mathbb{E}[Y].$$

Thus $g_{\lambda,\Gamma}[Y] \leq (1-\lambda)\mathbb{E}[Y] \leq 0$ for all $\lambda \leq 1$. The result then follows from Proposition 2. □

The non-decreasing, hence convexity preserving, property of g_{λ,δ_p}, g_{λ,h_α} (or, equivalently, $g_{\lambda,h_{\varepsilon_1,\varepsilon_2}}$) , and $g_{\lambda,\Gamma}$ may not hold for $\lambda > 1$, $\lambda > 1/(1-\alpha)$ (or, $\lambda > 1/\varepsilon_1$), and $\lambda > 1$, respectively. However, as shown in [9,10], these mean-risk objectives are guaranteed to be consistent with standard stochastic ordering rules only if $\lambda \in (0,1)$, $\lambda \in (0,1/(1-\alpha))$ (or, $\lambda \in (0,1/\varepsilon_1)$) and $\lambda \in (0,1)$, respectively[1].

4. Solving mean-risk stochastic programs

In this section, we discuss methods for solving stochastic programs with convexity preserving mean-risk objectives. We provide specific details for two particular classes of mean-risk stochastic programs, those that involve the semideviation risk measure δ_p with $p = 1$, i.e.,

$$\delta_1[Y] = \mathbb{E}[(Y - \mathbb{E}Y)_+]$$

which is referred to as the absolute semideviation (ASD), and those that involve the quantile deviation (QDEV) risk measure

$$h_{\varepsilon_1,\varepsilon_2}[Y] = \min_{z\in\mathbb{R}}\{\mathbb{E}[\varepsilon_1(z-Y)_+ + \varepsilon_2(Y-z)_+]\},$$

for $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$.

4.1. Deterministic equivalent formulations

Consider the stochastic program (1) where $f(x,\omega)$ is given as the value function of a second-stage optimization problem

$$f(x,\omega) = \min\{F_0(x,y,\omega):\ F_i(x,y,\omega) \le 0\ i=1,\dots,m,\ y \in Y\}.$$

Here we make the necessary assumptions on F_i for $i = 0,\dots,m$ and Y for $f(x,\omega)$ to be finite-valued, $\mathcal{F}$-measurable and P-integrable (cf. [13]). The two-stage stochastic linear programming objective function (2)-(3) is a special case. If the mean-risk objective is non-decreasing, then the corresponding mean-risk stochastic program (11) can be written as:

$$\begin{array}{ll} \min\limits_{x,y(\omega)} & g_{\lambda,\mathbb{D}}[F_0(x,y(\omega),\omega)] \\ \text{s.t.} & x \in X \\ & \left.\begin{array}{l} F_i(x,y(\omega),\omega) \le 0 \quad i=1,\dots,m \\ y(\omega) \in Y \end{array}\right\} \text{ for a.e. } \omega \in \Omega. \end{array} \tag{12}$$

When Ω is finite, problem (12) is a large-scale deterministic optimization problem. Furthermore, if F_i for $i = 0,1,\dots,m$ are convex functions of the decision variables x and $y(\omega)$ for a.e. $\omega \in \Omega$, and the sets X and Y are convex, then (12) is a convex program.

[1] Note that in our minimization setting smaller values of the random variables are preferable to larger, whereas in the maximization setting of [10], larger values are preferable. Consequently, for the quantile deviation risk function, [10] shows consistency with stochastic ordering for $\lambda \in (0,1/\alpha)$

4.1.1. The Mean-ASD objective. It can be verified that

$$g_{\lambda,\delta_1}[Y] = (1-\lambda)\mu[Y] + \lambda\nu[Y], \tag{13}$$

where $\mu[Y] = \mathbb{E}[Y]$ and $\nu[Y] = \mathbb{E}[\max\{Y, \mathbb{E}Y\}]$. The deterministic equivalent (12) of the mean-ASD stochastic program is then (cf. [16]):

$$\begin{array}{ll} \min\limits_{x,y(\omega),\nu(\omega)} & (1-\lambda)\int F_0(x,y(\omega),\omega)dP(\omega) + \lambda\int \nu(\omega)dP(\omega) \\ \text{s.t.} & x \in X \\ & \left.\begin{array}{l} F_i(x,y(\omega),\omega) \le 0 \quad i=1,\dots,m \\ \nu(\omega) \ge F_0(x,y(\omega),\omega) \\ \nu(\omega) \ge \int F_0(x,y(\xi),\xi)dP(\xi) \\ y(\omega) \in Y \end{array}\right\} \text{ for a.e. } \omega \in \Omega. \end{array} \tag{14}$$

Note that, if $\lambda \in [0,1]$, F_i for $i = 0,1,\dots,m$ are convex functions of the decision variables x and $y(\omega)$ for a.e. $\omega \in \Omega$, and the sets X and Y are convex, then the objective function and constraints of (14) are all convex. In case of two-stage stochastic linear programs, the functions F_i for $i = 0,\dots,m$ are linear and the sets X and Y are polyhedral. Then, in case of a finite set Ω of scenarios, (14) is a large-scale linear program. Unfortunately, this linear program does not posses the dual-block angular structure that is natural in the deterministic equivalents of standard two-stage stochastic linear programs [16]. Consequently, decomposition methods (such as the Benders or L-shaped algorithm) (cf. Chapter 3 of [13]) used for solving standard two-stage stochastic linear programs cannot be directly applied. In the next section, we show that problem decomposition can be achieved by using slight variations of these methods.

4.1.2. The Mean-QDEV objective. The mean-QDEV objective function can be written as

$$\begin{aligned} g_{\lambda,h_{\varepsilon_1,\varepsilon_2}}[Y] &= \mathbb{E}[Y] + \lambda \min_{z\in\mathbb{R}}\{\mathbb{E}[\varepsilon_1(z-Y)_+ + \varepsilon_2(Y-z)_+]\} \\ &= \min_{z\in\mathbb{R}}\{\mathbb{E}[Y + \lambda\varepsilon_1(z-Y)_+ + \lambda\varepsilon_2(Y-z)_+]\} \\ &= \min_{z\in\mathbb{R}}\{\mathbb{E}[Y + \lambda\varepsilon_1(Y-z)_+ - \lambda\varepsilon_1(Y-z) + \lambda\varepsilon_2(Y-z)_+]\} \\ &= \min_{z\in\mathbb{R}}\{\lambda\varepsilon_1 z + (1-\lambda\varepsilon_1)\mathbb{E}[Y] + \lambda(\varepsilon_1+\varepsilon_2)\mathbb{E}[(Y-z)_+]\}. \end{aligned} \tag{15}$$

From the fact that $\mathbb{E}[Y]$ and $\mathbb{E}[(Y-z)_+]$ are non-decreasing in Y and $\lambda \in [0, 1/\varepsilon_1]$, it follows that deterministic equivalent (12) of the mean-QDEV stochastic program is

$$\begin{array}{ll} \min\limits_{z,x,y(\omega),\nu(\omega)} & \lambda\varepsilon_1 z + (1-\lambda\varepsilon_1)\int F_0(x,y(\omega),\omega)dP(\omega) + \lambda(\varepsilon_1+\varepsilon_2)\int \nu(\omega)dP(\omega) \\ \text{s.t.} & z \in \mathbb{R},\ x \in X \\ & \left.\begin{array}{l} F_i(x,y(\omega),\omega) \le 0 \quad i=1,\dots,m \\ \nu(\omega) \ge F_0(x,y(\omega),\omega) - z \\ \nu(\omega) \ge 0,\ y(\omega) \in Y \end{array}\right\} \text{ for a.e. } \omega \in \Omega. \end{array} \tag{16}$$

If F_i for $i = 0, 1, \dots, m$ are convex functions of the decision variables x and $y(\omega)$ for a.e. $\omega \in \Omega$, and the sets X and Y are convex, then the objective function and constraints of (16) are all convex. Moreover, in case of two-stage stochastic linear programs, with a finite set Ω of scenarios, (16) is a large-scale linear program with dual-block angular structure and is directly amenable to standard decomposition methods [17]. In the following two sections, we propose such a decomposition algorithm and its parametric extension to efficiently generate the entire mean-QDEV efficient frontier.

4.2. Cutting plane methods

When $f(\cdot, \omega)$ is convex for a.e. $\omega \in \Omega$ and the mean-risk function $g_{\lambda,\mathbb{D}}[\cdot]$ is convexity-preserving, the mean-risk stochastic program (11) involves minimizing a convex (often non-smooth) objective function $\phi(x) = g_{\lambda,\mathbb{D}}[f(x, \omega)]$. Effective solution schemes for such problems are subgradient-based methods, such as various cutting plane algorithms, bundle methods, or level methods [4,5]. Given a candidate solution $x \in \mathbb{R}^n$, these methods require the calculation of a subgradient s of the composite function $\phi(\cdot) = g_{\lambda,\mathbb{D}}[f(\cdot, \omega)]$ at x, i.e., $s \in \partial\phi(x)$. Note that a subgradient of $g_{\lambda,\mathbb{D}}$ always exists if $g_{\lambda,\mathbb{D}}$ is real-valued, convex, and continuous. Consequently, given subgradients of f and $g_{\lambda,\mathbb{D}}$, a subgradient of $\phi(\cdot)$ can be calculated using the chain rule of subdifferentiation. Subgradient formulas for various mean-risk objectives, including mean-ASD and mean-QDEV, are derived in detail in [14]. These subgradient formulas can be used for the optimization of the mean-risk stochastic program (11) using, for example, a generic cutting plane algorithm, such as Algorithm 1, or some variant of it.

Algorithm 1 A cutting plane algorithm for the mean-risk stochastic program (11).

```
let ϵ ≥ 0 be a pre-specified tolerance; set i = 1, LB = −∞ and UB = +∞.
while UB − LB ≥ ϵ do
   solve the following master problem
         LB = min_{x,θ} {θ : x ∈ X, θ ≥ φ(x^j) + (s^j)^T (x − x^j), j = 1, ..., i − 1};
   and let x^i be its optimal solution (or any feasible solution if an optimal does not exist).
   for ω ∈ Ω do
      compute f(x^i, ω) and π^i(ω) ∈ ∂f(x^i, ω).
   end for
   compute φ(x^i) and s^i ∈ ∂φ(x^i), from f(x^i, ω) and π^i(ω).
   set UB = min{UB, φ(x^i)}.
end while
```

In case of two-stage stochastic linear programs of the form (2)-(3), a subgradient $\pi(\omega) \in \partial f(x, \omega)$ is given by $c + T^T\vartheta(\omega)$ where $\vartheta(\omega)$ is a dual optimal solution to the second-stage linear program (3) for given x and realization ω. When Ω is finite, the second-stage subproblems corresponding to a given x can be solved independently for each realization $\omega \in \Omega$ allowing for a computationally convenient decomposition. The optimal objective values and the dual solutions for the subproblems can then be used to compute the function value and its subgradient. This scheme is a slight variation of the well-known L-shaped (or Benders) decomposition method for solving standard two-stage stochastic linear programs involving an expected value objective. If the mean-risk function is polyhedral (as in case of the mean-ASD and mean-QDEV objectives), then Algorithm 1 is guaranteed to terminate in a finite number of iterations with an ϵ-optimal solution for any $\epsilon \geq 0$.

4.3. Parametric cutting plane algorithms for mean-ASD and mean-QDEV stochastic linear programs

Often, it is necessary to solve the mean-risk stochastic program (11) for many different values of the mean-risk tradeoff parameter λ so as to trace out the mean-risk efficient frontier. This can, of course, be accomplished by repeatedly solving the problem for different values of λ. However, more efficient parametric optimization schemes may be possible. For the case when $f(x,\omega)$ is linear in x, as in the portfolio optimization setting, a parametric simplex algorithm for generating mean-ASD and mean-QDEV frontiers has been proposed in [15]. In this section, we propose parametric cutting plane algorithms to generate mean-ASD and mean-QDEV frontiers when $f(x,\omega)$ corresponds to two-stage stochastic linear programs. Note that in Algorithm 1, the tradeoff parameter λ is embedded within the subgradient s^i calculated in step 7, and hence in the cut coefficients. As such, this algorithm is not suitable for an efficient parametric analysis of the mean-risk model with respect to λ.

Recall from (13), that the mean-ASD stochastic program is

$$\min\Big\{g_{\lambda,\delta_1}[f(x,\omega)] = (1-\lambda)\mu[f(x,\omega)] + \lambda\nu[f(x,\omega)] : \ x \in X\Big\}, \tag{17}$$

where $\mu[Y] = \mathbb{E}Y$ and $\nu[Y] = \mathbb{E}[\max\{Y, \mathbb{E}Y\}]$. Note that both μ and ν are convex non-decreasing functions, and so the composite functions $\mu[f(x,\omega)]$ and $\nu[f(x,\omega)]$ are convex in x that can be approximated by subgradient inequalities (cuts) within a cutting plane scheme. If we use a cutting plane scheme for (17) where separate cuts are used to approximate the functions μ and ν, then the cut coefficients are independent of the tradeoff weight λ which only appears in the objective function of the master problem. Thus the cuts are valid for any $\lambda \in [0,1]$. If X is polyhedral, then the master problem is a linear program for which a parametric analysis with respect to the objective coefficient λ can be easily carried out to detect the range of λ for which the current master problem basis remains optimal. We can then choose a λ outside this range and reoptimize. In this way we can construct the efficient frontier for the entire range of $\lambda \in [0,1]$. This parametric modification of the cutting plane Algorithm 1 is summarized in Algorithm 2.

From (15), we have that the mean-QDEV stochastic program is

$$\begin{aligned}\min\Big\{g_{\lambda,h_{\varepsilon_1,\varepsilon_2}}[f(x,\omega)] = \min_{z\in\mathbb{R}}\{&\lambda\varepsilon_1 z + (1-\lambda\varepsilon_1)\mu[f(x,\omega)] \\ &+\lambda(\varepsilon_1+\varepsilon_2)\nu[f(x,\omega),z]\} : x \in X\Big\},\end{aligned} \tag{18}$$

Algorithm 2 A parametric cutting plane algorithm for the mean-ASD stochastic program (17) (the set X is polyhedral).

1: let $\epsilon \geq 0$ be a pre-specified tolerance; set $\lambda = 0$, $i = 1$, $LB = -\infty$ and $UB = +\infty$.
2: **while** $UB - LB \geq \epsilon$ **do**
3: solve the following *master* linear program

$$LB = \min_{x,\theta,\eta} \{(1-\lambda)\theta + \lambda\eta : \ x \in X, \ \theta \geq \mu(x^j) + (u^j)^T(x - x^j) \ j = 1,\ldots,i-1, \\ \eta \geq \nu(x^j) + (v^j)^T(x - x^j) \ j = 1,\ldots,i-1\};$$

and let x^i be its optimal solution (or any feasible solution if an optimal does not exist).
4: **for** $\omega \in \Omega$ **do**
5: compute $f(x^i,\omega)$ and $\pi^i(\omega) \in \partial f(x^i,\omega)$.
6: **end for**
7: compute $\mu(x^i) = \mathbb{E}[f(x^i,\omega)]$ and a subgradient $u^i = \mathbb{E}[\pi^i(\omega)]$.
8: compute $\nu(x^i) = \mathbb{E}[\max\{f(x^i,\omega), \mu(x^i)\}]$
and a subgradient $v^i = \mathbb{E}[\iota(\omega)\pi^i(\omega)] + u^i\mathbb{E}[1 - \iota(\omega)]$ where $\iota(\omega) = 1$ if $f(x^i,\omega) > \mu(x^i)$ and 0 otherwise.
9: set $UB = \min\{UB, (1-\lambda)\mu(x^i) + \lambda\nu(x^i)\}$.
10: **end while**
11: use parametric analysis on the master problem to find the range $[\lambda, \lambda^*]$ for which the current master problem basis remains optimal.
12: if $\lambda^* < 1$, set $\lambda = \min\{1, \lambda^* + \delta\}$ (where $\delta > 0$ is very small), $LB = -\infty$ and $UB = +\infty$ and return to step 2.

Algorithm 3 A parametric cutting plane algorithm for the mean-QDEV stochastic program (18) (the set X is polyhedral).

1: let $\epsilon \geq 0$ be a pre-specified tolerance; set $\lambda = 0$, $i = 1$, $LB = -\infty$ and $UB = +\infty$.
2: **while** $UB - LB \geq \epsilon$ **do**
3: solve the following *master* linear program

$$LB = \min_{z,x,\theta,\eta} \{\lambda\varepsilon_1 z + (1-\lambda\varepsilon_1)\theta + \lambda(\varepsilon_1+\varepsilon_2)\eta : \ z \in \mathbb{R}, \ x \in X, \\ \theta \geq \mu(x^j) + (u^j)^T(x - x^j) \ \ j = 1,\ldots,i-1, \\ \eta \geq \nu(z^j, x^j) + (v_z^j)^T(z - z^j) + (v_x^j)^T(x - x^j) \ \ j = 1,\ldots,i-1\};$$

and let (z^i, x^i) be its optimal solution (or any feasible solution if an optimal does not exist).
4: **for** $\omega \in \Omega$ **do**
5: compute $f(x^i,\omega)$ and $\pi^i(\omega) \in \partial f(x^i,\omega)$.
6: **end for**
7: compute $\mu(x^i) = \mathbb{E}[f(x^i,\omega)]$ and a subgradient $u^i = \mathbb{E}[\pi^i(\omega)]$.
8: compute $\nu(x^i,z^i) = \mathbb{E}[(f(x^i,\omega) - z^i)_+]$, a subgradient with respect to x: $v_x^i = \mathbb{E}[\iota(\omega)\pi^i(\omega)]$, and a subgradient with respect to z: $v_z^i = -\mathbb{E}[\iota(w)]$, where $\iota(\omega) = 1$ if $f(x^i,\omega) > z^i$ and 0 otherwise.
9: set $UB = \min\{UB, \lambda\varepsilon_1 z^i + (1-\lambda\varepsilon_1)\mu(x^i) + (\lambda\varepsilon_1+\varepsilon_2)\nu(x^i,z^i)\}$.
10: **end while**
11: use parametric analysis on the master problem to find the range $[\lambda, \lambda^*]$ for which the current master problem basis remains optimal.
12: if $\lambda^* < 1/\varepsilon_1$, set $\lambda = \min\{1/\varepsilon_1, \lambda^* + \delta\}$ (where $\delta > 0$ is very small), $LB = -\infty$ and $UB = +\infty$ and return to step 2.

where $\mu[Y] = \mathbb{E}Y$ and $\nu[Y,z] = \mathbb{E}[(Y-z)_+]$. Note that both μ and ν are convex non-decreasing in Y, and so the composite functions $\mu[f(x,\omega)]$ and $\nu[f(x,\omega),z]$ are convex in x. Moreover ν is also convex in z. Thus, μ and ν can be approximated separately by cutting planes whose coefficients are independent of λ. Consequently, we can develop a parametric cutting plane algorithm similar to Algorithm 2 for the mean-QDEV objective. A key difference is that the optimization is with respect to x and z. The details are provided in Algorithm 3.

4.4. Computational illustration

We implemented enhanced versions of Algorithms 2 and 3 to generate the mean-ASD frontier and the mean-QDEV frontier, respectively. Our implementations extend the basic schemes of Algorithm 2 and Algorithm 3 with ℓ_∞–trust-region based regularizations as described in [6]. We used the GNU Linear Programming Kit (GLPK) [7] library routines to solve linear programming subproblems. All computations were carried out on a Linux workstation with dual 2.4 GHz Intel Xeon processors and 2 GB RAM. We considered a 100 scenario instance of the standard stochastic programming test problem `gbd`. Data for this instance is available from the website:

`http://www.cs.wisc.edu/~swright/stochastic/sampling`

Figures 1 and 2 show the mean-ASD efficient frontier obtained by Algorithm 2 and the mean-QDEV efficient frontier (corresponding to $\varepsilon_1 = 1$ and $\varepsilon_2 = 9$, i.e., $\alpha = 0.9$) obtained by Algorithm 3, respectively. In both cases, the parametric strategy was significantly more efficient than resolving the problem from scratch for different values of λ. For the mean-ASD problem, 5 unique solutions were identified for $\lambda \in [0, 1]$. For the mean-QDEV, 18 unique solutions were identified for $\lambda \in [0, 10]$. It is evident that the mean-QDEV model offers more flexibility in the mean-risk trade-off than the mean-ASD model. Computational results using Algorithm 2 on additional stochastic programming test problems are provided in [1].

Acknowledgements

This research has been supported by the National Science Foundation under CAREER Award number DMII-0133943. The author is grateful to two anonymous reviewers and Alexander Shapiro for constructive comments.

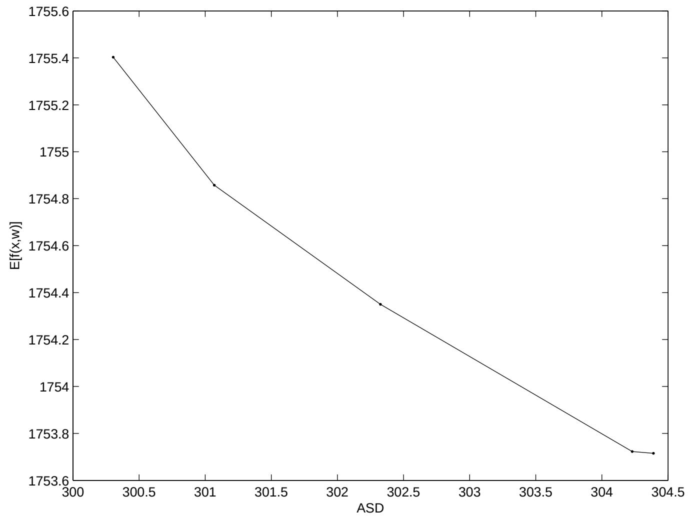

Fig. 1. Mean-ASD frontier for `gbd`

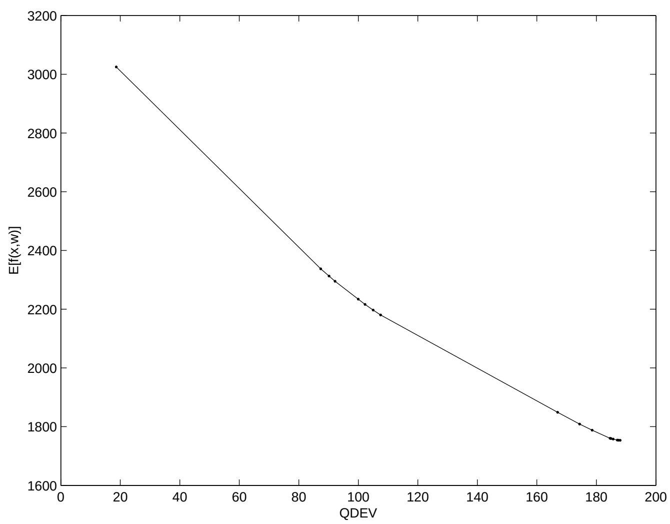

Fig. 2. Mean-QDEV frontier for `gbd` ($\alpha = 0.9$)

References

1. S. Ahmed. Mean-risk objectives in stochastic programming. Technical report, Georgia Institute of Technology, 2004. E-print available at www.optimization-online.org.
2. J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, New York, NY, 1997.
3. M.R. Garey and D.S. Johnson. *Computers and intractability: A guide to the theory of NP-completeness*. W. H. Freeman and Co., 1979.
4. J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms II*. Springer-Verlag, Berlin Heidelberg, 1996.
5. C. Lemaréchal, A. Nemirovski, and Yu. Nesterov. New variants of bundle methods. *Mathematical Programming*, 69:111–148, 1995.
6. J. Linderoth and S. Wright. Decomposition algorithms for stochastic programming on a computational grid. *Computational Optimization and Applications*, 24:207–250, 2003.
7. A. Makhorin. GNU Linear Progamming Kit, Reference Manual, Version 3.2.3. http://www.gnu.org/software/glpk/glpk.html, 2002.
8. H. M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments*. John Wiley & Sons, 1959.
9. W. Ogryczak and A. Ruszczyński. On consistency of stochastic dominance and mean-semideviation models. *Mathematical Programming*, 89:217–232, 2001.
10. W. Ogryczak and A. Ruszczyński. Dual stochastic dominance and related mean-risk models. *SIAM Journal on Optimization*, 13:60–78, 2002.
11. R. B. Porter. Semivariance and stochastic dominance. *American Economic Review*, 64:200–204, 1974.
12. R.T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *The Journal of Risk*, 2:21–41, 2000.
13. A. Ruszczyński and A. Shapiro, editors. *Stochastic Programming*, volume 10 of *Handbooks in Operations Research and Management Science*. North-Holland, 2003.
14. A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. Submitted for publication. E-print available at www.optimization-online.org, 2004.
15. A. Ruszczyński and R. Vanderbei. Frontiers of stochastically nondominated portfolios. *Econometrica*, 71:1287–1297, 2004.
16. R. Schultz. Stochastic programming with integer variables. *Mathematical Programming*, 97:285–309, 2003.
17. A. Shapiro and S. Ahmed. On a class of minimax stochastic programs. *SIAM Journal on Optimization*, 14:1237–1249, 2004.
18. S. Takriti and S. Ahmed. On robust optimization of two-stage systems. *Mathematical Programming*, 99:109–126, 2004.
19. S. Yitzhaki. Stochastic dominance, mean variance, and Gini's mean difference. *American Economic Review*, 72(1):178–185, 1982.