Shabbir Ahmed

Mean-risk objectives in stochastic programming

April 20, 2004

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 a parametric cutting plane algorithm to generate the entire mean-risk efficient frontier for a particular mean-risk objective.

Key words. Stochastic programming, mean-risk objectives, computational complexity, 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 all $\omega \in \Omega$, 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. [1, 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).

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

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

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 alternative 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 a parametric algorithm to generate the mean-risk efficient frontier for a particular mean-risk objective in the context of stochastic linear programs. Computational results involving standard stochastic programming test problems are reported.

2. Complexity of mean-variance stochastic programming

In this section we show that mean-variance extensions of even very simple stochastic linear programs lead to NP-hard optimization problems.

Consider the class of two-stage stochastic linear programs with *simple recourse* (cf. [1]), 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 $k = 1,\ldots,K$. The variance of $Q(x,\xi(\omega))$ is then given by*

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

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 $b_k \geq 0$ 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) 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 $Ax \leq b$? (8)

The binary integer feasibility problem (8) is known to be NP-complete [2]. 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 $j = 1,\ldots,n$,

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

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

Note that for all $j = 1,\ldots,n$, $\mathbb{E}[Q_j(x_j,\xi_j(\omega))]+\lambda\mathbb{V}[Q_j(x_j,\xi_j(\omega))] \ge (3+\frac{3}{4\lambda}+\frac{3\lambda}{2})$ 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 $p \ge 1$, a scalar $\lambda \ge 0$, 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 all $\omega \in \Omega$, 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 all $\omega \in \Omega$ 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 [11]. For $Y \in \mathcal{X}_p$ and a fixed target $T \in \mathbb{R}$, the p-th Semideviation from T 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 $S_{T,p}[Y]$ convex and non-decreasing in Y for all $p \geq 1$, the result follows from Proposition 1. □

Conditional value-at-risk [12]. For $Y \in \mathcal{X}_1$ and $\alpha \in (0,1)$ the α-Conditional value at risk is defined as

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

Proposition 4. *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. Since $\mathrm{CVaR}_\alpha[\cdot]$ convex and non-decreasing in Y [12], the result follows from Proposition 1. □

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

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

Proposition 5. *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 [10]. Given $\alpha \in (0,1)$ and $Y \in \mathcal{X}_1$, let

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

where $\kappa_\alpha[Y]$ is the α-th quantile of Y, i.e., $\Pr(Y < \kappa_\alpha[Y]) \leq \alpha \leq \Pr(Y \leq \kappa_\alpha[Y])$.

Proposition 6. *The mean-quantile deviation objective*

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

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

Proof. Ogryczak and Ruszczyński [10] have shown that $h_\alpha[Y]$ is convex and positively homogenous. Thus $g_{\lambda,h_\alpha}[Y]$ is convex and positively homogenous for all $\lambda \geq 0$. Moreover, For any $Y \leq 0$, we have $\kappa_\alpha[Y] \leq 0$. Thus $\kappa_\alpha[Y] - Y \leq -Y$, and $h_\alpha[Y] \leq \alpha\mathbb{E}[(\kappa_\alpha[Y] - Y)] \leq -\alpha\mathbb{E}[Y]$. Therefore, $g_{\lambda,h_\alpha}[Y] \leq (1-\lambda\alpha)\mathbb{E}[Y] \leq 0$, since $\lambda \leq 1/\alpha$. The result then follows from Proposition 2. □

Gini mean difference [16]. For $Y \in \mathcal{X}_1$ with distribution function F_Y, the Gini mean difference is defined as

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

Proposition 7. *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. □

Remark 1. Note that the non-decreasing, hence convexity preserving, property of g_{λ,δ_p}, g_{λ,h_α}, and $g_{\lambda,\Gamma}$ may not hold for $\lambda > 1$, $\lambda > 1/\alpha$, 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 $\lambda \in (0,1)$, $\lambda \in (0, 1/a)$ and $\lambda \in (0,1)$, respectively.

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 a particular class 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).

4.1. Deterministic equivalent formulation

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

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

For example, 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:

$$\begin{aligned} \min_{x,y(\omega)} \ & g_{\lambda,\mathbb{D}}[F_0(x,y(\omega),\omega)] \\ \text{s.t.} \ & x \in X \\ & \left.\begin{aligned} & F_i(x,y(\omega),\omega) \le 0 \quad i=1,\dots,m \\ & y(\omega) \in Y \end{aligned}\right\} \forall\, \omega \in \Omega. \end{aligned} \tag{12}$$

When Ω is finite, problem (12) is a large-scale deterministic optimization problem which may be solved by standard methods.

In particular, consider the mean-ASD objective function g_{λ,δ_1}. 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. [15]):

$$\begin{aligned} \min_{x,y(\omega),\nu(\omega)} \ & (1-\lambda)\int F_0(x,y(\omega),\omega)dP(\omega) + \lambda \int \nu(\omega)dP(\omega) \\ \text{s.t.} \ & x \in X \\ & \left.\begin{aligned} & F_i(x,y(\omega),\omega) \le 0 \quad i=1,\dots,m \\ & \nu(\omega) \ge F_0(x,y(\omega),\omega) \\ & \nu(\omega) \ge \int F_0(x,y(\xi),\xi)dP(\xi) \\ & y(\omega) \in Y \end{aligned}\right\} \forall\, \omega \in \Omega. \end{aligned} \tag{14}$$

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, the mean-ASD problem (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 [15]. 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.

4.2. Cutting plane methods

When $f(\cdot,\omega)$ is convex for all $\omega \in \Omega$ 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 [3,4]. 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., $s \in \partial\phi(x)$. Such a subgradient can be calculated as follows. Let $Y(\omega) = f(x,\omega)$ and $\pi(\omega) \in \mathbb{R}^n$

be a subgradient of $f(\cdot,\omega)$ at x, i.e., $\pi(\omega) \in \partial f(x,\omega)$, for all $\omega \in \Omega$. Note that $\gamma : \mathcal{X}_p^* \mapsto \mathbb{R}$ (here $\mathcal{X}_p^*$ is the dual space of $\mathcal{X}_p$, and in this case $\mathcal{X}_p^* = \mathcal{X}_p$) is a subgradient of $g_{\lambda,\mathbb{D}}$ at $Y \in \mathcal{X}_p$, i.e., $\gamma \in \partial g_{\lambda,\mathbb{D}}[Y]$ if

$$g_{\lambda,\mathbb{D}}[Y'] \geq g_{\lambda,\mathbb{D}}[Y] + \gamma[Y' - Y] \quad \forall\, Y' \in \mathcal{X}_p, \tag{15}$$

and such a subgradient always exists if $g_{\lambda,\mathbb{D}}$ is real-valued, convex, and continuous. Given a $\gamma \in \partial g_{\lambda,\mathbb{D}}[Y]$, it is easily verified that

$$s = (\gamma[\pi_1], \gamma[\pi_2], \ldots, \gamma[\pi_n])^T \in \partial\phi(x). \tag{16}$$

In fact, it can be shown that $\partial\phi(x)$ is given by the convex hull of all such s (cf. [14]). The above subgradient can now 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)/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 Y^i(ω) = f(x^i, ω) and π^i(ω) ∈ ∂f(x^i, ω).
  end for
  compute φ(x^i) = g_{λ,D}[Y^i] and s^i = (γ^i[π_1^i], ..., γ^i[π_n^i])^T, where γ^i ∈ ∂g_{λ,D}[Y^i].
  set UB = min{UB, φ(x^i)}.
end while
```

In case of the Mean-ASD objective, a subgradient $\gamma \in \partial g_{\lambda,\delta_1}[\widehat{Y}]$ can be calculated as follows.

Proposition 8. *Given $\widehat{Y} \in \mathcal{X}_1$ with probability distribution P, the function*

$$\gamma[Y] = \mathbb{E}Y + \lambda \int_{\widehat{Y} \geq \mathbb{E}\widehat{Y}} (Y - \mathbb{E}Y)\,dP$$

is a subgradient of $g_{\lambda,\delta_1}[Y]$ at $\widehat{Y}$.

Proof. Note that

$$\delta_1[Y] = \mathbb{E}[(Y - \mathbb{E}Y)_+] = \int_{Y \geq \mathbb{E}Y} (Y - \mathbb{E}Y)\,dP.$$

Let $\widehat{\Omega}_+ = \{\omega \in \Omega : \widehat{Y} \geq \mathbb{E}\widehat{Y}\}$ and $\Omega_+ = \{\omega \in \Omega : Y \geq \mathbb{E}Y\}$. Then

$$\begin{aligned}\textstyle\int_{\widehat{\Omega}_+} (Y - \mathbb{E}Y)\,dP &= \underbrace{\int_{\widehat{\Omega}_+ \setminus (\Omega_+ \cap \widehat{\Omega}_+)} (Y - \mathbb{E}Y)\,dP}_{\leq 0} + \underbrace{\int_{\Omega_+ \cap \widehat{\Omega}_+} (Y - \mathbb{E}Y)\,dP}_{\leq \int_{\Omega_+} (Y - \mathbb{E}Y)\,dP} \\ &\leq \textstyle\int_{\Omega_+} (Y - \mathbb{E}Y)\,dP.\end{aligned}$$

Thus

$$\begin{aligned} g_{\lambda,\delta_1}[\widehat{Y}] + \gamma[Y - \widehat{Y}] &= \mathbb{E}Y + \lambda \textstyle\int_{\widehat{Y} \geq \mathbb{E}\widehat{Y}} (Y - \mathbb{E}Y) dP \\ &\leq \mathbb{E}Y + \lambda \textstyle\int_{Y \geq \mathbb{E}Y} (Y - \mathbb{E}Y) dP \\ &= g_{\lambda,\delta_1}[Y], \end{aligned}$$

and the result follows from the definition (15) of a subgradient. □

From the above result and (16), a subgradient $s \in \partial\phi(x)$, where $\phi(x) = g_{\lambda,\delta_1}[f(x,\omega)]$, is given by

$$s = \mathbb{E}[\pi] + \lambda \int_{f(x,\omega) \geq \mathbb{E}[f(x,\omega)]} (\pi(\omega) - \mathbb{E}[\pi]) dP,$$

where $\pi(\omega) \in \partial f(x,\omega)$ for all $\omega \in \Omega$.

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 compute the function value and its subgradient (see Algorithm 1 for details). 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 objective), 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. A parametric cutting plane algorithm for mean-ASD 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, a more efficient parametric optimization scheme is possible. Note that in Algorithm 1, the tradeoff parameter λ is embedded within the cut coefficients, and as such, this algorithm is not suitable for an efficient parametric analysis of the mean-risk model with respect to λ. Fortunately, in case of mean-ASD, this issue can be resolved. Recall from (13), that the mean-ASD stochastic program is

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

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 supporting hyperplanes (cuts) within a cutting plane scheme. A subgradient for ν is given as follows.

Proposition 9. *Given $\widehat{Y} \in \mathcal{X}_1$ with probability distribution P, the function*

$$v[Y] = \int_{\omega:\widehat{Y}(\omega) > \mathbb{E}\widehat{Y}} Y dP + \mathbb{E}Y \int_{\omega:\widehat{Y}(\omega) \leq E\widehat{Y}} dP$$

is a subgradient of $\nu[Y] = \mathbb{E}[\max\{Y, \mathbb{E}Y\}]$.

Proof. Let $\widehat{\Omega}_+ = \{\omega \in \Omega : \widehat{Y} \geq \mathbb{E}\widehat{Y}\}$, $\Omega_+ = \{\omega \in \Omega : Y \geq \mathbb{E}Y\}$, $\widehat{\Omega}_+^c = \Omega \setminus \widehat{\Omega}_+$ and $\Omega_+^c = \Omega \setminus \Omega_+$. Note that, for any $Y \in \mathcal{X}_1$,

$$\begin{aligned}
&\nu[\widehat{Y}] - v[Y - \widehat{Y}] \\
&= v[Y] \\
&= \textstyle\int_{\widehat{\Omega}_+} Y dP + \int_{\widehat{\Omega}_+^c} \mathbb{E}Y dP \\
&= \textstyle\int_{\widehat{\Omega}_+ \cap \Omega_+} Y dP + \underbrace{\int_{\widehat{\Omega}_+ \cap \Omega_+^c} Y dP}_{\leq \int_{\widehat{\Omega}_+ \cap \Omega_+^c} \mathbb{E}Y dP} + \underbrace{\int_{\widehat{\Omega}_+^c \cap \Omega_+} \mathbb{E}Y dP}_{\leq \int_{\widehat{\Omega}_+^c \cap \Omega_+} Y dP} + \textstyle\int_{\widehat{\Omega}_+^c \cap \Omega_+^c} \mathbb{E}Y dP \\
&\leq \textstyle\int_{\Omega_+} Y dP + \int_{\Omega_+^c} \mathbb{E}Y dP \\
&= \nu[Y],
\end{aligned}$$

and the result follows from the definition (15) of a subgradient. □

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 chose 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.

We implemented an enhanced version of Alogrithm 2 and applied it to generate the mean-ASD frontier for five standard stochastic linear programming test problems. Our implementation extends the basic scheme of Algorithm 2 with an ℓ_∞–trust-region based regularization 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. The stochastic linear programming test problems in our experiments are derived from those used in [5]. We consider the problems `LandS`, `gbd`, `20term`, `storm` and `ssn`. Data for these instances are available from the website:

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

These problems involve extremely large number of scenarios (joint realizations of the uncertain problem parameters). Consequently, we consider instances with 500, 100, 50, 50, and 25 equi-probable sampled scenario for the problems `LandS`,

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** $\frac{UB-LB}{LB} \geq \epsilon$ **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 $Y^i(\omega) = f(x^i, \omega)$ and $\pi^i(\omega) \in \partial f(x^i, \omega)$.
6: **end for**
7: compute $\mu(x^i) = \mathbb{E}[Y^i]$ and $u^i = \mathbb{E}[\pi^i]$.
8: compute $\nu(x^i) = \mathbb{E}[\max\{Y^i, \mu(x^i)\}]$
and $v^i = \int_{\omega: Y^i(\omega) > \mu(x^i)} \pi^i(\omega) dP(\omega) + \int_{w: Y^i(\omega) \leq \mu(x^i)} u^i dP(\omega)$ (cf. Proposition 9).
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 remain optimal.
12: if $\lambda^* < 1$, set $\lambda = \min\{1, \lambda^* + \varepsilon\}$ (where $\varepsilon > 0$ is very small), $LB = -\infty$ and $UB = +\infty$ and return to step 2.

`gbd`, `20term`, `storm` and `ssn`, respectively. Figures 1-5 shows the mean-ASD efficient frontier obtained by the parametric cutting plane algorithm (Algorithm 2) for the five problems. In each case, the parametric strategy was significantly more efficient than resolving the problem for scratch for different values of λ.

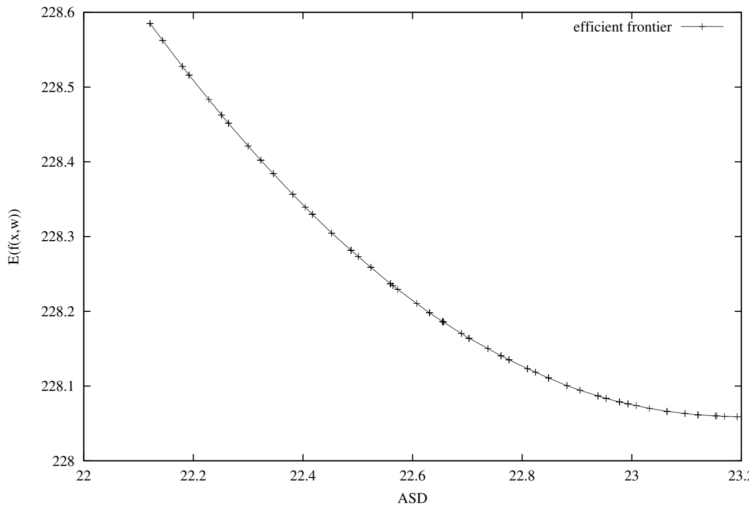

Fig. 1. Mean-ASD frontier for the `LandS` problem

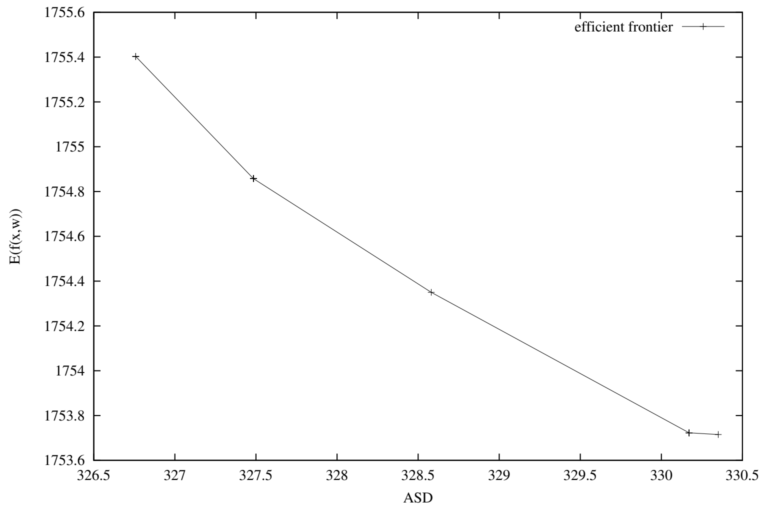

Fig. 2. Mean-ASD frontier for the `gbd` problem

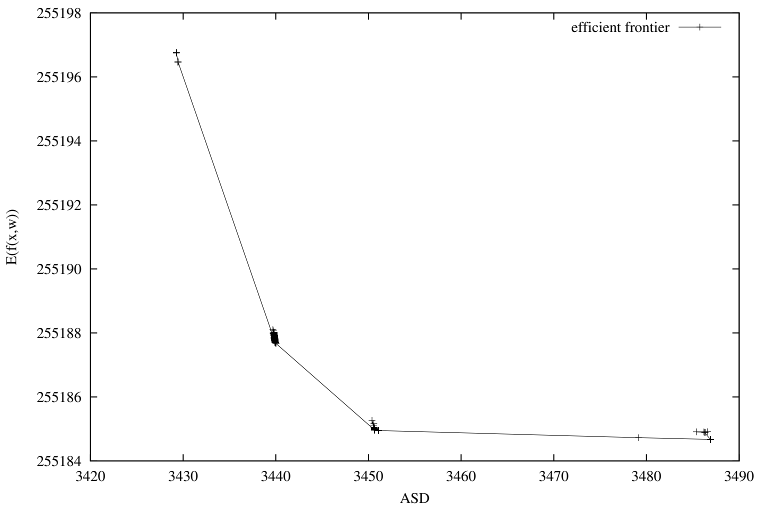

Fig. 3. Mean-ASD frontier for the `20term` problem

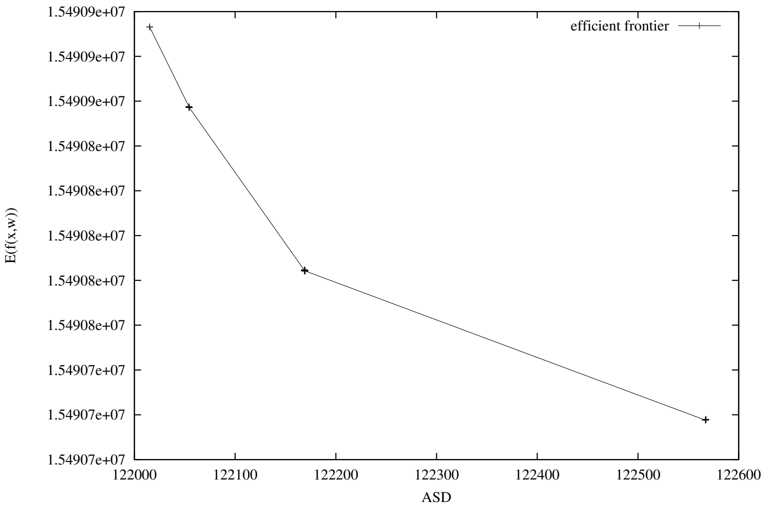

Fig. 4. Mean-ASD frontier for the `storm` problem

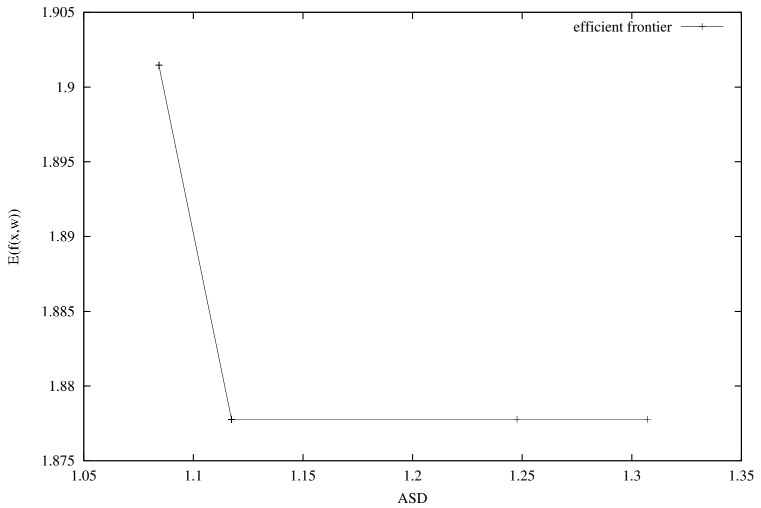

Fig. 5. Mean-ASD frontier for the `ssn` problem

Acknowledgements

This research has been supported by the National Science Foundation under CAREER Award number DMII-0133943.

References

1. J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, New York, NY, 1997.
2. M.R. Garey and D.S. Johnson. *Computers and intractability: A guide to the theory of NP-completeness*. W. H. Freeman and Co., 1979.
3. J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms II*. Springer-Verlag, Berlin Heidelberg, 1996.
4. C. Lemaréchal, A. Nemirovski, and Yu. Nesterov. New variants of bundle methods. *Mathematical Programming*, 69:111–148, 1995.
5. J. Linderoth, A. Shapiro, and S. Wright. The empirical behavior for sampling methods for stochastic programming. To appear in the *Annals of Operations Research*, 2004.
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. R. Schultz. Stochastic programming with integer variables. *Mathematical Programming*, 97:285–309, 2003.
16. S. Yitzhaki. Stochastic dominance, mean variance, and Gini's mean difference. *The American Economic Review*, 72(1):178–185, 1982.