Stochastic Programming with Multivariate Second Order Stochastic Dominance Constraints with Applications in Portfolio Optimization

Rudabeh Meskarian[1]
Department of Engineering Systems and Design,
Singapore University of Technology and Design,
Singapore
Email:rudabeh@sutd.edu.sg

Jörg Fliege
School of Mathematics
University of Southampton
Southampton, SO17 1BJ, UK
Email:J.Fliege@soton.ac.uk

Huifu Xu
School of Engineering and Mathematical Sciences
City University London
EC1V 0HB, London, UK
Email:Huifu.Xu.1@city.ac.uk

January 6, 2014

Abstract

In this paper we study optimization problems with multivariate stochastic dominance constraints where the underlying functions are not necessarily linear. These problems are important in multicriterion decision making, since each component of vectors can be interpreted as the uncertain outcome of a given criterion. We propose a penalization scheme for the multivariate second order stochastic dominance constraints. We solve the penalized problem by the level function methods, and a modified cutting plane method and compare them to the cutting surface method proposed in the literature. The proposed numerical schemes are applied to a generic budget allocation problem and a real world portfolio optimization problem.

AMS Classification: 90C15 and 90C90

Keywords: Multivariate stochastic dominance, second order dominance, Slater constraint qualification, level function method, penalty method, portfolio optimization

1 Introduction

Stochastic dominance is used to compare the distribution of two random variables [20], thus providing a way to measure risks. Over the past few years the discussion on stochastic programs with stochastic dominance constraints has garnered more and more attention. Dentcheva and Ruszczyński [3, 4] first introduced optimization problems with stochastic dominance constraints. This is an attractive approach for managing risks in an optimization setting. While pursuing

---

[1]Corresponding author.

1

expected profits, one avoids high risks by choosing options that are preferable to a random benchmark.

Much of the work on optimization with stochastic dominance has focused on the case where the underlying random quantities being compared are unidimensional [5, 6, 18, 21]. More recently, Dentcheva and Ruszczyński [7] proposed the concept of positive linear second order stochastic dominance which is a special case of multivariate stochastic dominance and obtained necessary conditions of optimality for non-convex problems. The notion of multivariate stochastic dominance refers to the stochastic ordering of random vectors. It can be used as a tool for multicriterion decision making, since each component of vectors can be interpreted as the uncertain outcome of a given criterion. Homem-de-Mello and Mehrota [10] expanded the definition of positive linear second order dominance to polyhedral second order dominance and called it P-dominance. They proposed a sample average cutting-surface algorithm for optimization problems with multidimensional polyhedral linear second order stochastic dominance constraints. Hu et al. [14] extended this work and presented a more general definition of stochastic dominance over random vectors as natural extension of the polyhedral linear stochastic dominance concept. More recently, Hu et al. [13] proposed a new concept of stochastically weighted dominance, in which they treat the vector of weights as a random vector to deal with large number of weights for bigger problems. They showed that such an approach is much less restrictive than the deterministic weighted approach. More recently, Armbruster and Luedtke [1] derived a linear formulation for multivariate second order stochastic dominance which can be solved with off the shelf linear programming solvers.

Inspired by the notion of multivariate stochastic dominance and our earlier work on unidimensional second order stochastic dominance constraints particularly dealing with nonlinear underlying functions, we study stochastic optimization problems with multivariate second order stochastic dominance (SSD) constraints. Sun et al. [25] proposed an exact penalization scheme for unidimensional second order stochastic dominance. In this paper we effectively extend the methods proposed in [25] to stochastic programs with multivariate second order stochastic dominance constraints. We propose an exact penalization scheme for such problems and solve the penalized problem by the level function method proposed by Lemarechal et al. [17] and extended by Xu [27] and a modified cutting plane method and compare them to the cutting surface method proposed by Homem-de-Mello et al. [10] and the linearized method proposed by Armbruster and Luedtke [1].

The main contribution of this paper can be summarized as follows:

- We develop a penalization scheme for stochastic programs with multivariate second order stochastic dominance constraints. We do so by exploiting Clarke's exact penalty function theorem [2, Proposition 2.4.3] and Robinson's error bound [23]. We reformulate the multivariate stochastic dominance constraints and demonstrate that the reformulated problem satisfies the Slater Constraint Qualification under some moderate conditions. Furthermore, an exact penalization scheme based on L_∞-norm is derived. Based on the exact penalization formulations, we apply a well known level function method in nonsmooth optimization as discussed in [27, 17] to the penalized problems. An obvious advantage of this approach is that we can effectively deal with excessive number of constraints, non-smoothness in the constraints and nonlinearity of the underlying functions.

- A modified cutting plane method is also proposed. This cutting plane method differs from those in the literature [24] in that it applies to the maximum of the constraint functions rather than each constraint function. Moreover, our modified cutting plane method uses the cutting plane representation proposed in [16], so it differ from the methods proposed in [10, 14]. The idea of applying the cutting-plane method to the maximum of the constraint

2

functions is similar to the idea in algorithm proposed by Fábián et al. [8]. However, their method is applied to linear models while our modified cutting plane method is also applicable to nonlinear case. Therefore we may regard our algorithm as an extension of theirs. Furthermore, the proposed numerical methods provides an alternative approach to the existing cutting surface method for multivariate stochastic dominance introduced by Homem-de-Mello and Mehrota [10] and the linearized method proposed by Armbruster and Luedtke [1].

- We examine the efficiency of the penalization scheme and the numerical methods by presenting an academic problem, a generic budget allocation problem, and a real world portfolio optimization problem. The budget allocation model is inspired by the homeland security application of Hu et al. [12] and the budget allocation example of Armbruster and Luedtke [1], in which a limited budget must be allocated to a set of possible projects, and the allocation must stochastically dominate a given benchmark. The proposed method proved to be more efficient in sense of CPU time when solving larger problems compared to the linearized method proposed in [1]. For the portfolio optimization problem, we use real world test data of three indices to set up backtest and out-of-sample test to inspect the performance of the generated portfolio and compare it to the benchmark portfolio and a portfolio strategy generated by Markowitz model. The results suggested that the portfolio strategy generated by the proposed model significantly outperforms the benchmark portfolio and the portfolio generated by Markowitz model.

The rest of this paper is organized as follows. In Section 2, we define the optimization problem and introduce the exact penalization schemes. In Section 3, we discuss the solution method and correspondingly the algorithms. In Section 4, we apply the proposed method to some portfolio optimization problems and report some numerical test results. Finally, we present conclusions in Section 5.

2 Problem formulation

Let's start with a note of notation that are used in the following sections. We write $x^T y$ for the scalar product of two vectors x and y, and $\| \cdot \|$ for the Euclidean norm, while $\| \cdot \|_\infty$ for the maximum norm of continuous functions defined over a set T. $d(x, \mathcal{D}) := \inf_{x' \in \mathcal{D}} \|x - x'\|$ and $d_\infty(x, \mathcal{D}) := \inf_{x' \in \mathcal{D}} \|x - x'\|$ denote the distance from a point x to a set $\mathcal{D}$ in Euclidean norm and $\mathcal{L}_\infty$-norm, respectively. For a real valued smooth function f, we use $\nabla f(x)$ to denote the gradient of f at x. The expected value operator is denoted by $\mathbb{E}$. The standard symbol $\mathcal{L}_1(\Omega, \mathcal{F}, P; \mathbb{R}^m)$ (shortly $\mathcal{L}_1^m$) denotes the space of all integrable mappings X from Ω to $\mathbb{R}^m$. If the values are taken in $\mathbb{R}$ the superscript m will be omitted.

The concept of stochastic ordering for scalar random variables has been introduced in statistics and further applied and developed in economics [9]. Let $g(x, \xi)$ be a concave function, with decision vector x and random variable ξ. It is said that $g(x, \xi)$ stochastically dominates a random variable $Y(\xi) \in \mathcal{L}_1$ in the first order, denoted by $g(x, \xi) \succeq_1 Y(\xi)$ if

$$F(g(x, \xi); \eta) \le F(Y(\xi); \eta), \ \forall \eta \in \mathbb{R}, \tag{2.1}$$

where $F(g(x, \xi); \eta)$ and $F(Y(\xi); \eta)$ are the cumulative distribution functions of $g(x, \xi)$ and $Y(\xi)$, respectively. Let $F_2(g(x, \xi); \cdot)$ be defined as

$$F_2(g(x, \xi); \eta) = \int_{-\infty}^{\eta} F(g(x, \xi); \alpha) d\alpha \text{ for } \eta \in \mathbb{R}.$$

3

Similarly, we say that $g(x,\xi)$ dominates in the second order a random variable $Y(\xi) \in \mathcal{L}_1$ if

$$F_2(g(x,\xi);\eta) \leq F_2(Y(\xi);\eta), \; \forall \eta \in \mathbb{R}. \tag{2.2}$$

We denote the relation (2.2) as $g(x,\xi) \succeq_{(2)} Y(\xi)$.

Definition 2.1 *A random vector $G(x,\xi) \in \mathcal{L}_1^m$ linearly dominates a random vector $Y(\xi) \in \mathcal{L}_1^m$ in positive linear second order, written as $G(x,\xi) \succeq_{(2)}^{Plin} Y(\xi)$, if*

$$\nu^T G(x,\xi) \succeq_{(2)}^{P} \nu^T Y(\xi), \; \forall \nu \in \mathbb{R}_+^m. \tag{2.3}$$

In the same manner one can define the first and higher order linear dominance relations: $G(x,\xi) \succeq_{(k)}^{Plin} Y(\xi)$, $k = 1, 2, \ldots$ provided that $(k-1)$-st moments of $G(x,\xi)$ and $Y(\xi)$ are finite.

It is clear that the set of scalarizing vectors ν in (2.3) can be truncated, by imposing the normalization constraint $\nu \in S$, where S is the simplex:

$$S = \left\{ \nu \in \mathbb{R}_+^m : \nu_1 + \nu_2 + \cdots + \nu_m = 1 \right\}. \tag{2.4}$$

This restriction does not change the relation ($\succeq_{(2)}^{Plin}$).

In this paper, we consider the following optimization problem with multivariate second order stochastic dominance (SSD) constraints:

$$\begin{aligned} \min_{x \in X} \quad & \mathbb{E}[f(x,\xi)] \\ \text{s.t} \quad & \nu^T G(x,\xi) \succeq_{(2)}^{Plin} \nu^T Y(\xi), \; \forall \nu \in S, \end{aligned} \tag{2.5}$$

where $f : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ is convex continuous function and $G : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ is concave continuous function, both in x and ξ, $x \in X$ is a decision vector with X being a nonempty convex subset of $\mathbb{R}^n$ and $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ is a random vector defined on probability space $(\Omega, \mathcal{F}, P)$ with support Ξ, $\mathbb{E}[\cdot]$ denotes the expected value w.r.t. the probability distribution of ξ. The random variable $Y(\xi)$ plays the role of a benchmark outcome. For example, one may consider $Y(\xi) = G(\bar{x}, \xi)$, where $\bar{x} \in X$ is some reasonable value of the decision vector, which is currently employed in the system. We shall investigate the case when $G(x,\xi)$ and $Y(\xi)$ are m-dimensional random vectors, rather than scalar variables.

Using the properties of second order dominance [22, 26] and the definition of positive linear dominance, we reformulate the multivariate stochastic dominance constraint in (2.5) as,

$$\mathbb{E}[(\nu^T \eta - \nu^T G(x,\xi))_+] \leq \mathbb{E}[(\nu^T \eta - \nu^T Y(\xi))_+], \; \forall (\eta, \nu) \in \mathbb{R}^m \times S,$$

where $(\nu^T \eta - \nu^T G(x,\xi))_+ = \max(\nu^T \eta - \nu^T G(x,\xi), 0)$.

As it was mentioned earlier, multivariate stochastic dominance can be used as a tool in multicriterion decision making, since each component of vectors can be interpreted as the uncertain outcome of a given criterion, so it would be beneficial to take η to be a vector instead of a scalar variable.

Consequently, we reformulate the optimization problem (2.5) as a stochastic semi-infinite programming problem:

$$\begin{aligned} \min_{x \in X} \quad & \mathbb{E}[f(x,\xi)] \\ \text{s.t.} \quad & H(x,\eta,\nu) := \mathbb{E}[(\nu^T \eta - \nu^T G(x,\xi))_+] - \mathbb{E}[(\nu^T \eta - \nu^T Y(\xi))_+] \leq 0, \; \forall (\eta, \nu) \in \mathbb{R}^m \times S. \end{aligned} \tag{2.6}$$

Our focus is on numerical methods for solving the stochastic semi-infinite programming problem (2.6). There are three issues to deal with: (a) the expectation of random functions in both the objective and constraints, (b) the infinite number of constraints, (c) the non-smoothness resulting from the max functions.

Homem-de-Mello et al. [10] introduced a more general notion of dominance which includes positive linear dominance as a particular case. They considered the case where the set S is assumed to be a polyhedron. By using the polyhedral properties they proposed a cutting-surface algorithm. They dealt with the constraints by considering a cut generation, and solved the problem by a branch-and-cut algorithm. Although the proposed cutting-surface method is effective, it is computationally demanding. In particular, even for the case of second-order stochastic dominance, which induces a convex feasible region, their algorithm requires global optimization of a nonconvex problem as a subproblem. Furthermore, Armbruster and Luedtke [1] proposed to use a different notion of multivariate stochastic dominance as a constraint in a stochastic optimization model. They derived an LP formulation for an SSD constraint which could be solved using off-the-shelf linear programming solvers.

In what follows, we propose an alternative approach for the stochastic programming problem with positive linear second order stochastic dominance constraint. Specifically, we propose an exact penalty function to move the infinite number of constraints to the objective and solve the penalized problem using the level function method and a modified cutting-plane method.

2.1 Exact penalization with L_∞-norm

In this section we develop an exact penalization scheme for solving problem (2.6). We do so through Robinson's error bound for convex systems [23] and Clarke's penalization theorem ([2, Proposition 2.4.3]). A crucial condition needed is Slater constraint qualification. Unfortunately the problem does not satisfy the condition and it may be satisfied through some reformulation. In this section, we go through these technical details, most of which are extended from similar results of single variate case in [25].

We focus on the case when ξ follows a discrete distribution, that is

$$
\begin{aligned}
\min_x \quad & \sum_{i=1}^{N} p_i f(x, \xi^i) \\
\text{s.t.} \quad & \sum_{i=1}^{N} p_i (\nu^T \eta - \nu^T G(x, \xi^i))_+ - \sum_{i=1}^{N} p_i (\nu^T \eta - \nu^T Y(\xi^i))_+ \leq 0, \ \forall (\eta, \nu) \in \mathbb{R}^m \times S, \\
& x \in X,
\end{aligned}
\tag{2.7}
$$

where the random variable ξ has a finite distribution, that is, $P(\xi = \xi^i) = p_i$, for $i = 1, \ldots, N$.

Problem (2.7) is said to satisfies the Slater Constraint Qualification (SCQ) if there exists $x_0 \in X$ and $\epsilon > 0$ such that

$$
\sum_{i=1}^{N} p_i (\nu^T \eta - \nu^T G(x_0, \xi))_+ - \sum_{i=1}^{N} (\nu^T \eta - \nu^T Y(\xi))_+ < -\epsilon, \ \forall (\eta, \nu) \in \mathbb{R}^m \times S. \tag{2.8}
$$

Unfortunately, this kind of constraint qualification is not satisfied. To see this, as discussed in [25], for a fixed $\nu \in S$ let

$$
\nu^T Y(\Xi) := \{\nu^T Y(\xi^i) : i = 1, \ldots, N\},
$$

5

and

$$C(\nu) := \min\{\nu^T Y(\xi^1), \ldots, \nu^T Y(\xi^N)\}. \tag{2.9}$$

For any $\nu^T \eta \leq C(\nu)$, it can be verified that $\mathbb{E}[(\nu^T \eta - \nu^T Y(\xi))_+] = 0$. For those η, the feasible constraint of problem (2.7) reduces to

$$\mathbb{E}[(\nu^T \eta - \nu^T G(x, \xi))_+] - \mathbb{E}[(\nu^T \eta - \nu^T Y(\xi))_+] = 0,$$

because the left hand term is non-negative. Consequently, there does not exist a feasible point $x_0 \in X$ such that (2.8) holds.

Dentcheva and Ruszczyński [3] tackled this issue in unidimensional case by considering a relaxed problem which restrict η to take value from a specified interval $[a, b]$. In other words, the feasible region of the original problem is enlarged. They showed that under some conditions, it is possible to choose a set C such that the relaxed problem satisfies SCQ. For example, if there exists a point $x_0 \in X$ such that

$$G(x_0, \xi) \succeq_{(1)} Y(\xi),,$$

and for every $\xi \in \Xi$, $G(x_0, \xi) > \eta^*$, where $\eta^* = \min\{Y(\xi^1), \ldots, Y(\xi^N)\}$, then x_0 is feasible point of the relaxed problem and

$$\int_{-\infty}^{\eta} F_1(G(x_0, \xi); \alpha) d\alpha < \int_{-\infty}^{\eta} F_1(Y(\xi); \alpha) d\alpha,$$

for all $\eta > \eta^*$. In such case, it is easy to verify that the SCQ holds for any $a > \eta^*$, while SCQ would fail if $[a, b]$ contains $Y(\Xi)$.

More recently, Homem-de-Mello and Mehrota [10] proposed an alternative approach to deal with the failure of constraint qualification by considering ϵ-feasible solutions:

$$\begin{aligned}
\min_x \quad & \mathbb{E}[f(x, \xi)], \\
\text{s.t.} \quad & \mathbb{E}[(\nu^T \eta - \nu^T G(x, \xi))_+] \leq \mathbb{E}[(\nu^T \eta - \nu^T T(\xi))_+] + \epsilon, \forall (\eta, \nu) \in \mathbb{R}^m \times S,
\end{aligned}$$

where ϵ is a small positive number. The relaxed problem (2.1) satisfies SCQ as long as the original problem is feasible. However, it must be shown that the feasible solution set of the relaxed problem approximates the feasible solution set of the original problem, which often in turn requires the original problem to satisfy certain regularity conditions such as lower semicontinuity of the feasible solution set of the relaxed problem.

In what follows, we propose an alternative way to tackle this problem by reformulating problem (2.6) using [3, Proposition 3.2] and [14] as follows:

$$\begin{aligned}
\min_x \quad & \sum_{i=1}^{N} p_i f(x, \xi^i) \\
\text{s.t.} \quad & \sum_{i=1}^{N} p_i (\nu^T \eta_j - \nu^T G(x, \xi^i))_+ \leq \gamma_j(\nu), \ \forall \nu \in S, \ j = 1, \cdots, N, \\
& x \in X,
\end{aligned} \tag{2.10}$$

where $\eta_j := Y(\xi^j)$ and $\gamma_j(\nu) := \sum_{i=1}^{N} p_i(\nu^T \eta_j - \nu^T Y(\xi^i))_+$. Note that, the reformulated problem does not satisfy the SCQ either.

Following [25], we use $\mathcal{N}$ to denote the power set of $\{1, \ldots, N\}$ excluding the empty set and for $j = 1, \ldots, N$, define

$$\psi_j(x, \nu) := \max_{\mathcal{J} \in \mathcal{N}} \sum_{i \in \mathcal{J}} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i)) - \gamma_j(\nu). \tag{2.11}$$

We would like to represent the constraints in (2.10) in terms of $\psi_j(x, \nu)$. The following lemma addresses this.

Lemma 2.1 *For $j = 1, \ldots, N$, let*

$$\varphi_j(x, \nu) := \max_{\mathcal{J} \in \mathcal{N}} \sum_{i \in \mathcal{J}} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i)).$$

Then

$$\sum_{i=1}^{N} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i))_+ = \max\{\varphi_j(x, \nu), 0\}, \tag{2.12}$$

for each fixed $\nu \in S$.

Proof. Let $\nu \in S$ be fixed, we consider two cases. Case 1. $\varphi_j(x, \nu) \leq 0$, and Case 2. $\varphi_j(x, \nu) > 0$.

Case 1. $\varphi_j(x, \nu) \leq 0$ implies that $\max\{\varphi_j(x, \nu), 0\} = 0$ and $\nu^T \eta_j - \nu^T G(x, \xi^i) \leq 0$, for $j \neq \{1, \cdots, N\}$. This implies that

$$\sum_{i=1}^{N} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i))_+ = 0,$$

hence (2.12) holds.

Case 2. Now consider case when $\varphi_j(x, \nu) > 0$. There exists a nonempty subset $\mathcal{J} \subseteq \{1, \ldots, N\}$ such that

$$\varphi_j(x, \nu) = \sum_{i \in \mathcal{J}} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i)) > 0.$$

It suffices to show that

$$\sum_{i \in \mathcal{J}} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i)) = \sum_{i=1}^{N} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i))_+,$$

or equivalently $\mathcal{J}$ consists of every index i with

$$\nu^T \eta_j - \nu^T G(x, \xi^i) > 0.$$

Note that, if $\mathcal{J}$ does not include such an index, then adding it to $\mathcal{J}$ would increase the quantity $\sum_{\mathcal{J} \in \mathcal{N}} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i))$. This contradicts the fact that $\varphi_j(x, \nu)$ is the maximum. Likewise, $\mathcal{J}$ does not consist of an index i with

$$\nu^T \eta_j - \nu^T G(x, \xi^i) < 0,$$

as removing the index will also increase the quantity $\sum_{\mathcal{J} \in \mathcal{N}} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i))$. This completes the proof. ∎

By Lemma 2.1, we can write (2.10) as

$$\begin{aligned}
\min_{x} \quad & \sum_{i=1}^{N} p_i f(x, \xi^i) \\
\text{s.t.} \quad & \psi_j(x, \nu) \leq 0, \ \forall \nu \in S, \ j = 1, \cdots, N, \\
& x \in X.
\end{aligned} \tag{2.13}$$

Compared to (2.10), a clear benefit of (2.13) is that it may satisfy SCQ under some circumstances.

Assumption 2.1 $f(x, \xi^i)$ and $G(x, \xi^i)$ are continuously differentiable w.r.t. x for $i = 1, \ldots, N$. Moreover, they are globally Lipschitz over X, that is, there exists $\kappa(\xi) < +\infty$ such that

$$\max \left(\left\| \nabla_x f(x, \xi^i) \right\|, \left\| \nabla_x G(x, \xi^i) \right\| \right) \leq \kappa(\xi^i), \ i = 1, \cdots, N.$$

We are now ready to state the main results.

Theorem 2.1 Let $G(x, \xi)$ and $Y(\xi)$ be defined as in problem (2.7) and $\psi_j(x, \nu)$ be defined as in (2.11). Let

$$\bar{\psi}_j(x) := \max_{\nu \in S} \psi_j(x, \nu) \text{ for } j = 1, \cdots, N.$$

Then

(i) $\nu^T G(x, \xi) \succeq_{(2)} \nu^T Y(\xi)$ for all $\nu \in S$ if and only if

$$\bar{\psi}_j(x) \leq 0, \text{ for } j = 1, \cdots, N; \tag{2.14}$$

(ii) problems (2.10) and (2.13) are equivalent;

(iii) if there exists a feasible point $\bar{x}$ such that $\nu^T G(\bar{x}, \xi) \succeq_{(1)} \nu^T Y(\xi)$ and $\nu^T G(\bar{x}, \xi) > C(\nu)$, where $C(\nu)$ is defined as in (2.9) for all $\xi \in \Xi$ and for all $\nu \in S$, then the system of inequalities (2.14) satisfies the SCQ.

Proof. The proof is similar to that of [25, Theorem 2.1] except that we have to deal with parameter ν.

Part (i). By [3, Proposition 3.2], $\nu^T G(x, \xi) \succeq_{(2)} \nu^T Y(\xi)$ for all $\nu \in S$ if and only if

$$\max_{\nu \in S} \sum_{i=1}^{N} p_i (\nu^T \eta_j - \nu^T G(x, \xi^i))_+ - \gamma_j(\nu) \leq 0, \text{ for } j = 1, \cdots, N, \tag{2.15}$$

or equivalently for $j = 1, \ldots, N$,

$$\max_{\nu \in S} \max \left\{ \sum_{i=1}^{N} p_i (\nu^T \eta_j - \nu^T G(x, \xi^i))_+ - \gamma_j(\nu), 0 \right\} = 0.$$

By (2.12),

$$\max \left\{ \sum_{i=1}^{N} p_i (\nu^T \eta_j - \nu^T G(x, \xi^i))_+ - \gamma_j(\nu), 0 \right\} = \max \left\{ \max \left\{ \varphi_j(x, \nu), 0 \right\} - \gamma_j(\nu), 0 \right\}.$$

Note that for any value $a \in \mathbb{R}$ and $r > 0$, it is easy to verify that

$$\max\{\max\{a, 0\} - r, 0\} = \max\{a - r, 0\}. \tag{2.16}$$

Using (2.16), we have that

$$\max\{\max\{\varphi_j(x, \nu), 0\} - \gamma_j(\nu), 0\} = \max\{\varphi_j(x, \nu) - \gamma_j(\nu), 0\} = \max\{\psi_j(x, \nu), 0\}.$$

The last equality is due to the definition of ψ_j. The discussion above demonstrates that (2.15) is equivalent to (2.14) and hence the conclusion.

Part (ii) follows straightforwardly from Part (i) in that the feasible sets of the two problems coincide, i.e.,

$$\left\{ x \in X : \max_{\nu \in S} \sum_{i=1}^{N} p_i (\nu^T \eta_j - \nu^T G(x, \xi^i))_+ - \gamma_j(\nu) \le 0 \right\} = \{ x \in X : \bar{\psi}_j(x) \le 0 \}.$$

Part (iii). Let $C(\nu)$ be defined as in (2.9) and $\hat{\gamma}(\nu) := \sum_{i=1}^{N} p_i (C(\nu) - \nu^T Y(\xi^i))_+$. Then $\hat{\gamma}(\nu) = \sum_{i=1}^{N} p_i (C(\nu) - \nu^T Y(\xi^i))_+ = 0$. Likewise, the assumption $\nu^T G(\bar{x}, \xi) > C(\nu)$ for $\xi \in \Xi$ implies that

$$\max_{\nu \in S} \max_{\mathcal{J} \in \mathcal{N}} \sum_{i \in \mathcal{J}}^{N} p_i (C(\nu) - \nu^T G(\bar{x}, \xi^i)) < 0.$$

This shows

$$\max_{\nu \in S} \left[\max_{\mathcal{J} \in \mathcal{N}} \sum_{i \in \mathcal{J}} p_i (C(\nu) - \nu^T G(\bar{x}, \xi^i)) - \hat{\gamma}(\nu) \right] < 0. \tag{2.17}$$

For each fixed $\nu \in S$, let $\nu^T \eta_1, \ldots, \nu^T \eta_N$, where $\eta_j := Y(\xi^j)$ denote the N elements in set $\nu^T Y(\Xi)$,

$$\nu^T \eta_1 \le \nu^T \eta_2 \le \cdots \le \nu^T \eta_N.$$

Then inequality (2.17) means that

$$\bar{\psi}_1(\bar{x}) := \max_{\nu \in S} \left[\max_{\mathcal{J} \in \mathcal{N}} \sum_{i \in \mathcal{J}} p_i (\nu^T \eta_1 - \nu^T G(\bar{x}, \xi^i)) - \sum_{i=1}^{N} p_i (\nu^T \eta_1 - \nu^T Y(\xi^i))_+ \right] < 0.$$

In what follows, we show that

$$\bar{\psi}_j(\bar{x}) < 0, \text{ for } j = 2, \cdots, N.$$

By definition, for $j = 2, \ldots, N$

$$
\begin{aligned}
\bar{\psi}_j(\bar{x}) \quad &= \quad \max_{\nu \in S} \left[\max_{\mathcal{J} \in \mathcal{N}} \sum_{i \in \mathcal{J}}^{N} p_i (\nu^T \eta_j - \nu^T G(\bar{x}, \xi^i)) - \sum_{i=1}^{N} p_i (\nu^T \eta_j - \nu^T Y(\xi^i))_+ \right] \\
&\le \quad \max_{\nu \in S} \left[\max \left\{ \max_{\mathcal{J} \in \mathcal{N}} \sum_{i \in \mathcal{J}}^{N} p_i (\nu^T \eta_j - \nu^T G(\bar{x}, \xi^i)), 0 \right\} - \sum_{i=1}^{N} p_i (\nu^T \eta_j - \nu^T Y(\xi^i))_+ \right] \\
&= \quad \max_{\nu \in S} \left[\max \left\{ \varphi_j(x, \nu), 0 \right\} - \gamma_j(\nu) \right] \\
&\underset{(2.12)}{=} \quad \max_{\nu \in S} \left[\sum_{i=1}^{N} p_i ((\nu^T \eta_j - \nu^T G(\bar{x}, \xi^i))_+ - (\nu^T \eta_j - \nu^T Y(\xi^i))_+) \right] \\
&= \quad \max_{\nu \in S} \left[\int_{-\infty}^{\nu^T \eta_j} (F_1(\nu^T G(\bar{x}, \xi), \alpha) - F_1(\nu^T Y(\xi), \alpha)) d\alpha \right]. \tag{2.18}
\end{aligned}
$$

The equality (2.18) is due to the equivalent representation of second order stochastic dominance [7].

Assume without loss of generality that $\nu^T \eta_1 < \nu^T \eta_2$ (otherwise $\bar{\psi}_2(\bar{x}) = \bar{\psi}_1(\bar{x}) < 0$). Let $\nu^T \eta_{min} \in (\nu^T \eta_1, \min\{\min_{\xi \in \Xi} G(\bar{x}, \xi^i), \nu^T \eta_2\})$ for a given ν. Note that by assumption $\nu^T \eta_1 <$

$\min\{\min_{\xi\in\Xi} G(\bar{x},\xi^i), \nu^T\eta_2\}$, $\nu^T\eta_{min}$ exists. Then

$$\max_{\nu\in S}\left[\int_{-\infty}^{\nu^T\eta_j}(F_1(\nu^T G(x,\xi),\alpha)-F_1(\nu^T Y(\xi),\alpha))d\alpha\right]$$

$$= \max_{\nu\in S}\left[\int_{-\infty}^{\nu^T\eta_{min}}(F_1(\nu^T G(\bar{x},\xi),\alpha)-F_1(\nu^T Y(\xi),\alpha))d\alpha\right.$$

$$\left.+\int_{\eta_{min}}^{\nu^T\eta_j}(F_1(\nu^T G(x,\xi),\alpha)-F_1(\nu^T Y(\xi),\alpha))d\alpha\right].$$

Note that

$$\max_{\nu\in S}\left[\int_{-\infty}^{\nu^T\eta_{min}}(F_1(\nu^T G(\bar{x},\xi),\alpha)-F_1(\nu^T Y(\xi),\alpha))d\alpha = 0 - p_1(\nu^T\eta_{min}-\nu^T\eta_1)\right] < 0,$$

where p_1 is the probability that $Y(\xi)$ takes value η_1. On the other hand, $\nu^T G(\bar{x},\xi) \succeq_{(1)} \nu^T Y(\xi)$ implies

$$\max_{\nu\in S}\left[\int_{\bar{\eta}}^{\nu^T\eta_j}(F_1(\nu^T G(\bar{x},\xi),\alpha)-F_1(\nu^T Y(\xi),\alpha))d\alpha\right] < 0.$$

This shows that

$$\max_{\nu\in S}\left[\int_{-\infty}^{\nu^T\eta_j}(F_1(\nu^T G(\bar{x},\xi),\alpha)-F_1(\nu^T Y(\xi),\alpha))d\alpha\right] < 0, \text{ for } j=2,\cdots,N. \tag{2.19}$$

The conclusion follows by combining (2.17)–(2.19). ∎

It might be helpful to discuss how strong the conditions in part (iii) of Theorem 2.1 are. Let us consider the case when ξ follows a finite distribution, that is, $\Xi = \{\xi^1,\ldots,\xi^N\}$. Suppose that there exist a point $\bar{x}\in X$ and $i_0 \in \{1,\cdots,N\}$ such that for $k=1,\cdots,m$

$$G_k(\bar{x},\xi^i) > Y_k(\xi^i), \tag{2.20}$$

which means at each scenario $G(\bar{x},\cdot)$ dominates the benchmark $Y(\cdot)$ in the multiobjective sense. Suppose further

$$G(\bar{x},\xi^i) \succ Y(\xi^{i_0}) \tag{2.21}$$

for $i=1,\cdots,N$, where $\succ$ is a natural order relation on the m-dimensional space, that is, there exists a scenario i_0 in which the benchmark value $Y(\xi^{i_0})$ is strictly dominated by the value of $G(\bar{x},\xi)$ in the multiobjective sense in any scenario $\xi \in \Xi$. Under condition (2.20), we have for any $\nu\in S$,

$$\nu^T G(\bar{x},\xi^i) > \nu^T Y(\xi^i)$$

and under additional condition (2.21)

$$\min_{i\in\{1,\ldots,N\}}\nu^T G(\bar{x},\xi^i) > \min_{i\in\{1,\ldots,N\}}\nu^T Y(\xi^i).$$

The assumptions made in the theorem above are not overly strong and depend mostly on the choice of benchmark Y.

10

We now move on to discuss penalty method for solving (2.13). One popular penalty scheme in optimization is based on the L_∞-norm. Here we consider such penalization scheme for (2.13) as follows:

$$\min_x \sum_{i=1}^N p_i f(x, \xi^i) + \rho \max_{j \in \{1, \cdots, N\}} (\max_{\nu \in S} \psi_j(x, \nu))_+, \tag{2.22}$$

and for problem (2.10)

$$\min_{x \in X} \sum_{i=1}^N p_i f(x, \xi^i) + \rho \max_{j \in \{1, \cdots, N\}} (\max_{\nu \in S} \sum_{i=1}^N p_i (\nu^T \eta_j - \nu^T G(x, \xi^i))_+ - \gamma_j(\nu))_+. \tag{2.23}$$

In what follows, we show that the two penalty schemes are equivalent, and estimate the penalty parameter. This will effectively justify the exact penalization function (2.23) for problem (2.10), although it does not satisfy the SCQ.

Theorem 2.2 *Assume that the problem (2.22) satisfies the SCQ and Assumption 2.1 holds; the feasible set of problem (2.13) is bounded. Then*

(i) problem (2.22) and (2.23) are equivalent;

(ii) there exist positive constants $\bar\delta$ and $\bar D$ such that when

$$\rho > \sum_{i=1}^N p_i \kappa(\xi^i) \bar\delta^{-1} \bar D, \tag{2.24}$$

the set of optimal solutions of (2.13) coincide with that of (2.22) and the set of optimal solutions of (2.10) coincides with that of (2.23).

Proof. Part (i). Through Lemma 2.1 and (2.16), the equivalence of the problem (2.22) and (2.23) can be verified as follows

$$\max_{j \in \{1, ..., N\}} (\max_{\nu \in S} \psi_j(x, \nu))_+ = \max_{j \in \{1, ..., N\}} \left[\max_{\nu \in S} \max_{\mathcal{J} \in \mathcal{N}} \sum_{i \in \mathcal{J}} p_i (\nu^T \eta_j - \nu^T G(x, \xi^i)) - \gamma_j(\nu) \right]_+$$

$$= \max_{j \in \{1, ..., N\}} \left[\max_{\nu \in S} \sum_{i=1}^N p_i (\nu^T \eta_j - \nu^T G(x, \xi^i))_+ - \gamma_j(\nu) \right]_+.$$

Part (ii). let $\mathcal{Q}$ denote the feasible set of problem (2.13) and define

$$\Psi(x, \nu) := (\psi_1(x), \cdots, \psi_N(x))^T.$$

Since $\mathcal{Q}$ is bounded, $\sum_{i=1}^N p_i f(x, \xi^i)$ is Lipschitz continuous with modulus $\sum_{i=1}^N p_i \kappa(\xi^i)$, problem (2.13) is convex and satisfies the SCQ. By virtue of [25, Lemma 3.1], there exists real number $\bar\delta > 0$ and $\bar D > 0$ such that

$$\bar\rho > \sum_{i=1}^N p_i \kappa(\xi^i) \bar\delta^{-1} \bar D,$$

the set of optimal solutions of problem (2.10) coincides with that of (2.23). Moreover, since problem (2.13) and (2.10) are equivalent, and problem (2.22) and (2.23) are also equivalent, the set of optimal solutions of problem (2.13) coincides with that of (2.22). ∎

In the rest of this paper, we apply the level function methods, and a modified cutting-plane method to solve the penalized optimization problem (2.22).

3   Solution methods

In this section we discuss numerical methods for solving optimization problem (2.22). Specifically, we apply the following methods: the level function method and a modified cutting plane method and compare them to the cutting surface algorithm in [10] and the linearized method proposed in [1].

3.1   Level function methods

The level function method is popular numerical approach for solving deterministic nonsmooth optimization problems. It has been proposed by Lemaréchal et al. [17] for solving nonsmooth convex optimization problems and extended by Xu [27] for solving quasiconvex optimization problems. Meskarian et al. [19] recently applied a level function method to sotchastic programming problems with scalar second order stochastic dominance constraints where the distribution of ξ is discrete. In this subsection, we apply the level function method in [27] to problems (2.22).

It is well known [27] that a subgradient of the convex function can be used to construct a level function. In what follows, we apply the level function method to the exactly penalized problems (2.22). Define $\vartheta(x, \rho)$ as follows:

$$\vartheta(x, \rho) := \sum_{i=1}^{N} p_i f(x_k, \xi^i) + \rho \max_{j \in \{1, \cdots, N\}} (\max_{\nu \in S} \psi_j(x, \nu))_+.$$

Let $\zeta_k \in \partial_x \vartheta(x_k, \rho)$, then

$$\sigma_{x_k}(x) = \zeta_k^T (x - x_k) / \|\zeta_k\|,$$

is a level function of $\vartheta(x, \rho)$ at x_k. Since the projected level function (PLF) Algorithm is a classical algorithm we are not going to present the algorithm, however we refer the reader to [27] for the outline of the algorithm steps.

Theorem 3.1 *Let $\{x_k\}$ be generated by (PLF) Algorithm. Assume that $f(x, \xi)$ and components of $G(x, \xi)$ are Lipschitz continuous functions with modulus $L_f(\xi)$ and $L_G(\xi)$ respectively, where $\mathbb{E}[L_f(\xi)] < +\infty$, $\mathbb{E}[L_G(\xi)] < +\infty$ and that the sequence of level functions $\{\sigma_{x_k}(x)\}$ is uniformly Lipschitz with constant M. Then*

$$\Delta(k) \leq \epsilon, \text{ for } k > M^2 \Upsilon^2 \epsilon^{-2} \lambda^{-2} (1 - \lambda^2)^{-1},$$

where Υ represents the diameter of the solution set X, ϵ and λ are given in (PLF) Algorithm.

Proof.   It is easy to observe that the Lipschitz continuity of $G(x, \xi)$ w.r.t. x with modulus $L_G(\xi)$ implies the Lipschitz continuity of $\psi_j(x, \nu)$ with the same Lipschitz modulus $\mathbb{E}[L_G(\xi)]$. Along with the Lipschitzness of $f(x, \xi)$, this shows $\vartheta(x, \rho)$ is Lipschitz continuous with modulus $\mathbb{E}[L_f(\xi)] + \rho \mathbb{E}[L_G(\xi)]$. On the other hand, since $\vartheta(x, \rho)$ is convex, the function $\sigma_{x_k}(x)$ constructed at each iterate is a level function with modulus 1. The rest follows from Xu [27, Theorem 3.3]. ∎

In (PLF) Algorithm, penalty parameter in $\vartheta(x, \rho)$ is fixed. In some cases, it might be difficult to estimate a good penalty parameter. One way to tackle this issue is to start with estimate of penalty parameter and solve the resulting penalized problem with the above algorithms. The feasibility of the obtained solution is checked: if it is feasible the optimal solution is obtained, otherwise, the penalty parameter is increased the process is repeated. This kind of procedure in known as *Simple Penalty Function Method* in the literature of optimization, see for instance

An alternative way to deal with the issue of penalty parameters is to solve the following problem

$$\min_{x \in X} \max_{j \in \{1,\ldots,N\}} (\max_{\nu \in S}(\sum_{i=1}^{N} p_i(\nu^T \eta_j - \nu^T G(x, \xi^i))_+ - \gamma_j(\nu))). \qquad (3.25)$$

This can be achieved by applying (PLF) Algorithm directly. The optimal value of (3.25) effectively gives an upper bound for parameter $\bar{\delta}$ (see Theorems 2.2. Note that these parameters are dependent of the Slater condition of (2.13).

3.2 Modified cutting plane algorithm

Rodulf and Ruszczyński [24] and Fábián et al. [8] proposed cutting plane methods to solve stochastic program with second order stochastic dominance constraints when the underlying random variable has finite distribution. This method is an extension of the cutting-plane method developed by Haneveld and Vlerk [16] for integrated chance constraints (ICC). In what follows, we consider a modification of the procedure where a cut is constructed.

Reformulate the optimization problem (2.13) as:

$$
\begin{aligned}
\min_{x,z} \quad & z \\
\text{s.t.} \quad & \psi(x, \nu) := \max_{j \in \{1,\ldots,N\}} \bar{\psi}_j(x) \leq 0, \\
& \sum_{i=1}^{N} p_i f(x, \xi^i) - z \leq 0, \\
& x \in X, \; z \in Z,
\end{aligned}
\qquad (3.26)
$$

where $\bar{\psi}_j(x) := \max_{\nu \in S} \psi_j(x, \nu)$, Z is a closed convex compact subset of $\mathbb{R}$ such that

$$\left\{ \sum_{i=1}^{N} p_i f(x, \xi^i) : x \in X \right\} \subset Z.$$

Note that, the existence of set Z is due to the fact that $f(x, \xi^i)$, $i = 1, \ldots, N$, is a continuous function and X is a compact set. Also the components of $G(x, \xi)$ are concave and $f(x, \xi)$ is convex w.r.t. x, which implies that $\psi(x, \nu)$ is convex w.r.t. x and $\sum_{i=1}^{N} p_i f(x, \xi^i) - z$ is convex w.r.t. (x, z). We apply the classical cutting-plane method [15] to both $\psi(x, \nu)$ and $\sum_{i=1}^{N} p_i f(x, \xi^i) - z$. Specifically, we propose the following algorithm.

Algorithm 3.1 *(Modified cutting plane algorithm)*
Define the optimization problem at iteration t as

$$
\begin{aligned}
\min_{x,z} \quad & z \\
\text{s.t.} \quad & x \in X, \; z \in Z, \\
& (x, z) \in P_t := \left\{ (x, z) \in X \times Z : a_l^T x \leq b_l, \; d_l^T x + e_l z \leq k_l, \; l = 1, \ldots, t \right\}.
\end{aligned}
\qquad (3.27)
$$

Set $t := 0$, $P_0 := X \times Z$. For each t, carry out the following.

Step 1. Solve the optimization problem (3.27), finding the optimal solution (x_t, z_t). If the problem (3.27) is infeasible, stop. Otherwise go to Step 2.

Step 2. Find the solution to $\max_{j \in \{1,\dots,N\}} \bar{\psi}_j(x)$ *and find optimal solution* (η^*, ν^*)*, and set* $\gamma(\nu_t^*) := \sum_{i=1}^{N} p_i(\nu_t^{*T}\eta_t^* - \nu_t^{*T}Y(\xi^i))_+$*. If*

$$\sum_{i=1}^{N} p_i(\nu^{*T}\eta^* - \nu^{*T}G(x,\xi^i))_+ \leq \gamma(\nu_t^*),$$

and

$$\sum_{i=1}^{N} p_i f(x_t, \xi^i) - z_t \leq 0,$$

stop: (x_t, z_t) *is an optimal solution. Otherwise go to Step 3.*

Step 3. Construct the set

$$\mathcal{J}_t := \left\{ i : (\nu^{*T}\eta^* - \nu^{*T}G(x,\xi^i)) > 0 \right\},$$

and the feasibility cuts $a_{t+1}^T x \leq b_{t+1}$*, and* $d_{t+1}^T x + e_{t+1}z \leq k_{t+1}$*, where*

$$a_{t+1} = -\sum_{i \in \mathcal{J}_t} p_i \nabla_x \nu^{*T} G(x_t, \xi^i),$$

$$b_{t+1} = \sum_{i \in \mathcal{J}_t} p_i(-\nabla_x \nu^{*T} G(x_t, \xi^i)^T x_t + \nu^{*T} G(x_t, \xi^i) - \eta^*) + \gamma(\nu^*),$$

$$d_{t+1} = -\nabla_x f(x, \xi), \ e_{t+1} = -1, \ k_{t+1} = -\nabla_x f(x_t, \xi)^T x_t + f(x_t, \xi).$$

and set

$$P_{t+1} = P_t \cap \left\{ (x, z) \in X \times Z : a_{t+1}^T x \leq b_{t+1}, \ d_{t+1}^T x + e_{t+1}z \leq k_{t+1} \right\}.$$

Proceed with iteration $t + 1$*.*

Remark 1 *We make a few comments about Algorithm 3.1.*

(i) *Algorithm 3.1 differs from the cutting-plane method discussed in [24, 8] in the way how feasible cuts are constructed. In the former, N constraints/cuts are added at each iteration, these cuts are not necessarily the extreme support of $\psi(x, \nu)$ at x_t. In Algorithm 3.1, we exclude all those non-support constraints, instead we include a cut at the extreme support (to $\psi(x, \nu)$ at x_t) which we believe is the most effective and a single linear cut is adequate to ensure the convergence. All other non-support constraints/cuts may potentially reduce numerical efficiency. This approach is similar to the algorithm proposed by Fábián et al. [8]. Note that, Fábián's algorithm is applied to linear models while Algorithm 3.1 is applicable to the nonlinear case. Therefore, we may regard the latter as an extension of the former.*

(ii) *In Step 2 of the above algorithm, we solve the following DC-program*

$$\max_{(\eta,\nu) \in [a,b] \times S} \psi(\eta, \nu) := \sum_{i=1}^{N} p_i[(\nu^T \eta - \nu^T G(x_t, \xi^i))_+ - (\nu^T \eta - \nu^T Y(\xi^i))_+],$$

where $[a, b] \supseteq \{\eta_1, \dots, \eta_N\}$ and S is defined as in (2.4).

A function is called DC if it minimizes difference of two convex functions over a closed convex set. There has been extensive work done on development of solution methods for such problems, we refer interested reader to [11] and references therein.

(iii) When f is linear w.r.t. x, there is no need to introduce additional variable z because the objective is linear.

We now present the convergence results.

Theorem 3.2 *Let $\{(x_t, z_t)\}$ be a sequence generated by the Algorithm 3.1. Let*

$$P := \{(x, z) \in X \times Z : \psi(x, \nu) \leq 0, \mathbb{E}[f(x, \xi)] - z \leq 0\} \subset X \times Z,$$

where $\psi(x, \nu)$ is defined in problem (3.26). Assume: (a) $f(x, \xi)$ and each of the component $g_i(x, \xi)$ of $G(x, \xi)$ are continuously differentiable and concave w.r.t. x for almost every ξ, (b) $X \times Z \in \mathbb{R}^n$ is a compact set, (c) there exists a positive constant L such that the Lipschitz modulus of $\mathbb{E}[f(x, \xi)]$ and $\psi(x, \nu)$ are bounded by L on $X \times Z$, (d) the set P is nonempty. Then, $\{(x_t, z_t)\}$ contains a subsequence which converges to a point $(x^*, z^*) \in P$, where (x^*, z^*) is the optimal solution.

The proof of Theorem 3.2 is included in the Appendix.

In next section, we investigate the efficiency of the above algorithms and compare them to the cutting surface algorithm proposed in [10] and the linearized method proposed in [1].

4   Numerical Tests

We have carried out an academic test, a budget allocation example as well as a real world portfolio optimization problem on the proposed model and algorithms by using MATLAB 7.10 and IBM ILOG CPLEX 12.4 installed on a HP Notebook PC with Windows 7 operating system, and Intel Core i7 processor.

We consider primarily an academic test problem introduced in [10, Section 2.2] to examine the penalization approach and efficiency of our proposed methods. Additionally, for comparison purposes we consider a budget allocation problem as discussed in [1]. Finally, we consider a portfolio optimization problem with real world test data to further investigate the efficiency of the proposed stochastic programming model with multivariate SSD constraint and compare it to the return generated by a Markowitz model and corresponding indices. Furthermore, to estimate the penalty parameter we have solved the optimization problem (3.24) using (PLF) Algorithm as discussed on page 12. Another approach is to integrate the *Simple Penalty Function Method* in (PLF) Algorithm, to find a suitable penalty parameter. We solved the reformulated problem with (PLF) Algorithms, 3.1, and the cutting-surface algorithm in [10]. For (PLF) Algorithm we use $\epsilon = 0.0001$ and $\lambda = 0.5$. In the rest of this section we report the corresponding results.

4.1   An academic example

Example 4.1 *Homem-de-Mello et at [10] considered the following model using stochastic dominance:*

$$
\begin{aligned}
\max_x \quad & 3x_1 + 2x_2, \\
\text{s.t.} \quad & \\
& -\begin{bmatrix} \xi^1 & 2 \\ 2 & \xi^2 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \succeq_{(2)} -\begin{bmatrix} \xi^3 \\ 160 \\ \xi^4 \end{bmatrix},
\end{aligned}
\tag{4.28}
$$

where, ξ^i, $i = 1, \ldots, 4$ denotes a random variable. Let $\xi^1 := 4 \pm \alpha$, $\xi^2 := 2 \pm \alpha$, $\xi^3 := 200 \pm 10\beta$, and $\xi^4 := 40 \pm 5\beta$ where α and β are equal to 1. We write $(a \pm b)$ to indicate that the actual value is random with two equally probable outcomes $a + b$ and $a - b$. Consequently, there are 16 scenarios to consider.

To solve the optimization problem (4.28), Homem-de-Mello et al. [10] linearized the program and eliminated the redundant constraints. They obtained the optimal solution of problem (4.28) (with $\alpha = \beta = 1$) to be $x = (28.18, 34.55)$ and the corresponding objective value to be $f = 153.44$. Here we reformulate the optimization problem (4.28) as discussed in Section 2.1 and solved the reformulated problem by the proposed algorithms. We set the penalty parameter $\rho = 50$ and double it at each iteration. The optimal solution obtained by the proposed algorithms after 4 iteration is $x = (27.99, 34.66)$ and the corresponding objective value is $f = 153.29$.

4.2 A budget allocation model

The purpose of this example is to compare the efficiency of the level function method based on the exact penalization scheme with the linearized method proposed by Armbruster and Luedtke [1].

In what follows we present the budget allocation problem and study the behavior of the proposed model and methods to solve a simple budget allocation problem. This example is inspired by the budget allocation problem of Armbruster and Luedtke [1] and the example in [12]. Here we restate the problem:

Example 4.2 *Given a fixed budget, the problem is to determine what fraction of the budget to allocate to a set of candidate projects, $t \in \mathcal{T}$ with $|\mathcal{T}| = T$. The quality of a budget allocation is characterized by d distinct objectives, for which larger values are preferred. Each project $t \in \mathcal{T}$ is characterized by a d-dimensional random vector of reward rates R_t for these objectives. Thus, given a feasible budget allocation $x \in X := \{x \in \mathbb{R}_+^T : x \cdot 1 = 1\}$, the values of the d objectives are $\sum_{t \in \mathcal{T}} R_t x_t$. We assume that we are given a d-dimensional random vector Y that indicates a minimal acceptable joint performance of these objectives, and we require the performance of the chosen budget allocation to stochastically dominate Y. Subject to this condition, the goal is to maximize a weighted combination of the expected values of the measures:*

$$
\begin{aligned}
\max_{x \in X} \quad & \sum_{t \in \mathcal{T}} w^T \mathbb{E}[R_t] x_t \\
\text{s.t.} \quad & \sum_{t \in \mathcal{T}} R_t x_t \succeq_{(2)}^{Plin} Y,
\end{aligned}
\tag{4.29}
$$

where $w \in \mathbb{R}_+^d$ is a given weight vector.

For the test instances, we assumed that the reward rate $R := [R_1, R_2, \ldots, R_T]$ are one of N equally likely scenarios $\{R^j : j \in \mathcal{N}\}$ sampled from a joint normal distribution with mean μ and covariance matrix Σ. The components of μ are chosen randomly from $U[10, 20]$ and the covariance matrix Σ was calculated as follows. The coefficient of variations were chosen from $U[0.2, 1.1]$. The correlation of any two distinct elements (t, k) and (t', k') were chosen from $U[-0.2, 0.4]$ if they share a project $(t = t')$ and from $U[-0.1, 0.1]$ if they share an objective $(k = k')$ and were 0 otherwise. The benchmark random vector Y was determined from an allocation in which all projects are allocated an equal fraction of the budget, but to avoid being overly conservative, was then reduced by a fixed fraction δ of its mean. Specifically, a given realizations $R_t^j \in \mathbb{R}^d$, for each scenario j and project t, realization j of Y has a probability

$q_y(j) = 1/N$ and is given by $Y^j = B^j - \delta(\frac{1}{N}\sum_{k=1}^{N} B^k)$ where $B^j = \frac{1}{T}\sum_{t\in\mathcal{T}} R_t^j$. In this example, we set the $\delta = 0.1$ and weight all objectives equally in the objective, $w = (1, \ldots, 1)$.

(d,T)	Algorithms	$N = 100$	$N = 300$	$N = 500$
(3,50)	PLF Algorithm	2.36	13.05	46.86
	Algorithm 3.1	24.13	219.67	739.27
(3,100)	PLF Algorithm	2.69	14.78	53.48
	Algorithm 3.1	104.80	-	-
(5,50)	PLF Algorithm	5.38	14.04	90.18
	Algorithm 3.1	24.10	236.28	527.14
(5,100)	PLF Algorithm	10.20	26.69	98.47
	Algorithm 3.1	117.94	-	-

Table 4.1: Average solution times in seconds of five instances solved by projected level function algorithm and the modified cutting plane algorithm. The "'-"' indicate that the algorithms could not solve the problem within 30 minutes limit.

Table 4.1 shows the computation times to solve these instances using the exact penalization scheme and solved by the projected level function algorithm and the modified cutting plane algorithm. For these experiments, we varied the number of objectives $d \in \{3, 5\}$, the number of projects $T \in \{50, 100\}$, and the number of scenarios $N = M \in \{100, 300, 500\}$. For each combination of these parameters we display the average computation time in seconds over five instances at that size.

(d,T)	$N = 100$	$N = 300$	$N = 500$
(3,50)	0.3	12.3	86.2
(3,100)	0.3	8.9	61.6
(5,50)	0.6	37.8	181.8
(5,100)	0.7	23.0	105.6

Table 4.2: Average solution times in seconds of five instances solved using linear SSD model [1].

These results indicate that with the exact penalization scheme and PLF Algorithm it is possible to solve instances with a relatively large number of scenarios with lower computation time compared to the linear SSD formulation model's results shown in Table 4.2. Although, the opposite is true for lower number of scenario, but one advantage of the proposed exact penalized model and the solution methods is that they can deal with both linear and nonlinear underlying functions. Furthermore, Algorithm 3.1 proved to be less efficient. This is because as the sample size increases, the construction of set $\mathcal{J}_t$ in Step 3 of the algorithm takes longer time.

4.3   Portfolio Performance

Suppose that we have a fixed capital to be invested in n assets. Let R_i, $i = 1, \ldots, n$, denote the return of asset i. In practice, the return is often uncertain and we use a random variable ξ to describe the uncertainty. Specifically, we write R_i as $R_i(\xi)$ and in doing so we are assuming that all n assets have an identical random factor depending on ξ.

To simplify the discussion, we normalize the capital to 1 and use x_i, $i = 1, \ldots, n$, to denote the fraction of capital to be invested in asset i. The portfolio return can then be formulated as:

$$f(x, \xi) := R_1(\xi)x_1 + R_2(\xi)x_2 + \cdots + R_n(\xi)x_n. \tag{4.30}$$

We use the optimization problem (2.5) to optimize our investment strategy. To ease the

presentation, we repeat the model:

$$\begin{aligned}
\min_{x \in X} \quad & -\mathbb{E}[f(x,\xi)] \\
\text{s.t} \quad & \nu^T g(x,\xi) \succeq_{(2)}^{Plin} \nu^T Y(\xi), \ \forall \nu \in S,
\end{aligned} \tag{4.31}$$

where f is defined by (4.30). We need to specify $g(x,\xi)$ and X. The random variable $Y(\xi)$ plays the role of a benchmark outcome. For example, one may consider $Y(\xi) = g(\bar{x}, \xi)$, where $\bar{x} \in X$ is some reasonable value of the decision vector, which is currently employed in the system. Note that $g(x,\xi)$ and $Y(\xi)$ are m-dimensional random vectors, rather than scalar variables. Additionally, we use set of linear constraints to define the set S, see (2.4).

To further examine the efficiency of the multivariate SSD model, we calculate the Conditional Value at Risk (CVaR) for random variable $f(x^*, \xi)$ where x^* is an approximate optimal solution obtained from solving (2.23). By definition for a specified probability level α, the Value at Risk (VaR) of a portfolio is the lowest amount C such that, with probability α, the profit does not fall below C. The CVaR$_\alpha$ is the conditional expectation of profit below C. In our context,

$$\text{CVaR}_\alpha(f(x^*,\xi)) = \sup_{C} \left\{ C - \frac{1}{\alpha}\mathbb{E}[(C - f(x^*,\xi))_+] \right\}, \tag{4.32}$$

where $\alpha \in (0,1)$ is a pre-specified constant. Three values of α are commonly considered: 0.90, 0.95, 0.99. However, in our analysis we focus on the case of $\alpha = 0.95$.

Let us now estimate the penalty parameter ρ through Theorem 2.2. Referring back to Lemma 2.2, we need to calculate κ, δ, and D. Let $x_0 \in X$ be the weights for an equally weighted portfolio, the $\sum_{i=1}^N p_i \kappa(\xi^i) = 0.0084$. The δ can be calculated as follows:

$$\delta := -\max_{j \in 1,\ldots,N} \left\{ \max_{\nu \in S} \left(\max_{\mathcal{J} \in \mathcal{N}} \sum_{i \in \mathcal{J}} p_i(\nu^T \eta_j - \nu^T G(x_0, \xi^i)) - \sum_{i=1}^N p_i(\nu^T \eta_j - \nu^T Y(\xi^i))_+ \right) \right\},$$

and we found $\delta \approx 4.865\text{E-}005$. We choose $D = 1$ and estimate the penalty parameter as follows:

$$\rho \geq \sum_{i=1}^N p_i \kappa(\xi^i)\delta^{-1}D = 192 \tag{4.33}$$

Note that the above calculation is accurate with 3 significant figures.

Example 4.3 *We consider m history of percentage returns, for three different group of n assets. Each of these groups could belong to a different Index. Our aim is to find an optimal investment strategy for a fixed capital in the n assets which maximized the expected profit subject to certain risk averse measures. Particularly we consider the following model:*

$$\begin{aligned}
\min_{x \in X} \quad & -\mathbb{E}[f(x,\xi)] \\
\text{s.t} \quad & \nu^T g(x,\xi) \succeq_{(2)} \nu^T Y(\xi),
\end{aligned}$$

where $g(x,\xi) = [g_1(x,\xi) \ g_2(x,\xi) \ g_3(x,\xi)]$ and $Y(\xi) = [Y_1(\xi) \ Y_2(\xi) \ Y_3(\xi)]$. We apply the exact penalization as explained in Section 2 and set the initial penalty parameter $\rho = 200$. We set the upper bound and lower bound for the capital invested equal to 0.2 and 0, respectively.

We collected 300 daily historical returns of 53 FTSE100, 53 Nasdaq100 and 30 Dow Jones assets prior to March 2011. We use the first 100 observations to construct the portfolio strategy. We solve the optimization problem using level function algorithms, modified cutting-plane

Algorithm	Iter.	Time	No.Assets	Return	CVaR
PLF	9	0.0174	6	0.034	0.015
3.1	4	0.0166	6	0.034	0.014
Cutting-surface [10]	6	0.653	6	0.034	0.015

Table 4.3: Time is in minutes. No.Assets refers to the number of assets in the optimal portfolio. The expected return of the benchmark portfolio $Y = [0.0051\ 0.0085\ 0.0069]$.

method, and the cutting surface method. Table 4.3 shows the result of this example. In this example each component of the vector $g(x, \xi)$ corresponds to the sum of return of the assets belonging to each of the three indices computed as described in (4.30).

As it can be seen all four algorithms result in very similar portfolios with identical expected return and number of assets in the portfolio.

We set up a backtest and use the remaining 200 observations to construct an out-of-sample test in order to investigate the performance of the selected portfolio. The Figures 4.1 and 4.2 shows the difference of return on selected portfolio and benchmak portfolio. The benchmark portfolio represent the average return of the three indices.
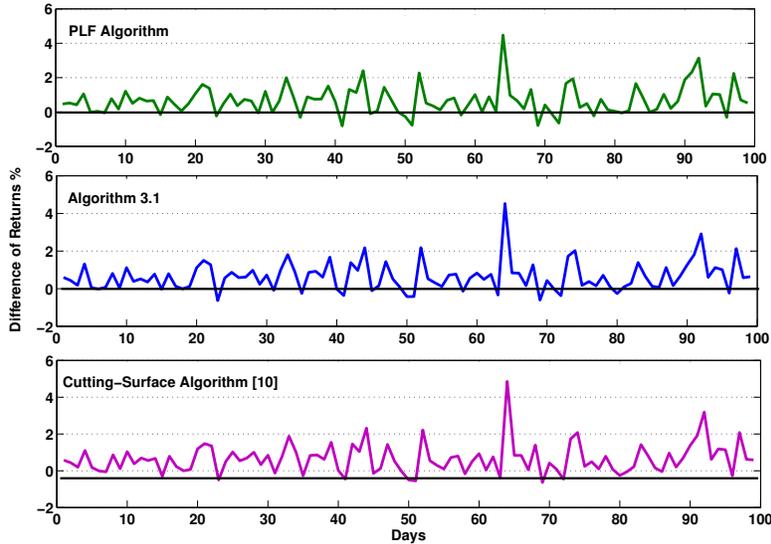

Figure 4.1: Backtesting of the difference of return on selected portfolios and indices. The benchmark portfolio is the average return of the indices.

It can be seen that in both Figures 4.1 and 4.2, the line lies mostly above the zero line which means that the generated portfolio return is higher than the benchmark portfolio.

To illustrate the benefit of using multivariate stochastic dominance constraints, we compare the portfolio strategy constructed by the optimization problem (2.22) with an investment strategy generated by Markowitz model as described below:

$$
\begin{aligned}
\max_{x \in X} \quad & \mathbb{E}[f(x, \xi)] - \lambda \mathbb{E}[R(x, \xi)], \\
\text{s.t.} \quad & \mathbb{E}[g_i(x, \xi)] \geq R_i^b,\ i = 1, \ldots, m, \\
& \sum_{i=1}^{n} x_i = 1,\ x \geq 0,\ x \in X,
\end{aligned}
\tag{4.34}
$$

where $\lambda = 1$ is a fixed nonnegative number, $\mathbb{E}[R(x, \xi)]$ is the portfolio variance, R_i^b is the
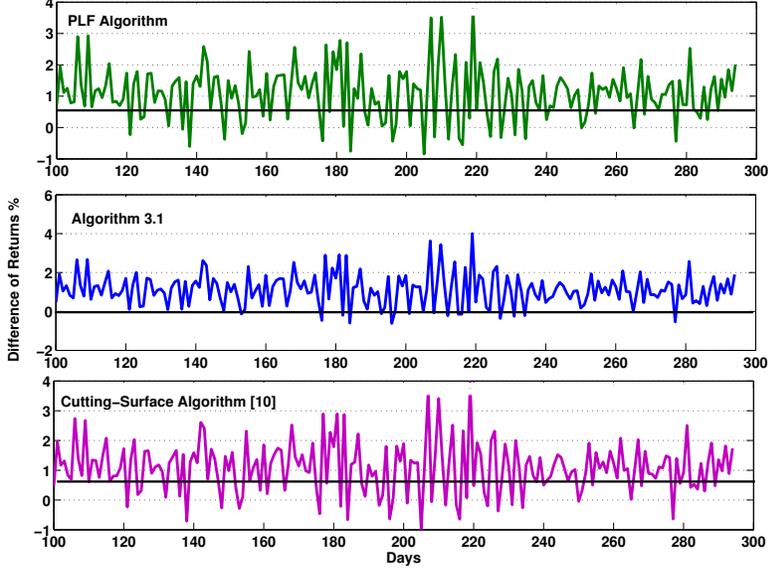
Figure 4.2: Out of sample test of the difference of return on selected portfolios and indices. The benchmark portfolio is the average return of the indices.

benchmark return set equal to the index i, $\mathbb{E}[g_i(x,\xi)]$ is the return of the asset belonging to index i, and $\mathbb{E}[f(x,\xi)]$ is the return defined as in (4.30).

Table 4.4 compares the portfolio generated by Markowitz model to the generated portfolio by the multivariate SSD model. As it can be seen, although the number of assets in the optimal portfolio are the same but the portfolio generated by the Markowitz model has a lower return and a higher CVaR.

Model	No.Assets	Return	CVaR
Multivariate SSD	6	0.034	0.014
Markowitz	6	0.032	0.018

Table 4.4: Time is in minutes. No.Assets refers to the number of assets in the optimal portfolio.

Figures 4.3 and 4.4 present the result of the backtest and out-of-sample test as described earlier. As it can be seen the portfolio generated by the optimization problem (2.23) outperforms the strategy generated by the Markowitz model (4.34) by having relatively higher returns both in-sample and out-of-sample.

To investigate the performance of the generated strategy out-of-sample we present graph of cumulative return of the of portfolio return generated by the multi-SSD model using the Algorithms 4.2-4.5, Markowtitz model and the benchmark portfolio in Figure 4.5. It can be seen that the return generated by the portfolio strategy based on the Multivariate SSD model is much higher compared to the Markowitz model and the benchmark portfolio. Moreover, we also use the Sortino ratio to further compare the generated strategies. The Sortino ratio measures the risk-adjusted return of an investment asset, portfolio or strategy. It is a modification of the Sharpe ratio but penalizes only those returns falling below a user-specified target, or required rate of return, while the Sharpe ratio penalizes both upside and downside volatility equally. We used risk free rate (0.5%) and the benchmark portfolio as the required rate of return. We calculated the Sortino ratio both at the 100th day and 300th day. The results are shown in Table 4.5. As it can be seen the portfolio generated by the multivariate SSD model outperforms the
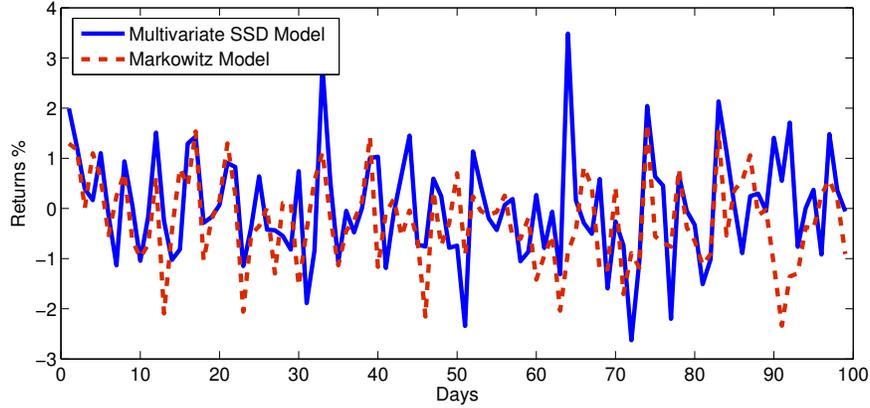
20

Figure 4.3: Comparing the backtest of the portfolio return of the optimization problem with multivariate SSD constraint and the Markowitz model.
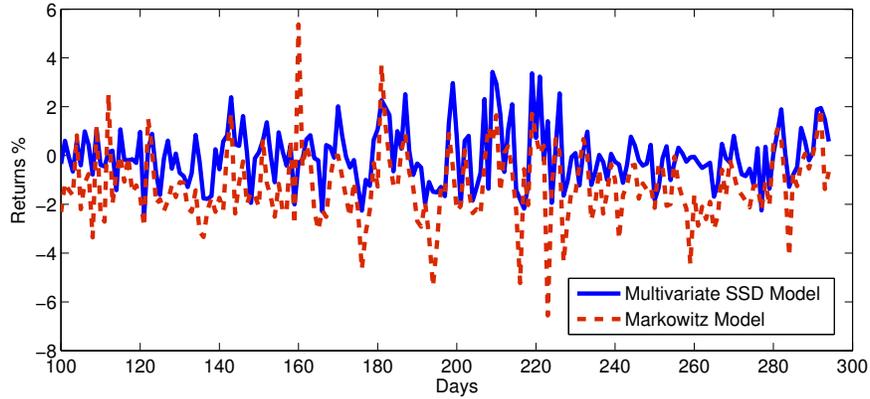

Figure 4.4: Comparing out-of-sample test of the portfolio return of the optimization problem with multivariate SSD constraint and the Markowitz model.

portfolio generated by the Markowitz model by having higher risk-adjusted return.

Model	Required return	100th day Sortino ratio	300th day Sortino ratio
Multivariate SSD Model	Benchmark	0.3969	0.3903
	Risk-free	0.2643	0.0749
Markowitz Model	Benchmark	0.1716	0.1308
	Risk-free	0.1795	0.0637

Table 4.5: Sortino ratio of the portfolio generated by optimization problem with multivariate SSD constraints and the Markowitz model.

Furthermore, we test the algorithms for various number of assets and record the CPU time. Figure 4.6 presents the result for this test. As it can be seen, all algorithms solve relatively large problems within a reasonable time. Additionally, we investigate the performance of the PLF Algorithm, Algorithm 3.1, and Cutting-surface algorithm [10] as the number of observations increases. This is illustrated in Figure 4.7. Although the Figure 4.7 shows that the cutting-surface algorithm [10] becomes inefficient, in our numerical tests increasing the number of observations did not result in a better portfolio.
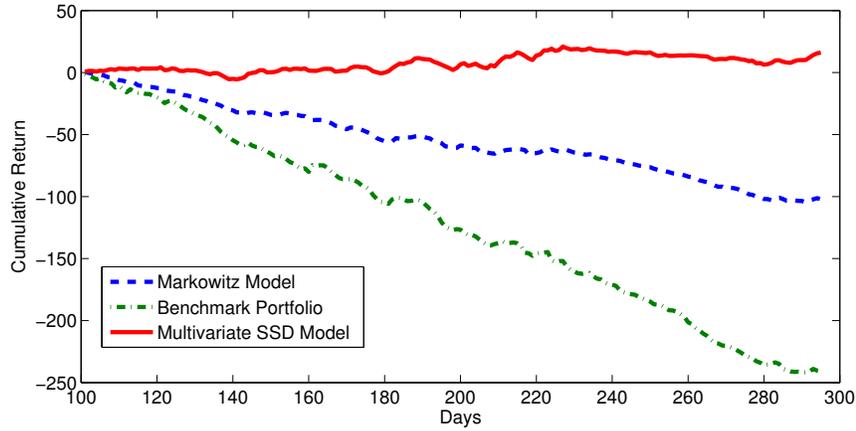
Figure 4.5: Out-of-sample cumulative return for the generated portfolio strategy based on the Multivariate SSD models, Markowitz model and the benchmark portfolio).
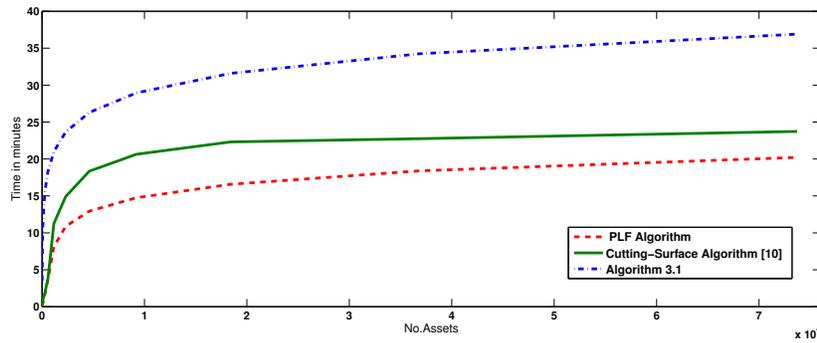

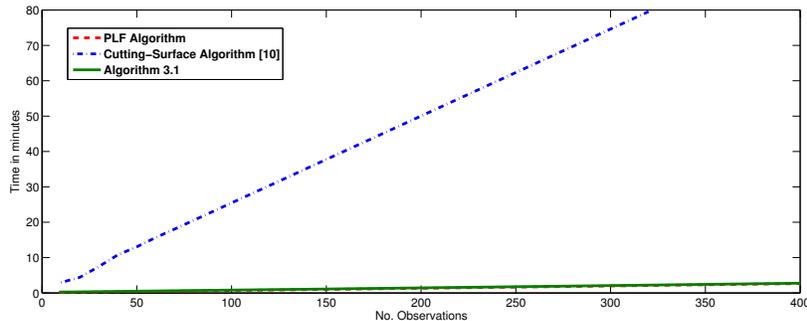Figure 4.6: Graph of CPU Time for various number of instruments for each algorithm.

Figure 4.7: Graph of CPU Time for various number of instruments for each algorithm.

5 Conclusion

In this paper we studied stochastic programming with multivariate second order stochastic dominance constraints. Specifically, we proposed an exact penalty method for second order multivariate stochastic dominance constraints. Furthermore, we solved the penalized problem

(2.22) using the level function method discussed by Meskarian et al. [19] for similar type of problem as well as a modified cutting-plane method inspired by the methods proposed in [24, 8]. These method were compared to the cutting surface method proposed in [10], and the linearized method proposed in [1].

We applied the penalization scheme and the numerical methods to an academic test problem, a budget allocation problem, and a portfolio optimization problem. The academic test results showed that the penalization approach and the numerical methods results in similar optimal solution as the solution generated in [10, Section 2.2]. The budget allocation problem showed that the proposed method solved with PLF Algorithm is more efficient compared to the linearized method when the sample size is large. However, this is not the case when sample size in relatively small. The main advantage of our proposed method to the linearized method is that it can deal with nonlinear underlying functions. In the portfolio optimization problem. we used data of 136 assets from three different indices (FTSE100, Nasdaq100, and Daw Jones). To investigate the performance of generated portfolio strategy, we set up a backtest and an out-of-sample test and compared the performance of the selected portfolio to the corresponding indices. We concluded that the generated portfolio performs better than the indices in sense of higher return both in-sample and out-of-sample.

Furthermore, to illustrate the benefit of considering multivariate stochastic dominance, we introduced the Markowitz model (4.34) and compared the performance of the two portfolio both in-sample and out-of-sample as well as based on the Sortino ratio. It was seen that the portfolio optimization problem with multivariate SSD constraints out perform the portfolio optimization problem based on Markowitz model by having higher risk-adjusted return.

Moreover, we performed a test to investigate the effect of the number of instruments on the computation time for each algorithm. These test suggested as anticipated that the projected level function algorithm and the cutting plane method can solve a large problem within reasonable time.

References

[1] B. Armbruster and J. Luedtke. Models and formulations for multivariate dominance constrained stochastic programs. *Northwestern University*, 2011.

[2] F. H. Clarke. *Optimization and non-smooth analysis*. Wiley, New York, 1983.

[3] D. Dentcheva and A. Ruszczyński. Optimization with stochastic dominance constraints. *SIAM Journal of Optimization*, 13:548–566, 2003.

[4] D. Dentcheva and A. Ruszczyński. Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints. *Mathematical Programming*, 99:329–350, 2004.

[5] D. Dentcheva and A. Ruszczyński. Portfolio optimization with stochastic dominance constraints. *journal of Banking and Finance*, 30:433–451, 2006.

[6] D. Dentcheva and A. Ruszczyński. composite semi-infinite optimization. *Control and Cybernetics*, 36:634–646, 2007.

[7] D. Dentcheva and A. Ruszczyński. Optimization with multivariate stochastic dominance constraints. *Mathematical Programming Series B*, 117:111–127, 2009.

[8] C. I. Fábián, G. Mitra, and D. Roman. Processing second order stochastic dominance models using cutting-plane representations. *Mathematical Programming Series A*, 2009.

[9] P. C. Fishburn. *Utility Theory for Decision Making.* John Wiley & Sons, New York, 1970.

[10] T. Homem-de Mello and S. Mehrota. A cutting surface method for uncertain linear programs with polyhedral stochastic dominance constraints. *SIAM Journal of Optimization*, 20:1250–1273, 2009.

[11] R. Horst and N. V. Thoai. Dc programming: Overview. *Journal of Optimization Theory and Applications*, 103:1–43, 1999.

[12] J. Hu, T. Homem-de Mello, and S. Mehrota. Risk adjusted budget allocation models with application in homeland security. *To appear in IIE Transactions*, 2011.

[13] J. Hu, T. Homem-de Mello, and S. Mehrota. Concepts and application of stochastically weighted stochastic dominance. April, 2011.

[14] J. Hu, T. Homem-de Mello, and S. Mehrota. Sample average approximation of stochastic dominance constrained programs. *Mathematical Programming, Series A*, October 2010.

[15] J. E. Kelley. A cutting plane method for solving convex programs. *SIAM Journal of Applied Mathematics*, 8:703–712, 1960.

[16] W. K. Klein Haneveld and M. H. Ven Der Vlerk. Integrated chance constraints: reduced forms and an algorithm. *Computational Management Science*, 3:254–269, 2006.

[17] C. Lemarechal, A. Nemirovskii, and Y. Nesterov. New variants of bundle methods. *Mathematical Programming*, 69:111–147, 1995.

[18] J. Luedtke. New formulation for optimization under stochastic dominance constraints. *SIAM Journal on Optimization*, 19:1433–1450, 2008.

[19] R. Meskarian, H. Xu, and J. Fliege. Numerical methods for stochastic programs with second order dominance constraints with applications to portfolio. *European Journal of Operational Research*, 216:376–385, 2012.

[20] A. Müller and D. Stoyan. *Comparison Methods for Stochastic Models and Risks.* Wiley, Chichester, 2002.

[21] Y. Nie, X. Wu, and T. Homem-de Mello. Optimal path problems with second-order stochastic dominance constraints. *Networks and Spatial Economics*, pages 1–27, 2011.

[22] W. Ogryczak and A. Ruszczyński. From stochastic dominance to mean-risk models: semideviations as risk measures. *European Journal of Operational Research*, 116:33–50, 1999.

[23] S. M. Robinson. An application of error bounds for convex programming in a linear space. *SIAM Journal of Control and Optimization*, 13:271–273, 1975.

[24] G. Rudolf and A. Ruszczyński. Optimization problems with second order stochastic dominance constraints: Duality, compact formulations, and cut generation methods. *SIAM Journal on Optimization*, 19:1326–1343, 2008.

[25] S. Sun, H. Xu, R. Meskarian, and Y. Wang. Exact penalization, level function method and modified cutting-plane method for stochastic programs with second order stochastic dominance constraints. October,2011.

[26] G. A. Whitmore and M. C. Findlay. *Stochastic Dominance: An Approach to Decison-Making Under Risk.* D. C. Health, Lexington, MA, 1978.

[27] H. Xu. Level function method for quasiconvex programming. *Journal of Optimization Theory and Applications*, 108:407–437, 2001.

6 Appendix

Proof of Theorem 3.2:

The proof is similar to the results in [15]. Note that, at each iteration $t > 0$, $a_{t+1} \in \partial_x \psi(x_t, \nu)$, $d_{t+1} = \nabla \mathbb{E}[f(x_t, \xi)]$, and $e_{t+1} = \nabla_z(\mathbb{E}[f(x_t, \xi)] - z_t = -1$. Then $a_{t+1}^T x - b_{t+1}$ and $d_{t+1}^T x + e_{t+1} z - k_{t+1}$ are the extreme support to the graphs of $\psi(x, \nu)$ and $\mathbb{E}[f(x, \xi)] - z$ at (x_t, z_t) respectively. By condition (a), $\psi(x, \nu)$ and $\mathbb{E}[f(x, \xi)]$ are convex and continuous w.r.t. (x, z). Consequently, if $(x_t, z_t) \in P$ and $\max\{\psi(x, \nu), \mathbb{E}[f(x, \xi)]\} \leq 0$, then

$$\max\{a_{t+1}^T x_t - b_{t+1}, \ d_{t+1}^T x_t + e_{t+1} z - k_{t+1}\} \leq 0.$$

Further, for all $(x_t, z_t) \notin P$,

$$\max\{a_{t+1}^T x_t - b_{t+1}, \ d_{t+1}^T x_t + e_{t+1} z - k_{t+1}\} = \max\{\psi(x_t, \nu), \ \mathbb{E}[f(x_t, \xi)] - z_t\} > 0.$$

Therefore, when $(x_t, z_t) \notin P$, the set P and the point (x_t, z_t) lie on opposite sides of the cutting angle $\max\{a_{t+1}^T x_t - b_{t+1}, \ d_{t+1}^T x_t + e_{t+1} z - k_{t+1}\} = 0$.

Note that, from the definition of P_t and (x_t, z_t), we know that $P \subset P_t \subset P_{t-1}$, (x_t, z_t) minimizes z in P_t and $z_{t-1} \leq z_t$. In the case when $(x_t, z_t) \in P$, it is easy to verify that (x_t, z_t) is the optimal solution of problem (3.26). Indeed, since (x_t, z_t) is an optimal solution, for every $(x, z) \in P_t$, we have $z \geq z_t$. Since $P \subset P_t$, then $z \geq z_t$ for $(x, z) \in P$, which implies optimality of (x_t, z_t) over P.

In what follows, we focus on the case when $(x_t, z_t) \notin P$ $\forall t$. Since $X \times Z$ is a compact set, the sequence $\{(x_t, z_t)\}$ contains a subsequence which converges to $(x^*, z^*) \in X \times Z$. Assume without loss of generality that $(x_t, z_t) \to (x^*, z^*)$. Let $P^* = \cap_t P_t$. Since P_t is compact and $P \subset P_t$, we have $P \subset P^*$ and $(x^*, z^*) \in P^*$. On the other hand, since

$$z \geq z_t, \forall (x, z) \in P_t,$$

then

$$z \geq z^*, \forall (x, z) \in P^*. \tag{6.35}$$

Indeed, if this is not true, then there exists $(\hat{x}, \hat{z}) \in P^*$ such that $\hat{z} < z^*$. Since $z_t \to z^*$, there exists some sufficiently large t such that $\hat{z} < z_t$. This is not possible because (x_t, z_t) is an optimal solution in P_t while $(\hat{x}, \hat{z}) \subset P^* \subset P_t$ is a feasible solutions. This shows that (6.35) holds. Since $P \subset P^*$, the inequality also holds for all $(x, z) \in P$, which implies (x^*, z^*) is an optimal solution of problem (3.27) if $(x^*, z^*) \in P$.

In what follows, we show that $(x^*, z^*) \in P$. Note that, (x_t, z_t) minimizes z in P_t, that is, it satisfies the inequalities:

$$a_{l+1}^T x - b_{l+1} \leq 0, \tag{6.36}$$

and

$$d_{l+1}^T x + e_{l+1} z - k_{l+1} \leq 0, \tag{6.37}$$

for $l = 0, \ldots, t - 1$ and by condition (c), $\max\{\|a_{l+1}\|, \|d_{l+1}\|\} \leq L$, $\forall l$. Let $\{x_t, z_t\}$ denote the subsequence. We claim that $\{\max\{\psi(x_t, \nu), \mathbb{E}[f(x_t, \xi)] - z_t\}\}$ must converge to 0. Note that since

$$
\begin{aligned}
b_{l+1} &= \sum_{i \in \mathcal{J}_l} p_i(-\nabla_x \nu^{*T} G(x_l, \xi^i)^T x_l + \nu^{*T} G(x_l, \xi^i) - \eta^*) + \gamma(\nu^*), \\
&= a_{l+1}^T x_l - \psi(x_l, \nu^*), \\
&= a_{l+1}^T x_l - \psi(x_l, \nu),
\end{aligned}
$$

then (6.36) implies
$$
\psi(x_l, \nu) + a_{l+1}^T(x - x_l) \leq 0.
$$

Similarly, by the definition of e_{l+1}, k_{l+1}, we have from (6.37) that
$$
\mathbb{E}[f(x_l, \xi)] + d_{l+1}^T(x - x_l) - z \leq 0.
$$

Assume that the desired convergence does not occur. Then there exists an $r > 0$ independent of t such that

$$
\begin{aligned}
r &\leq \max\{\psi(x_l, \nu), \mathbb{E}[f(x_l, \xi)] - z_l\} \\
&\leq \max\{a_{l+1}^T(x_l - x_t), d_{t+1}^T(x_l - x_t) - (z_l - z_t)\}, \\
&\leq (L+1)\|(x_l, z_l) - (x_t, z_t)\|,
\end{aligned}
$$

for all $0 \leq l \leq t - 1$, which shows that $\{(x_t, z_t)\}$ does not converge, a contradiction. This shows that
$$
\{\max\{\psi(x_t, \nu), \mathbb{E}[f(x_t, \xi)] - z_t\}\}
$$
converges to 0 and hence $(x_t, z_t) \in P$ is the optimal solution.

The proof is complete.