Sun, H., Xu, H., Meskarian, R. & Wang, Y. (2013). Exact penalization, level function method, and modified cutting-plane method for stochastic programs with second order stochastic dominance constraints. SIAM Journal on Optimization (SIOPT), 23(1), pp. 602-631. doi: 10.1137/110850815

EXACT PENALIZATION, LEVEL FUNCTION METHOD AND MODIFIED CUTTING-PLANE METHOD FOR STOCHASTIC PROGRAMS WITH SECOND ORDER STOCHASTIC DOMINANCE CONSTRAINTS

HAILIN SUN*, HUIFU XU†, RUDABEH MESKARIAN‡, AND YONG WANG§

Abstract. Level function methods and cutting plane methods have been recently proposed to solve stochastic programs with stochastic second order dominance (SSD) constraints. A level function method requires an exact penalization setup because it can only be applied to the objective function, not the constraints. Slater constraint qualification (SCQ) is often needed for deriving exact penalization. It is well known that SSD usually does not satisfy SCQ and various relaxation schemes have been proposed so that the relaxed problem satisfies the SCQ. In this paper, we show that under some moderate conditions the desired constraint qualification can be guaranteed through some appropriate reformulation of the constraints rather than relaxation. Exact penalization schemes based on L_1-norm and L_∞-norm are subsequently derived through Robinson's error bound on convex system and Clarke's exact penalty function theorem. Moreover, we propose a modified cutting plane method which constructs a cutting plane through the maximum of the reformulated constraint functions. In comparison with the existing cutting plane methods, it is numerically more efficient because only a single cutting plane is constructed and added at each iteration. We have carried out a number of numerical experiments and the results show that our methods display better performances particularly in the case when the underlying functions are nonlinear w.r.t. decision variables.

Key words. Slater constraint qualification, exact penalization, modified cutting-plane method, level function method

AMS subject classifications. 90-08, 90C15, 90C30, 90C90

1. Introduction.

Stochastic dominance is a fundamental concept in decision theory and economics. A random outcome $a(\omega)$ is said to dominate another random outcome $b(\omega)$ in the second order, written as $a(\omega) \succeq_2 b(\omega)$, if $\mathbb{E}[v(a(\omega))] \geq \mathbb{E}[v(b(\omega))]$ for every concave nondecreasing function $v(\cdot)$, for which the expected values are finite, see monograph [17] for the recent discussions of the concept.

In their pioneering work [4], Dentcheva and Ruszczyński introduced a stochastic programming model with second order stochastic dominance constraints :

$$\begin{array}{ll} \min\limits_{x} & \mathbb{E}[F(x,\xi(\omega))] \\ \text{s.t.} & G(x,\xi(\omega)) \succeq_2 Y(\xi(\omega)), \\ & x \in \mathcal{X}, \end{array} \tag{1.1}$$

where $\mathcal{X}$ is a closed convex subset of $\mathbb{R}^n$, $\xi : \Omega \to \Xi$ is a vector of random variables defined on probability space $(\Omega, \mathcal{F}, P)$ with support set $\Xi \subset \mathbb{R}^q$, $F : \mathbb{R}^n \times \Xi \to \mathbb{R}$ is convex continuous function w.r.t. x, $G : \mathbb{R}^n \times \Xi \to \mathbb{R}$ is concave continuous function w.r.t. x and $\mathbb{E}[\cdot]$ denotes the expected value with respect to the distribution of $\xi(\omega)$.

*Department of Mathematics, Harbin Institute of Technology, Harbin, 150001, China (mathhlsun@gmail.com).

†School of Engineering and Mathematical Sciences, City University London, London, EC1V 0HB, London, UK; School of Maritime and Transportation, Ningbo University, China. (Huifu.Xu.1@city.ac.uk).

‡School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK (rm3g08@soton.ac.uk).

§Department of Mathematics, Harbin Institute of Technology, Harbin, 150001, China (mathwy@hit.edu.cn).

Here we make a blanket assumption that the expected value of the random function is well defined.

A simple economic interpretation of the model can be given as follows. Let $G(x, \xi(\omega))$ be a profit function which depends on decision vector x and a random variable $\xi(\omega)$, let $F = -G$ and $Y(\xi(\omega))$ be a benchmark profit. Then (1.1) can be viewed as an expected profit maximization problem subject to the constraint that the profit dominates the benchmark profit in second order.

Let $F_1(X; \eta)$ denote cumulative distribution function of random variable X, that is,

$$F_1(X; \eta) := P(X \leq \eta),$$

and

$$F_2(G(x, \xi(\omega)); \eta) := \int_{-\infty}^{\eta} F_1(G(x, \xi(\omega)); t)dt.$$

$G(x, \xi(\omega))$ is said to dominate $Y(\xi(\omega))$ in first order, denoted by $G(x, \xi(\omega)) \succeq_1 Y(\xi(\omega))$, if for all $\eta \in \mathbb{R}$,

$$F_1(G(x, \xi(\omega)); \eta) \leq F_1(Y(\xi(\omega)); \eta).$$

$G(x, \xi(\omega))$ is said to dominate $Y(\xi(\omega))$ in second order, denoted by $G(x, \xi(\omega)) \succeq_2 Y(\xi(\omega))$, if

$$F_2(G(x, \xi(\omega)); \eta) \leq F_2(Y(\xi(\omega)); \eta), \forall \eta \in \mathbb{R}. \tag{1.2}$$

It is easy to observe that first order stochastic dominance implies second order stochastic dominance. It is well known that second order dominance constraint in (1.1) can be reformulated as

$$\mathbb{E}[(\eta - G(x, \xi(\omega)))_+] \leq \mathbb{E}[(\eta - Y(\xi(\omega)))_+], \ \forall \eta \in \mathbb{R}, \tag{1.3}$$

where $(x)_+ = \max(0, x)$; see [6]. Ogryczak and Ruszczyński [18] investigated the relationship between stochastic second order dominance and mean-risk models. In a more recent development [5, Theorem 3.1], the second order dominance is shown to be equivalent to a continuum of conditional value at risk constraints: one for each probability level. Using the reformulation of the second order dominance constraints, Dentcheva and Ruszczyński [6] reformulated (1.1) as:

$$\begin{array}{ll} \min\limits_{x} & \mathbb{E}[F(x, \xi(\omega))] \\ \text{s.t.} & \mathbb{E}[(\eta - G(x, \xi(\omega)))_+] \leq \mathbb{E}[(\eta - Y(\xi(\omega)))_+], \ \forall \eta \in \mathbb{R}, \\ & x \in \mathcal{X}. \end{array} \tag{1.4}$$

To ease notation, we will use ξ to denote the random vector $\xi(\omega)$ and a deterministic vector, depending on the context. It is well known that (1.4) does not satisfy the well known Slater constraint qualification, a condition that is often needed for deriving first order optimality conditions of the problem and developing a numerically stable method for solving the problem. Subsequently, a so-called *relaxed* form of (1.4) is proposed:

$$\begin{array}{ll} \min\limits_{x} & \mathbb{E}[F(x, \xi)] \\ \text{s.t.} & \mathbb{E}[(\eta - G(x, \xi))_+] \leq \mathbb{E}[(\eta - Y(\xi))_+], \ \forall \eta \in [a, b], \\ & x \in \mathcal{X}, \end{array} \tag{1.5}$$

where $[a, b]$ is a closed interval in $\mathbb{R}$. Over the past few years, Dentcheva and Ruszczyński have developed comprehensive theory of optimality and duality for (1.5), see [4, 5, 6].

Unfortunately, problem (1.5) is difficult to solve since it is a stochastic semi-infinite nonsmooth programming problem. In the case when $G(x, \xi)$ and $F(x, \xi)$ are linear w.r.t. x and ξ has finitely many scenarios, Dentcheva and Ruszczyński [5] reformulated problem (1.5) as a linear programming (LP) problem by introducing new variables which represent positive parts in each constraint of problem (1.5). The reformulation effectively tackles the nonsmoothness in the second order constraints and the approach can easily be used to the case when G and F are nonlinear. This reformulation, however, introduces many new variables particularly when the random variable ξ has many distributional scenarios. Apparently this does not have significant impact on numerical implementation as the existing solvers for LP are very powerful (can deal with millions of variables). However, the impact will be much more significant when F and G are nonlinear and this is indeed one of the key issues this paper is to address.

Rudolf and Ruszczyński [22] and Fábián et al [8] proposed cutting-plane methods for solving a stochastic program with second order stochastic dominance constraints. A crucial element of the method in [8] is based on an observation that when F and G are linear w.r.t. x and probability space Ω is finite, the constraint function in the second order dominance constraint is the convex envelope of a finitely many linear functions, which is called cutting-plane representation and observed by Haneveld and van der Vlerk in [11]. Subsequently, an iterative scheme which exploits the fundamental idea of classical cutting-plane method is proposed where at each iterate "cutting-plane" constraints are constructed and added. This also effectively tackles the nonsmoothness issue caused by the plus function. While the method displays strong numerical performance, it relies on discreteness of the probability space as well as the linearity of F and G. Hu, Homem-de-Mello and Mehrotra [10] and Homem-de-Mello and Mehrotra [9] also proposed a cut generation algorithm for solving a sample average approximation (SAA) problem of stochastic program with multivariate stochastic dominance constraints. Different from the cutting plane method in [22] and [8], they reformulated every subproblem as linear programming problem by introducing some new variables when F and G are linear.

In this paper we consider problem (1.4) with a focus on the case when ξ has a discrete distribution, that is,

$$\begin{array}{ll} \min\limits_{x} & \sum\limits_{i=1}^{N} p_i F(x, \xi^i) \\ \text{s.t.} & \sum\limits_{i=1}^{N} p_i (\eta - G(x, \xi^i))_+ \leq \sum\limits_{i=1}^{N} p_i (\eta - Y(\xi^i))_+, \ \eta \in \mathbb{R}, \\ & x \in \mathcal{X}. \end{array} \tag{1.6}$$

Here the random variable ξ has a finite distribution, that is, $P(\xi = \xi^i) = p_i$, for $i = 1, \cdots, N$. When $p_i = \frac{1}{N}$, problem (1.6) can be viewed as a sample average approximation of problem (1.4). We investigate the Slater constraint qualification of the problem and its reformulation, exact penalization schemes and numerical methods. Specifically, we make the following contributions.

- We develop penalization schemes for problem (1.6). We do so by exploiting Clarke's exact penalty function theorem [2, Proposition 2.4.3] and Robinson's

error bound [19]. The latter requires SCQ. Unfortunately, problem (1.6) or its reformulation (2.5) does not satisfy the constraint qualification (see discussions by Dentcheva and Ruszczyński at pages 558-559 in [4]). Here we propose an *alternative way* to reformulate the constraints of problem (2.5). We then demonstrate that the newly reformulated problem (see (2.7)) satisfies the SCQ under some moderate conditions (see Theorem 2.1). Two exact penalization schemes based on L_1-norm and L_∞-norm are subsequently derived and are shown that they are exact penalization of problem (2.5) although the latter does not satisfy SCQ (see Theorems 3.1 and 3.2). Note that Liu and Xu [14] proposed an exact penalization scheme with L_∞-norm for the relaxed problem (1.5). A crucial condition is the SCQ of problem (1.5) which relies on the *relaxation* because the original problem (1.4) may not satisfy the SCQ. Our penalization schemes in this paper differ from Liu and Xu's in that they are proposed for the original problem rather than for the relaxed problem, which means that they are established without the SCQ of the original problem. This makes the penalization scheme more appealing given the fact that the original problem usually does not satisfy the SCQ particularly in the case when ξ satisfies discontinuous distribution. Based on the exact penalization formulations, we apply a well known level function method in nonsmooth optimization [13, 25] to the penalized problems. An obvious advantage of this approach is that we can effectively deal with excessive number of constraints, nonsmoothness in the constraints and nonlinearity of the underlying functions.

- We also propose a modified cutting-plane method to solve the problem. The cutting-plane method differs from those in the literature [22] in that it applies to the maximum of the constraint functions rather than each constraint function. This saves considerable computation time because at each iteration, our cutting-plane method requires to add a couple of linear constraints whereas the cutting-plane method in [22] requires to add N constraints (N is the cardinality of the support set Ξ). The approach also differs from that in [9, 10] because our modified cutting-plane method uses the cutting-plane representation proposed in [11]. The idea of applying the cutting-plane method to the maximum of the constraint functions is similar to the idea in algorithm proposed by Fábián, Mitra and Roman, see the algorithm at page 48 in [8]. Note that Fábián, Mitra and Roman's algorithm is applied to linear models while Algorithm 4.1 is applicable to nonlinear case. Therefore we may regard our algorithm as an extension of theirs.
- We have carried out extensive numerical tests on our proposed methods in comparison with the cutting plane method in [8]. The numerical results show that our proposed methods are more efficient. Specifically, we have discovered that level function method based on exact penalization scheme with L_∞-norm is most efficient in terms of computation time; the modified cutting-plane method (Algorithm 4.1) performs also efficiently.

It might be interesting to extend the proposed penalty schemes and algorithms to the stochastic problem with multivariate dominance constraints when the random variables in the problem have finite support sets, we will do this in our follow-up work [15]. Another interesting extension will be to develop similar schemes and algorithms of second order dominance programs with continuous distributions although we may apply them to the sample average approximated problems [9, 23]. A potential advan-

tage of doing this is to apply stochastic approximation methods (e.g. [7]) other than SAA.

The rest of the paper is organized as follows. In section 2, we discuss the SCQ of problem (1.6) and its reformulation. In section 3, we propose two exact penalization schemes for problem (2.5) and apply a level function method to solve them. In section 4, a modified cutting-plane method has been proposed for solving the problem and finally in section 5, we report some numerical test results.

Throughout this paper, we use the following notation. $x^T y$ denotes the scalar product of two vectors x and y, $\|\cdot\|$, $\|\cdot\|_1$ and $\|\cdot\|_\infty$ denote the Euclidean norm, L_1-norm and L_∞-norm of a vector and a compact set of vectors respectively. We also use $\|\cdot\|$ to denote the infinity norm of a continuous function space and its induced norm of a linear operator. $d(x,\mathcal{D}) := \inf_{x'\in\mathcal{D}} \|x-x'\|$, $d_1(x,\mathcal{D}) := \inf_{x'\in\mathcal{D}} \|x-x'\|_1$ and $d_\infty(x,\mathcal{D}) := \inf_{x'\in\mathcal{D}} \|x-x'\|_\infty$ denote the distance from point x to set $\mathcal{D}$ in Euclidean norm, L_1-norm and L_∞-norm respectively. For a real valued-function $h(x)$, we use $\nabla h(x)$ to denote the gradient of h at x.

2. Slater constraint qualification. In the literature of stochastic programs with second order dominance constraints, SCQ has been used as a key condition for deriving optimality conditions and exact penalization, see for instances [4, 14].

Recall that problem (1.6) is said to satisfies the SCQ if there exists $x_0 \in \mathcal{X}$ such that

$$\sum_{i=1}^N p_i(\eta - G(x_0,\xi^i))_+ - \sum_{i=1}^N p_i(\eta - Y(\xi^i))_+ < 0,\ \eta \in \mathbb{R}. \tag{2.1}$$

Unfortunately, this kind of constraint qualification is not satisfied. To see this, let

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

and

$$\underline{y} := \min\{Y(\xi^1),\cdots,Y(\xi^N)\}. \tag{2.2}$$

For any $\eta \le \underline{y}$, it is easy to verify that $\mathbb{E}[(\eta - Y(\xi))_+] = 0$. For those η, the feasible constraint of problem (1.5) reduces to

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

because the term at the left hand side of the equation is non-negative. This means that there does not exist a feasible point $x_0 \in \mathcal{X}$ such that (2.1) holds.

Dentcheva and Ruszczyński [4] observed this issue and tackled it by considering a relaxed problem (1.5) which effectively restricts η to take value from a specified $[a,b]$. In other words, the feasible region of the original problem (1.6) is enlarged. In this context, their relaxation scheme can be written as follows:

$$\begin{array}{ll} \min\limits_x & \sum\limits_{i=1}^N p_i F(x,\xi^i) \\ \text{s.t.} & \sum\limits_{i=1}^N p_i(\eta - G(x,\xi^i))_+ \le \sum\limits_{i=1}^N p_i(\eta - Y(\xi^i))_+,\ \eta \in [a,b], \\ & x \in \mathcal{X}. \end{array} \tag{2.3}$$

Under some circumstance, it is possible to choose a proper value a such that problem (2.3) satisfies the SCQ. For instance, if there exists a point $x_0 \in \mathcal{X}$ such that

$$G(x_0, \xi) \succeq_1 Y(\xi)$$

and for every $\xi \in \Xi$, $G(x_0, \xi) > \underline{y}$, then x_0 is a feasible point of problem (2.3) and

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

for all $\eta > \underline{y}$. In such a case, it is easy to verify that the SCQ holds for any $a > \underline{y}$. Note that problem (2.3) is a relaxed problem of (1.6) which depends on $[a, b]$ and when $[a, b]$ contains $Y(\Xi)$, the SCQ fails.

Hu, Homem-de-Mello and Mehrotra [10] proposed an alternative approach to deal with the SCQ issue by considering ϵ-feasible solutions:

$$\begin{array}{ll} \min\limits_{x} & \mathbb{E}[F(x, \xi(\omega))] \\ \text{s.t.} & \mathbb{E}[(\eta - G(x, \xi(\omega)))_+] \le \mathbb{E}[(\eta - Y(\xi(\omega)))_+] + \epsilon, \ \forall \eta \in \mathbb{R}, \\ & x \in \mathcal{X}. \end{array} \tag{2.4}$$

where ϵ is a small positive number. It is easy to observe that the relaxed problem satisfies the SCQ for any positive ϵ as long as the original problem is feasible. A theoretical issue to be addressed in the aforementioned relaxation schemes is to show 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 (to ensure lower semicontinuity of the feasible solution set of the relaxed problems).

In this section, we propose an *alternative way* to address the issue of SCQ of problem (1.6) without relaxation. To this end, let us use [4, Proposition 3.2] to reformulate problem (1.6) as follows:

$$\begin{array}{ll} \min\limits_{x} & \sum\limits_{i=1}^{N} p_i F(x, \xi^i) \\ \text{s.t.} & \sum\limits_{i=1}^{N} p_i (Y_j - G(x, \xi^i))_+ \le \gamma_j, \ j = 1, \ldots, N, \\ & x \in \mathcal{X}, \end{array} \tag{2.5}$$

where $Y_j := Y(\xi^j)$, $\gamma_j := \sum_{i=1}^{N} p_i (Y_j - Y_i)_+$. Like the original problem (1.6), the reformulated problem (2.5) does not satisfy SCQ.

We consider the power set of $\{1, \ldots, N\}$, that is, a collection of all subsets of $\{1, \ldots, N\}$ including empty set and itself. For the simplicity of notation, let $\mathcal{N}$ denote the power set excluding the empty set and for $j = 1, \ldots, N$,

$$\psi_j(x) := \max_{\mathcal{J} \in \mathcal{N}} \sum_{i \in \mathcal{J}} p_i (Y_j - G(x, \xi^i)) - \gamma_j. \tag{2.6}$$

Consider problem

$$\begin{array}{ll} \min\limits_{x} & \sum_{i=1}^{N} p_i F(x, \xi^i) \\ \text{s.t.} & \psi_j(x) \le 0, \text{for } j = 1, \ldots, N, \\ & x \in \mathcal{X}. \end{array} \tag{2.7}$$

Note that (2.7) is similar to the cutting plane representation in [8, 22]. The key difference is that here $\mathcal{J}$ is restricted to take a nonempty subset of $\{1,\dots,N\}$, which allows $\max_{\mathcal{J}\in\mathcal{N}}\sum_{i\in\mathcal{J}}p_i(Y_j-G(x,\xi^i))$ to take a negative value (otherwise the term is nonnegative as $\sum_{i\in\mathcal{J}}p_i(Y_j-G(x,\xi^i))=0$ for $\mathcal{J}=\emptyset$). This paves the way for (2.7) to satisfy the SCQ in later discussions. In what follows, we will show that problem (2.7) is equivalent to problem (2.5) but, under some circumstance, the former satisfies the SCQ.

LEMMA 2.1. *For $j=1,\cdots,N$, let*

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

Then

$$\sum_{i=1}^{N}p_i(Y_j-G(x,\xi^i))_+=\max\{\varphi_j(x),0\}. \tag{2.8}$$

Proof. We prove the claim by going through two cases: 1. $\varphi_j(x)\le 0$; 2. $\varphi_j(x)>0$.

Case 1. Since $\varphi_j(x)\le 0$, then $\max\{\varphi_j(x),0\}=0$ and $Y_j-G(x,\xi^i)\le 0$ for $j\in\{1,\dots,N\}$. The latter implies

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

and hence (2.8).

Case 2. Since $\varphi_j(x)>0$, then there exists a nonempty subset $\mathcal{J}\subseteq\{1,\dots,N\}$ such that

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

It suffices to show that

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

or equivalently $\mathcal{J}$ contains every index i with $Y_j-G(x,\xi^i)>0$. Indeed, if $\mathcal{J}$ does not include such an index, then adding it to $\mathcal{J}$ would increase the quantity $\sum_{i\in\mathcal{J}}p_i(Y_j-G(x,\xi^i))$ and this contradicts the fact that $\varphi_j(x)$ is the maximum. Likewise, $\mathcal{J}$ does not contain an index i with $Y_j-G(x,\xi^i)<0$, because, otherwise, removing the index will also increase the quantity $\sum_{i\in\mathcal{J}}p_i(Y_j-G(x,\xi^i))$. This completes the proof. □

We are now ready to state the main result in this section.

THEOREM 2.1. *Let $G(x,\xi)$ and $Y(\xi)$ be defined as in problem (1.6) and ψ_j be defined as in (2.6). Then*

(i) $G(x,\xi)\succeq_2 Y(\xi)$ if and only if

$$\psi_j(x)\le 0,\ \forall j=1,\dots,N; \tag{2.9}$$

(ii) problems (2.5) and (2.7) are equivalent;

(iii) if there exists a feasible point x_0 such that $G(x_0,\xi)\succeq_1 Y(\xi)$ and $G(x_0,\xi)>\underline{y}$ for all $\xi\in\Xi$, then the system of inequalities (2.9) satisfies the SCQ.

Proof. Part (i). By [4, Proposition 3.2], $G(x,\xi) \succeq_2 Y(\xi)$ if and only if

$$\sum_{i=1}^{N} p_i(Y_j - G(x,\xi^i))_+ \leq \gamma_j,\ j = 1,\ldots,N, \tag{2.10}$$

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

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

By (2.8)

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

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

$$\max\{\max\{a,0\} - r, 0\} = \max\{a - r, 0\}. \tag{2.11}$$

Using (2.11), we have that

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

The last equality is due to the definition of ψ_j. The discussion above demonstrates that (2.10) is equivalent to (2.9) and hence the conclusion.

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

$$\left\{x \in \mathcal{X} : \sum_{i=1}^{N} p_i(Y_j - G(x,\xi^i))_+ - \gamma_j \leq 0\right\} = \{x \in \mathcal{X} : \psi_j(x) \leq 0\}.$$

Part (iii). Let $\gamma_{\underline{y}} := \sum_{i=1}^{N} p_i(\underline{y} - Y(\xi^i))_+$, where $\underline{y}$ is defined in (2.2). By the definition of $\underline{y}$, the right hand side equals to 0. Therefore $\gamma_{\underline{y}} = 0$. Likewise, the assumption $G(x_0,\xi) > \underline{y}$ for $\xi \in \Xi$ implies

$$\sum_{i\in\mathcal{J}} p_i(\underline{y} - G(x_0,\xi^i)) < 0$$

for every nonempty index set $\mathcal{J} \subseteq \{1,\ldots,N\}$. This shows

$$\max_{\mathcal{J}\in\mathcal{N}} \sum_{i\in\mathcal{J}} p_i(\underline{y} - G(x_0,\xi^i)) - \gamma_{\underline{y}} < 0. \tag{2.12}$$

Assume without loss of generality that the N elements in set $Y(\Xi)$ satisfies the following order

$$Y_1 \leq Y_2 \leq \cdots \leq Y_N.$$

By the definition of $\psi_j(x)$ (see (2.9), inequality (2.12) means that

$$\psi_1(x_0) = \max_{\mathcal{J}\in\mathcal{N}} \sum_{i\in\mathcal{J}} p_i(Y_1 - G(x_0,\xi^i)) - \sum_{i=1}^{N} p_i(Y_1 - Y(\xi^i))_+ < 0.$$

In what follows, we show $\psi_j(x_0) < 0$, for $j = 2, \cdots, N$. By definition

$$\begin{aligned}
\psi_j(x_0) &= \max_{\mathcal{J}\in\mathcal{N}} \sum_{i\in\mathcal{J}} p_i(Y_j - G(x_0,\xi^i)) - \sum_{i=1}^{N} p_i(Y_j - Y(\xi^i))_+ \\
&\le \max\left\{ \max_{\mathcal{J}\in\mathcal{N}} \sum_{i\in\mathcal{J}} p_i(Y_j - G(x_0,\xi^i)), 0 \right\} - \sum_{i=1}^{N} p_i(Y_j - Y(\xi^i))_+ \\
&= \sum_{i=1}^{N} p_i((Y_j - G(x_0,\xi^i))_+ - (Y_j - Y(\xi^i))_+) \\
&= \int_{-\infty}^{Y_j} \big(F_1(G(x_0,\xi),t) - F_1(Y(\xi),t)\big)dt. \qquad (2.13)
\end{aligned}$$

The second last equality follows from Lemma 2.1 and the last equality is due to the equivalent representation of second order dominance between (1.2) and (1.3).

Assume without loss of generality that $Y_2 > Y_1$ (otherwise $\psi_2(x_0) = \psi_1(x_0) < 0$). Let $\bar\eta \in (Y_1, \min\{\min_{\xi\in\Xi} G(x_0,\xi), Y_2\})$. Note that, by assumption, $Y_1 < \min\{\min_{\xi\in\Xi} G(x_0,\xi), Y_2\}$, $\bar\eta$ exists. Then

$$\begin{aligned}
\int_{-\infty}^{Y_j} F_1(G(x,\xi),t) - F_1(Y(\xi),t)dt &= \int_{-\infty}^{\bar\eta} F_1(G(x,\xi),t) - F_1(Y(\xi),t)dt \\
&+ \int_{\bar\eta}^{Y_j} F_1(G(x,\xi),t) - F_1(Y(\xi),t)dt.
\end{aligned}$$

Note that

$$\int_{-\infty}^{\bar\eta} F_1(G(x,\xi),t) - F_1(Y(\xi),t)dt = 0 - p_1(\bar\eta - Y_1) < 0$$

where p_1 is the probability that $Y(\xi)$ takes value Y_1. On the other hand, $G(x_0,\xi) \succeq_1 Y(\xi)$ implies

$$\int_{\bar\eta}^{Y_j} (F_1(G(x_0,\xi),t) - F_1(Y(\xi),t))dt \le 0.$$

This shows

$$\int_{-\infty}^{Y_j} \big(F_1(G(x_0,\xi),t) - F_1(Y(\xi),t)\big)dt < 0, \text{ for } j = 2, \cdots, N. \qquad (2.14)$$

The conclusion follows by combining (2.12), (2.13) and (2.14). □

Theorem 2.1 says that although problems (1.6) and (2.5) do not satisfy SCQ, the reformulated problem (2.7) may do under some circumstance. The fundamental reason behind this has to do with the plus function $(\cdot)_+$. Consider a single variate function $a(x) = x$. It is easy to see that the single inequality $a(x) \le 0$ satisfies SCQ but $(a(x))_+ \le 0$ does not although the two inequalities represent the same set $(-\infty, 0]$. Clearly, the constraint qualification is closely related to the function which represents the feasible set. In problem (2.7), we give an alternative presentation of the feasible constraints of (1.6) and (2.5) without the plus function (which could potentially destroy the SCQ).

It might be helpful to discuss how strong the conditions in part (iii) of Theorem 2.1 are. Assume that ξ has a finite distribution, that is, $\Xi = \{\xi^1, ...\xi^N\}$. If there exists a point $x_0 \in X$ such that $G(x_0, \xi^i) \geq Y(\xi^i)$ for each i and

$$\min_{i\in\{1,...,N\}} G(x_0, \xi^i) > \min_{i\in\{1,...,N\}} Y(\xi^i),$$

then the conditions are satisfied. In the context of portfolio optimization, this means there exists a feasible strategy which exhibits a return not worse than the benchmark strategy in any scenario, and the worst outcome of the return based on this feasible strategy is strictly better than the worst outcome that the benchmark strategy could possibly generate.

3. Exact penalization schemes and level function method. Problem (2.7) is an ordinary nonlinear programming problem with finite number of constraints. This means that we can apply any existing NLP code to solve it. However, from numerical point of view, problem (2.7) is difficult to solve because every constraint $\psi_j(x)$ is a maximum function of $2^N - 1$ functions. That means problem (2.7) contains $(2^N - 1)N$ constraints which depends on N, the cardinality of support set Ξ, and this may make the problem difficult to solve by well-known NLP methods such as the active set method even when N is not very large.

This motivates us to consider an exact penalty function method which is well known in nonlinear programming. Liu and Xu [14] proposed an L_∞-norm based penalization scheme for the relaxed problem (1.5). In this context, their penalization scheme can be written as follows:

$$\min_x \quad \sum_{i=1}^N p_i F(x, \xi^i) + \rho \max_{\eta\in[a,b]} \left(\sum_{i=1}^N p_i((\eta - G(x, \xi^i))_+ - (\eta - Y(\xi^i))_+) \right)_+ . \tag{3.1}$$

Justification of the penalty scheme (the equivalence of problem (1.5) and (3.1)) requires SCQ but the constraint qualification is not satisfied when $Y(\xi) \subset [a, b]$.

In this section, we apply the penalty function method to problem (2.7). There are essentially two ways to apply the penalty function method in this paper. One is to apply an exact penalty function method with L_∞-norm to problem (2.7). The other is to use an exact penalty function method with L_1-norm. In this section, we consider both of them.

To this end, we need the following technical result.

LEMMA 3.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function and $g : \mathbb{R}^n \to \mathbb{R}^m$ be a continuous vector-valued function whose components are convex. Let $X \subseteq \mathbb{R}^n$ be a compact and convex set. Consider the following constrained minimization problem*

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) \leq 0, \\ & x \in X. \end{array} \tag{3.2}$$

(i) If $g(x)$ satisfies SCQ, that is, there exists a point x_0 and a real number $\delta > 0$ such that

$$\delta\mathcal{B} \subset g(x_0) + K,$$

and the feasible set, denoted by S, is bounded, then

$$d(x, S) \leq \delta^{-1} D d_1(0, g(x) + K)$$

and

$$d(x,S) \le \delta^{-1} D d_\infty(0, g(x)+K),$$

where $\mathcal{B}$ denotes the closed unit ball in $\mathbb{R}^m$ and $K := [0,+\infty)^m$, and D denotes the diameter of S.

(ii) If $f(x)$ is Lipschitz continuous on X with modulus κ, then for

$$\rho > \kappa\delta^{-1} D,$$

the set of optimal solutions of (3.2) coincides with the set of optimal solutions of problem

$$\begin{array}{ll} \min & f(x) + \rho\|(g(x))_+\|_1 \\ \text{s.t.} & x \in X, \end{array} \tag{3.3}$$

and that of

$$\begin{array}{ll} \min & f(x) + \rho\|(g(x))_+\|_\infty \\ \text{s.t.} & x \in X. \end{array} \tag{3.4}$$

Proof. Part (i) follows from Robinson's error bound for convex systems [19] and Part (ii) follows from Part (i) and Clarke's exact penalty function theorem [2, Proposition 2.4.3]. □

3.1. Exact penalization with L_1-norm. A popular exact penalty scheme in optimization is based on L_1-norm. Here we consider the penalization scheme for (2.7):

$$\begin{array}{ll} \min\limits_x & \sum\limits_{i=1}^N p_i F(x,\xi^i) + \bar\rho \sum\limits_{j=1}^N (\psi_j(x))_+ \\ \text{s.t.} & x \in \mathcal{X}, \end{array} \tag{3.5}$$

and for (2.5):

$$\begin{array}{ll} \min\limits_x & \vartheta_{\bar\rho}(x) := \sum\limits_{i=1}^N p_i F(x,\xi^i) + \bar\rho \sum\limits_{j=1}^N (\sum\limits_{i=1}^N p_i (Y_j - G(x,\xi^i))_+ - \gamma_j)_+ \\ \text{s.t.} & x \in \mathcal{X}. \end{array} \tag{3.6}$$

In what follows, we show that the two penalty schemes are equivalent, and estimate the penalty parameter. As we discussed in the preceding section, (2.5) does not satisfy the SCQ, but (2.7) does under some moderate conditions. A key point we want to make here is that exact penalization function scheme (3.6) is justified despite (2.5) does not satisfy the SCQ.

We need the following assumption on the underlying random functions of problem (1.6).

ASSUMPTION 3.1. *$F(x,\xi^i)$, $G(x,\xi^i)$ are continuously differentiable w.r.t. x in an open neighborhood of $\mathcal{X}$, for $i = 1,\cdots,N$. Moreover, they are globally Lipschitz over $\mathcal{X}$, that is, there exists $\kappa(\xi) < +\infty$ such that*

$$\max(\|\nabla_x F(x,\xi^i)\|, \|\nabla_x G(x,\xi^i)\|) \le \kappa(\xi^i)$$

for $i = 1,\cdots,N$.

THEOREM 3.1. *Assume: (a) problem (2.7) satisfies SCQ, (b) Assumption 3.1 holds; (c) the feasible set of problem (2.7) is bounded. Then*

(i) problems (3.5) and (3.6) are equivalent;
(ii) there exist positive constants $\bar{\delta}$ and $\bar{D}$ such that when

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

the set of optimal solutions of (2.7) coincides with that of (3.5), and the set of optimal solutions of (2.5) coincides with that of (3.6).

Proof. Part (i). Through Lemma 2.1 and (2.11), it is easy to verify that

$$\sum_{j=1}^{N} (\psi_j(x))_+ = \sum_{j=1}^{N} (\varphi_j(x) - \gamma_j)_+ = \sum_{j=1}^{N} (\sum_{i=1}^{N} p_i (Y_j - G(x, \xi^i))_+ - \gamma_j)_+, \tag{3.7}$$

where $\varphi_j(\cdot)$ is defined in Lemma 2.1. The conclusion follows from (3.7).

Part (ii). Let

$$\Phi(x) := (\psi_1(x), \cdots, \psi_N(x))^T$$

and $\mathcal{Q}$ be the feasible set of problem (2.7). 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 (2.7) is a convex programming problem and satisfies SCQ, by Lemma 3.1, there exists real numbers $\bar{\delta} > 0$ and $\bar{D} > 0$ such that when

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

the set of optimal solutions of problem (2.7) coincides with that of (3.5). Moreover, since problem (2.5) and (2.7) are equivalent, while problem (3.5) and (3.6) are equivalent, the set of optimal solutions of problem (2.5) coincides with that of (3.6). □

Theorem 3.1 shows that the exact penalization (with L_1-norm) of problem (2.5) can be derived although it does not satisfy SCQ. This is achieved through problem (2.7). Since the reformulation of (2.5) depends on the distribution of random variable ξ, it is unclear whether Theorem 3.1 can be generalized to the case when ξ satisfies continuous distribution.

3.2. Exact penalization with L_∞-norm. Another popular penalty scheme in optimization is based on L_∞-norm. Here we consider the penalization scheme for (2.7)

$$\begin{array}{ll} \min\limits_{x} & \sum\limits_{i=1}^{N} p_i F(x, \xi^i) + \hat{\rho} \max\limits_{j \in \{1,\ldots,N\}} (\psi_j(x))_+ \\ \text{s.t.} & x \in \mathcal{X}, \end{array} \tag{3.8}$$

and for (2.5)

$$\begin{array}{ll} \min\limits_{x} & \hat{\vartheta}_{\hat{\rho}}(x) := \sum\limits_{i=1}^{N} p_i F(x, \xi^i) + \hat{\rho} \max\limits_{j \in \{1,\ldots,N\}} (\sum\limits_{i=1}^{N} p_i (Y_j - G(x, \xi^i))_+ - \gamma_j)_+ \\ \text{s.t.} & x \in \mathcal{X}. \end{array} \tag{3.9}$$

Similar to the discussions in the preceding subsection, we need to show that the two penalty schemes are equivalent and give an estimate of the penalty parameter $\bar{\rho}$.

THEOREM 3.2. *Assume: (a) problem (2.7) satisfies SCQ, (b) Assumption 3.1 holds; (c) the feasible set of problem (2.7) is bounded. Then*

(i) problems (3.8) and (3.9) are equivalent;

(ii) there exist positive constants $\hat{\delta}$ and $\hat{D}$ such that when

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

the set of optimal solutions of (2.7) coincides with that of (3.8), and the set of optimal solutions of (2.5) coincides with that of (3.9).

Proof. The proof is similar to Theorem 3.1, we omit details here. □

Analogous to the comments following Theorem 3.1, we note that a main contribution of Theorem 3.2 is to show exact penalization scheme with L_∞-norm can be established for problem (2.5) despite it does not satisfy SCQ. The observation makes the exact penalization schemes more appealing because they can be applied to a fairly large class of problems.

Note that our exact penalty schemes are established through Clarke's penalty function theorem [2, Proposition 2.4.3] and Robinson's error bound [19] for convex systems. The latter requires SCQ as a key condition. It is unclear if the exact penalization schemes can be derived through other avenues. For instance, Dentcheva and Ruszczyński [4] observed that first order optimality conditions of (2.5) may be established without SCQ. It might be interesting to explore whether this can be exploited to derive the error bound and exact penalization. We leave this for our future research.

3.3. Level function methods. Level function method is popular numerical approach for solving deterministic nonsmooth optimization problems. It is proposed by Lemaréchal et al [13] for solving nonsmooth convex optimization problems and extended by Xu [25] for solving quasiconvex optimization problems. Meskarian et al [16] recently applied a level function method to (3.1) where the distribution of ξ is discrete. In this subsection, we apply the level function method in [25] to problems (3.6) and (3.9).

Let $v : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function. Recall that the Clarke generalized derivative of v at a point x in direction d is defined as

$$v^o(x, d) := \limsup_{y \to x, t \downarrow 0} \frac{v(y + td) - v(y)}{t}.$$

The Clarke generalized gradient (also known as Clarke subdifferential) is defined as

$$\partial v(x) := \{\zeta : \zeta^T d \leq v^o(x, d)\}.$$

See [2, Chapter 2] for the details of the concepts. In the case when v is convex, the Clarke subdifferential coincides with the usual convex subdifferential [20]. It is well known [25] 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 (3.6) and (3.9). Let $\vartheta_{\bar{\rho}}(x)$ and $\hat{\vartheta}_{\hat{\rho}}(x)$ be defined as in (3.6) and (3.9) respectively. In the algorithm to be stated below, we need to calculate an element of the subdifferential of $\partial\vartheta_{\bar{\rho}}(x)$ and $\partial\hat{\vartheta}_{\hat{\rho}}(x)$ at each iteration.

ALGORITHM 3.1. *(Level function method for penalized problem (3.6) (or (3.9)))*
Step 1. Let $\epsilon > 0$ be a constant and select a constant $\tau \in (0,1)$ and a starting point $x_0 \in \mathcal{X}$; set $k := 0$.
Step 2. Calculate $\zeta_k \in \partial_x \vartheta_{\bar{\rho}}(x_k)$ (for problem (3.9), $\zeta_k \in \partial_x \hat{\vartheta}_{\hat{\rho}}(x_k)$). Set

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

and

$$\sigma_k(x) = \max\{\sigma_{k-1}(x), \sigma_{x_k}(x)\}$$

where $\sigma_{-1}(x) := -\infty$. Let

$$x_k = \operatorname{argmin}\{\vartheta(x_j) : j \in 0, \ldots, k\}$$

and

$$x_{k+1} \in \pi(x_k, Q_k),$$

where

$$Q_k = \{x \in \mathcal{X} : \sigma_k(x) \le -\tau\Delta(k)\},\ \Delta(k) = -\min_{x\in\mathcal{X}} \sigma_k(x),$$

and $\pi(x, Q_k)$ is the Euclidean projection of the point x on a set Q.
Step 3. If $\Delta(k) \le \epsilon$, stop; otherwise, set $k := k + 1$; go to step 2.

THEOREM 3.3. *Let $\{x_k\}$ be generated by Algorithm 3.1. Assume that $F(x,\xi)$ and $G(x,\xi)$ are Lipschitz continuous function 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,*

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

where Υ represents the diameter of X, ϵ and τ are given in Algorithm 3.1.

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 $\psi_j(x)$ with the same Lipschitz modulus $\mathbb{E}[L_G(\xi)]$. Along with the Lipschitzness of $F(x,\xi)$, this shows $\vartheta_{\bar{\rho}}(x)$ is Lipschitz continuous with modulus $\mathbb{E}[L_F(\xi)] + \bar{\rho}\mathbb{E}[L_G(\xi)]$. Similar conclusion can be drawn for $\hat{\vartheta}_{\hat{\rho}}(x)$. On the other hand, since $\vartheta_{\bar{\rho}}(x)$ and $\hat{\vartheta}_{\hat{\rho}}(x)$ are convex, then the function $\sigma_{x_k}(x)$ constructed at each iterate is a level function with modulus 1. The rest follows from Xu [25, Theorem 3.3]. □

In Algorithm 3.1, the penalty parameters in $\vartheta_{\bar{\rho}}(x)$ and $\hat{\vartheta}_{\hat{\rho}}(x)$ are fixed. In some cases, it might be difficult to compute/estimate these parameters. A simple way to tackle this issue is to start with an estimate of the penalty parameter and solve the resulting penalized problem with Algorithm 3.1. We then check the feasibility of the obtained solution: if it is infeasible, then increase the penalty parameter and repeat the process, otherwise it is an optimal solution. This kind of procedure in known as *Simple Penalty Function Method* in the literature of optimization, see for instance [24, Algorithm 10.2.3]. We describe the aforementioned procedure formally in the following algorithm for the penalized problem (3.6). Similar scheme can be applied to (3.9)).

ALGORITHM 3.2. *(Simple Penalty Function Method for penalized problem (3.6)).*

Step 1. Let $\bar{\epsilon}$ be a positive number. Let ρ_0 be an intial estimate of the penalty parameter. Set $t := 0$.

Step 2. For $\bar{\rho} := \rho_t$, apply Algorithm 3.1 to solve problem (3.6). Let x_t denote the solution obtained from solving the problem.

Step 3. If $\max_{j\in\{1,\ldots,N\}}(\sum_{i=1}^{N} p_i(Y_j - G(x_t,\xi^i))_+ - \gamma_j)_+ \le \bar{\epsilon}$, stop; otherwise, set $x_{t+1} := x_t$, $\rho_{t+1} := 10\rho_t$ and $t := t+1$, go to step 2.

Algorithm 3.2 terminates in a finite number of iterations in that the exact penalty parameters for problems (3.6) and (3.9) are finite, see Theorems 3.1 and 3.2.

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

$$\min_{x\in X} \max_{j\in\{1,\ldots,N\}} (\sum_{i=1}^{N} p_i(Y_j - G(x,\xi^i))_+ - \gamma_j). \tag{3.10}$$

This can be achieved by applying Algorithm 3.1 directly. The optimal value of (3.10) effectively gives an upper bound for parameters $\hat{\delta}$ and $\bar{\delta}$ (see the definition in Theorems 3.1 and 3.2). To see this argument clearly, we refer readers to Lemma 3.1. From the error bound established in the lemma and the penalty parameter estimated there, it is evident that a smaller δ gives a tighter error bound and hence a small lower bound for the penalty parameter. Problem (3.10) is to find an optimal interior point x_0 in the feasible set of (2.7) which minimizes the constant δ.

4. A modified cutting plane method. Rudolf and Ruszczyński [22] and Fábián et al [8] proposed cutting plane methods for solving stochastic programs with second order stochastic dominance constraints when the underlying random variables satisfy finite distribution. The methods are extension of the cutting-plane method developed by Haneveld and van der Vlerk in [11] for integrated chance constraints (ICC). Here we revisit the cutting-plane methods [22, 8] by considering a modification of the procedure where a cut is constructed.

The cutting plane methods in [8, 22] are essentially based on the reformulation of problem (2.5) as follows:

$$\begin{array}{ll} \min\limits_{x} & \sum\limits_{i=1}^{N} p_i F(x,\xi^i) \\ \text{s.t.} & \max\limits_{\mathcal{J}\subseteq\{1,\ldots,N\}} \sum\limits_{i\in\mathcal{J}} p_i(Y_j - G(x,\xi^i)) - \gamma_j \le 0, \forall j = 1,\ldots,N, \\ & x \in \mathcal{X}. \end{array} \tag{4.1}$$

At iteration t, the authors considered a collection of subsets (events)

$$\{\mathcal{J}_{j,t} \subseteq \{1,\ldots,N\} : j = 1,\ldots,N\},$$

which depend on the t-th iterate, denoted by x_t, and solve subproblem

$$\begin{array}{ll} \min\limits_{x} & \sum\limits_{i=1}^{N} p_i F(x,\xi^i) \\ \text{s.t.} & \sum\limits_{i\in\mathcal{J}_{j,l}} p_i(Y_j - G(x,\xi^i)) - \gamma_j \le 0, \text{ for } j = 1,\ldots,N \text{ and } l = 1,\ldots,t, \\ & x \in \mathcal{X}. \end{array} \tag{4.2}$$

Specifically, in [8], constraints

$$\sum_{i\in\mathcal{J}_{j,t}} p_i(Y_j - G(x,\xi^i)) - \gamma_j \le 0, \text{ for } j = 1,\dots,N \tag{4.3}$$

are added at iteration t (each of which corresponds to a cut) to the $N \times (t-1)$ constraints

$$\sum_{i\in\mathcal{J}_{j,l}} p_i(Y_j - G(x,\xi^i)) - \gamma_j \le 0, \text{ for } j = 1,\dots,N \text{ and } l = 1,\dots,t-1,$$

inherited from the previous iterations. Here $\mathcal{J}_{j,t}$ is the index set such that

$$\sum_{i\in\mathcal{J}_{j,t}} p_i(Y_j - G(x_t,\xi^i)) = \max_{\mathcal{J}\subseteq\{1,\dots,N\}} \sum_{i\in\mathcal{J}} p_i(Y_j - G(x_t,\xi^i)).$$

It is observed that such $\mathcal{J}_{j,t}$ can be identified as follows:

$$\mathcal{J}_{j,t} = \{i : Y_j - G(x_t,\xi^i) > 0, \text{ for } i = 1,\cdots,N\},$$

see a comment at page 45 in [8]. Note that Fábián, Mitra and Roman's [8] did not propose a detailed algorithm, instead, they observed that the cutting plane method due to Haneveld and Vlerk [11] can be easily applied to (4.1). To distinguish this method from the classical cutting method and Rudolf and Ruszczyński's cut generation method [22], we call it generalized cutting-plane method although this is fundamentally similar to the latter.

In [22], $\mathcal{J}_{j,t} = \mathcal{J}_{\hat{j},t}$ for $j = 1,\cdots,N$, where $\hat{j}$ is any $j \in \{1,\dots,N\}$ such that

$$\sum_{i=1}^{N} p_i(Y_j - G(x_t,\xi^i))_+ - \gamma_j > 0.$$

This differs from the previous approach because here the index set $\mathcal{J}_{\hat{j},t}$ is constructed for the $\hat{j}$-th constraint which is violated at x_t, rather than for every j-th constraint.

The focus of [8, 22] is on the case when F and G are linear functions of x and subsequently subproblem (4.2) is a linear programming problem. In the case when G is nonlinear w.r.t. x, (4.2) is a nonlinear program and therefore the approach is not a cutting plane method in the classical sense.

In what follows, we reformulate problem (2.7) as follows:

$$\begin{array}{ll} \min\limits_{x,y} & y \\ \text{s.t.} & \psi(x) := \max\limits_{j\in\{1,\dots,N\}} \psi_j(x) \le 0, \\ & \sum_{i=1}^{N} p_i F(x,\xi^i) - y \le 0, \\ & x \in \mathcal{X}, y \in \mathcal{Y}, \end{array} \tag{4.4}$$

where $\mathcal{Y}$ is a closed convex compact set such that

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

Existence of $\mathcal{Y}$ is due to the fact that $F(x,\xi^i)$, $i = 1,\cdots,N$, is a continuous function and $\mathcal{X}$ is a compact set. Note that $G(x,\xi)$ is concave and $F(x,\xi)$ is convex w.r.t. x,

which implies that $\psi(x)$ is convex w.r.t. x and $\sum_{i=1}^N p_i F(x,\xi^i) - y$ is convex w.r.t. (x,y). Moreover problem (4.4) is equivalent to problem (2.7). We apply the classical cutting-plane method [12] to both $\psi(x)$ and $\sum_{i=1}^N p_i F(x,\xi^i) - y$. Specifically, we propose the following algorithm.

ALGORITHM 4.1 (A modified cutting plane method). Define the current problem CP_t at iteration t as

$$\begin{array}{ll}\min\limits_{x,y} & y \\ \text{s.t.} & x \in \mathcal{X}, y \in \mathcal{Y} \\ & (x,y) \in S_t := \{(x,y) \in \mathcal{X}\times\mathcal{Y} : (a_l^{j_{l-1}^*})^T x \le b_l^{j_{l-1}^*}, d_l^T x + e_l y \le k_l, \quad l = 1,\dots,t,\},\end{array} \tag{4.5}$$

set $t := 0$, $S_0 := \mathcal{X}\times\mathcal{Y}$. For each t, carry out the following.

Step 1. Solve the LP problem (4.5) and let (x_t, y_t) denote the optimal solution. If problem (4.5) is infeasible, Stop: the original problem is infeasible.

Step 2. Find j_t^* such that

$$j_t^* = \mathrm{argmax}\{\psi_j(x_t), j = 1,\dots,N\}$$

and construct an index set

$$\mathcal{J}_t := \{i : Y_{j_t^*} - G(x_t,\xi^i) > 0\}.$$

Step 3. If

$$\sum_{i=1}^N p_i (Y_{j_t^*} - G(x_t,\xi^i))_+ - \gamma_{j_t^*} \le 0$$

and

$$\sum_{i=1}^N p_i F(x_t,\xi^i) - y_t \le 0,$$

stop: (x_t, y_t) is the optimal solution. Otherwise, construct feasible cuts

$$(a_{t+1}^{j_t^*})^T x \le b_{t+1}^{j_t^*}$$

and

$$d_{t+1}^T x + e_{t+1} y \le k_{t+1},$$

where

$$\begin{aligned} a_{t+1}^{j_t^*} &:= -\sum_{i\in\mathcal{J}_t} p_i \nabla_x G(x_t,\xi^i), \\ b_{t+1}^{j_t^*} &:= \sum_{i\in\mathcal{J}_t} p_i(-\nabla_x G(x_t,\xi^i)^T x_t + G(x_t,\xi^i) - Y_{j_t^*}) + \gamma_{j_t^*}, \\ d_{t+1} &:= \sum_{i=1}^N p_i \nabla_x F(x_t,\xi^i), \\ e_{t+1} &:= -1, \\ k_{t+1} &:= \sum_{i=1}^N p_i(\nabla_x F(x_t,\xi^i)^T x_t - F(x_t,\xi^i)). \end{aligned}$$

and set

$$S_{t+1} := S_t \cap \left\{(x,y) \in \mathcal{X} \times \mathcal{Y} : (a_{t+1}^{j_t^*})^T x \le b_{t+1}^{j_t^*}, d_{t+1}^T x + e_{t+1} y \le k_{t+1}\right\}.$$

Proceed with iteration $t+1$.

REMARK 4.1. We make a few comments on Algorithm 4.1.

(i) The main difference between Algorithm 4.1 and the cutting plane methods in [22] is in the way how feasible cuts are constructed. In [22], N constraints/cuts are added at iteration t, see (4.3). These constraints/cuts are not necessarily the extreme support (tangent plane) of $\psi(x)$ at x_t. In Algorithm 4.1, we exclude all those non-support constraints because we believe a cut at the extreme support (to $\psi(x)$ at x_t) is most effective and a single linear cut is adequate to ensure the convergence as we will demonstrate in Theorem 4.1; all other non-support constraints/cuts may potentially reduce numerical efficiency. Our approach is similar to the algorithm proposed by Fábián, Mitra and Roman, see the algorithm at page 48 in [8]. Note that Fábián, Mitra and Roman's algorithm is applied to linear models while Algorithm 4.1 is applicable to nonlinear case. Therefore we may regard the latter as an extension of the former.

(ii) In order to apply Algorithm 4.1 to problem (4.4), we need to identify index j_t^* where

$$j_t^* = \text{argmax}\{\psi_j(x_t) : j = 1, \ldots, N\}$$

at iteration t. This requires to evaluate $\psi_j(x_t)$ for $j \in \{1, \ldots, N\}$ and identify the index corresponding to the maximum. The procedure is also needed to verify the feasibility of x_t.

(iii) At Step 2, if there exists more than one index j_t^* such that the maximum is achieved, then we just pick up anyone of them. In such a case, the graph of function $\psi(x)$ has a kink at x_t^*. Our algorithm requires to construct a support plane to one of the active piece (note that $\psi(x)$ is piecewise smooth) and such a support plane is also a support to $\psi(x)$ at x_t.

(iv) When F is linear w.r.t. x, there is no need to introduce additional variable y because the objective is linear.

(v) Note that although problem (2.5) does not satisfy the SCQ, under some moderate conditions, problem (2.7) may satisfy it (Theorem 2.1). Moreover, since each subproblem (4.5) is a relaxation of problem (2.7), it also satisfies the SCQ.

The following theorem states convergence of the algorithm.

THEOREM 4.1. *Let $\{(x_t, y_t)\}$ be a sequence generated by Algorithm 4.1. Let*

$$S := \{(x,y) \in \mathcal{X} \times \mathcal{Y} : \psi(x) \le 0, \mathbb{E}[F(x,\xi)] - y \le 0\} \subset \mathcal{X} \times \mathcal{Y},$$

where $\psi(x)$ is defined in problem (4.4) and ξ has finite support set. Assume: (a) $F(x,\xi)$ is continuously differentiable and convex and $G(x,\xi)$ is continuously differentiable and concave w.r.t. x for almost every ξ; (b) $\mathcal{X} \times \mathcal{Y} \in \mathbb{R}^n$ is a compact set; (c) there exists a positive constant L such that the Lipschitz modulus of $\mathbb{E}[F(x,\xi)]$ and $\psi(x)$ are bounded by L on $\mathcal{X} \times \mathcal{Y}$; (d) S is nonempty. Then $\{(x_t, y_t)\}$ contains a subsequence which converges to a point $(x^*, y^*) \in S$, where (x^*, y^*) is the optimal solution and y^* is the optimal value of (4.4).

Proof. The proof is similar to [12, Theorem]. Note that in every iteration $t > 0$, $a_{t+1}^{j_t^*} \in \partial_x \psi(x_t)$, $d_{t+1} = \nabla \mathbb{E}[F(x_t, \xi)]$ and $e_{t+1} = \nabla_y(\mathbb{E}[F(x_t, \xi)] - y_t) = -1$. Then $(a_{t+1}^{j_t^*})^T x - b_{t+1}^{j_t^*}$ and $d_{t+1}^T x + e_{t+1} y - k_{t+1}$ are the extreme support to the graphs of $\psi(x)$ and $\mathbb{E}[F(x, \xi)] - y$ at point (x_t, y_t) respectively. By condition (a), $\psi(x)$ and $\mathbb{E}[F(x, \xi)] - y$ are convex and continuous functions w.r.t. (x, y). Thus, if $(x, y) \in S$ and $\max(\psi(x), \mathbb{E}[F(x, \xi)] - y) \leq 0$, then

$$\max\left((a_{t+1}^{j_t^*})^T x - b_{t+1}^{j_t^*}, d_{t+1}^T x + e_{t+1} y - k_{t+1}\right) \leq 0.$$

On the other hand, for $(x_t, y_t) \notin S$,

$$\max\{(a_{t+1}^{j_t^*})^T x_t - b_{t+1}^{j_t^*}, d_{t+1}^T x_t + e_{t+1} y_t - k_{t+1}\} = \max\{\psi(x_t), \mathbb{E}[F(x_t, \xi)] - y_t\} > 0.$$

Thus, when $(x_t, y_t) \notin S$, the set S and the point (x_t, y_t) lie on opposite sides of the cutting angle $\max\{(a_{t+1}^{j_t^*})^T x - b_{t+1}^{j_t^*}, d_{t+1}^T x + e_{t+1} y - k_{t+1}\} = 0$.

Note that from the definition of S_t and (x_t, y_t), we know that $S \subset S_t \subset S_{t-1}$, (x_t, y_t) minimizes y in S_t and $y_{t-1} \leq y_t$. In the case when $(x_t, y_t) \in S$, it is easy to verify that (x_t, y_t) is the optimal solution of problem (4.4). Indeed, since (x_t, y_t) is an optimal solution, for every $(x, y) \in S_t$, we have $y \geq y_t$. Since $S \subset S_t$, then $y \geq y_t$ for $(x, y) \in S$, which implies optimality of (x_t, y_t) over S.

In the rest of proof, we focus on the case when $(x_t, y_t) \notin S$ for all t. Since $\mathcal{X} \times \mathcal{Y}$ is a compact set, the sequence $\{(x_t, y_t)\}$ contains a subsequence which converges to $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$. Assume without loss of generality that $(x_t, y_t) \to (x^*, y^*)$. Let $S^* = \cap_t S_t$. Since S_t is compact and $S \subset S_t$, we have $S \subset S^*$ and $(x^*, y^*) \in S^*$. On the other hand, since

$$y \geq y_t, \ \forall (x, y) \in S_t,$$

then

$$y \geq y^*, \ \forall (x, y) \in S^*. \tag{4.6}$$

Indeed, if this is not true, then there exists $(\hat{x}, \hat{y}) \in S^*$ such that $\hat{y} < y^*$. Since $y_t \to y^*$, there exists some sufficiently large t such that $\hat{y} < y_t$.

This is not possible because (x_t, y_t) is an optimal solution in S_t while $(\hat{x}, \hat{y}) \subset S^* \subset S_t$ is a feasible solution! This shows (4.6) holds. Since $S \subset S^*$, the inequality also holds for all $(x, y) \in S$, which implies (x^*, y^*) is an optimal solution of problem (4.4) if $(x^*, y^*) \in S$.

In what follows, we want to show that $(x^*, y^*) \in S$. Observe that (x_t, y_t) minimizes y in S_t, that is, it satisfies the inequalities:

$$(a_{l+1}^{j_l^*})^T x - b_{l+1}^{j_l^*} \leq 0, \tag{4.7}$$

and

$$d_{l+1}^T x + e_{l+1} y - k_{l+1} \leq 0, \tag{4.8}$$

for $l = 0, \ldots, t-1$ and by condition (c), $\max\{||(a_{l+1}^{j_l^*})||, ||d_{l+1}||\} \leq L$ for all l. For the simplicity of notation, let $\{x_t, y_t\}$ denote the subsequence. We claim that $\{\max(\psi(x_t), \mathbb{E}[F(x_t, \xi)] - y_t)\}$ converges to 0. Note that since

$$\begin{array}{rcl} b_{l+1}^{j_l^*} & = & \sum_{i \in \mathcal{J}_l} p_i(-\nabla_x G(x_l, \xi^i)^T x_l + G(x_l, \xi^i) - Y_{j_l^*}) + \gamma_{j_l^*} \\ & = & (a_{l+1}^{j_l^*})^T x_l - \psi_{j_l^*}(x_l) \\ & = & (a_{l+1}^{j_l^*})^T x_l - \psi(x_l), \end{array}$$

then (4.7) implies

$$\psi(x_l) + (a_{l+1}^{j_l^*})^T(x - x_l) \leq 0.$$

Likewise, by the definition of e_{l+1}, k_{l+1}, we have from (4.8) that

$$\mathbb{E}[F(x_l, \xi)] + d_{l+1}^T(x - x_l) - y \leq 0.$$

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

$$\begin{aligned} r &\leq \max(\psi(x_l), \mathbb{E}[F(x_l, \xi)] - y_l) \\ &\leq \max((a_{l+1}^{j_l^*})^T(x_l - x_t), d_{l+1}^T(x_l - x_t) - (y_l - y_t)) \\ &\leq (L+1)||(x_l, y_l) - (x_t, y_t)||, \end{aligned}$$

for all $0 \leq l \leq t-1$, which means $\{(x_t, y_t)\}$ does not converge, a contradiction! This shows $\{\max(\psi(x_t), \mathbb{E}[F(x_t, \xi)] - y_t)\}$ converges to 0 and hence $(x^*, y^*) \in S$ is the optimal solution and y^* is the optimal value of (4.4). □

5. Numerical tests. In this section, we investigate the numerical performance of Algorithms 3.1 and 4.1 along with the generalized cutting-plane method. We do so by applying them to an academic problem, a practical portfolio selection problem and a supply chain problem.

The tests are carried out in MATLAB 7.10 installed on a HP Notebook PC with Windows 7 operating system, Intel Core i7 processor. We use IBM ILOG CPLEX Studio 12.4 to solve the subproblems within the Algorithm 3.1 and the Algorithm 4.1, while the subproblem within the generalized cutting-plane method is solved by "fmincon" due to nonlinearity. Furthermore, Algorithm 3.2 is integrated in the Algorithm 3.1, to find a suitable penalty parameter. The initial penalty parameter is set as 100. For Algorithm 3.1, we use the stopping criteria parameters $\epsilon = 10^{-4}$ and $\tau = 0.5$. Moreover, the Algorithm 4.1 and the generalized cutting-plane method proposed in [8] terminates when the solution of any subproblem become a feasible point of the original problem, see Step 3 of Algorithm 4.1 for details.

EXAMPLE 5.1. Consider problem (2.5) with $F(x, \xi) = -x\xi$, $G(x, \xi) = x\xi - \frac{1}{2}x^2$, $Y(\xi) = G(1, \xi)$, $\mathcal{X} = [0, 50]$, where ξ is a random variable with finite distribution $P(\xi = 2 + \frac{i-1}{N}) = \frac{1}{N}$ for $i = 1, \cdots, N$ and $N = 101$. The problem can be specifically presented as:

$$\begin{aligned} \min_x \quad & -\frac{1}{101}\sum_{i=1}^{101} x(2 + \frac{i-1}{101}), \\ \text{s.t.} \quad & \frac{1}{101}\sum_{i=1}^{101}(\eta - x(2 + \frac{i-1}{101}) + \frac{1}{2}x^2)_+ \leq \frac{1}{101}\sum_{i=1}^{101}(\eta - (2 + \frac{i-1}{101}) + \frac{1}{2})_+, \forall \eta \in \mathbb{R}, \\ & x \in \mathcal{X}. \end{aligned} \tag{5.1}$$

It is difficult to work out the feasible set precisely. Here we only need to find out the optimal solution of problem (5.1). For $x \in [1, 3]$,

$$P(G(x, \xi) \leq \eta) \leq P(Y(\xi) \leq \eta), \text{for all } \eta \in \mathbb{R},$$

which means $G(x, \xi) \succeq_1 G(1, \xi)$, and hence $G(x, \xi) \succeq_2 G(1, \xi)$. When $x > 3$,

$$\sum_{i=1}^{N} p_i(1.5 - G(x, \xi^i))_+ > \sum_{i=1}^{N} p_i(1.5 - Y(\xi^i))_+,$$

which implies $G(x,\xi) \not\succeq_2 G(1,\xi)$. This shows that any point in the interval $[1,3]$ is a feasible point of problem (5.1) whereas any point x with $x > 3$ is infeasible. It is easy to see that $x^* = 3$ is the optimal solution (with corresponding optimal value -7.5) because the objective function is linear w.r.t. x.

We apply the L_1-norm based penalty scheme (3.6) and L_∞-norm based penalty scheme (3.9) to problem (5.1) respectively. To justify the application, we examine the SCQ of problem (5.1) and estimate the penalty parameter ρ. Consider formulation (2.7) for problem (5.1). Let $\underline{y} := \min_{i=1}^N Y(\xi^i)$. It is easy to show that $\underline{y} = Y(2)$. For $x_0 = 2$, it is a feasible point. Moreover, $G(x_0,\xi) \succeq_1 Y(\xi)$, and $G(x_0,\overline{\xi}^i) > \underline{y}$ for all $i = 1,\cdots,N$. This verifies the conditions of Theorem 2.1. Hence formulation (2.7) of problem (5.1) satisfies the SCQ.

Next we estimate the penalty parameter ρ through Theorem 3.1. We need to work out κ, δ and D defined in Lemma 3.1. Observe first that the objective function of problem (5.1) is Lipschitz continuous function with modulus $\kappa = 3$. Let

$$\delta := -\max_{j\in\{1,\ldots,N\}} \left(\frac{1}{N} \max_{\mathcal{J}\in\mathcal{N}} \sum_{i\in\mathcal{J}} (Y_j - G(x_0,\xi^i)) - \frac{1}{N}\sum_{i=1}^N (Y_j - Y(\xi^i))_+ \right).$$

It is easy to calculate that $\delta = 0.005$. On the other hand, it is easy to verify that the feasible set of problem (5.1) is contained in $[0,3]$. Let $D = 3$. We obtain an estimate of penalty parameter ρ, that is, $\rho > \kappa\delta^{-1}D = 1800$.

We have carried out numerical tests on four algorithms for problem (5.1): Algorithm 3.1 based on exact penalization with L_1-norm (Algorithm 3.1 (L_1) for short), Algorithm 3.1 based on exact penalization with L_∞-norm (Algorithm 3.1 (L_∞) for short), Algorithm 4.1 and the generalized cutting-plane method.

The numerical results are displayed in Table 5.1. A few words about the notation. *Opt.Sol* denotes the numerical optimal solution and *Opt. Val* denotes the corresponding optimal value. To check the efficiency of the algorithms, we have recorded the CPU time (in minutes) for each of the algorithms.

TABLE 5.1
Numerical results for (5.1), Example 5.1.

	CPU Time	*Opt.Sol*	*Opt. Val*
Algorithm 3.1 (L_1)	0.0374	3.000	-7.4850
Algorithm 3.1 (L_∞)	0.0148	3.000	-7.4850
Generalized cutting-plane method	0.0463	3.003	-7.5099
Algorithm 4.1	0.0196	3.000	-7.4850

The results show that all four algorithms perform efficiently albeit Algorithm 3.1 (L_1) and generalized cutting-plane method takes slightly more CPU time. To further investigate the performance of the algorithms, we propose to run a portfolio optimization problem of larger size.

EXAMPLE 5.2. Consider a portfolio optimization problem with nonlinear transaction cost where short selling is allowed. Let $r_j : \Omega \to \mathbb{R}$, denote the random return rate of asset j for $j = 1,\cdots,n$, and $R := (r_1,\ldots,r_n)$. We assume that $\mathbb{E}[|r_j|] < \infty$. Denoting by x_j the fraction of the initial capital invested in asset j, we can easily derive the total return rate:

$$R(\omega)^T x = r_1(\omega)x_1 + r_2(\omega)x_2 + \cdots + r_n(\omega)x_n,$$

and the total return with transaction costs:

$$G(x,R) := R(\omega)^T x - 0.01 \times (|x_1| + |x_2| + \cdots + |x_n|), \tag{5.2}$$

where the number 0.01 is the unit transaction cost. Here x_i may take a negative value due to the practice of short selling which allows an investor to borrow some shares of stock from a broker and sell them. The investor must return the same shares by purchasing them back from the market over a specified time period (in hope that the share price would drop). This is also known as covering the short position. See [1].

Our aim is to investigate the optimal investment policy for a fixed capital in the n assets with some desirable characteristics of the portfolio using second order stochastic dominance model.

Let $G(x,R)$ be defined as in (5.2), associated with the decision vector $x \in \mathcal{X} \subset \mathbb{R}^n$, where $\mathcal{X}$ is a compact and convex set. In the literature of portfolio optimization, x is known as a portfolio and $\mathcal{X}$ is the set of feasible portfolios.

Let $Y(R)$ denote the benchmark return. Our aim is to find an optimal portfolio x such that $G(x,R)$ dominates $Y(R)$ in second order and $\mathbb{E}[G(x,R)]$ is maximized. We formulate the decision making problem as a stochastic minimization problem with second order dominance constraints, namely (1.4), with $F(x,\xi) = -G(x,R)$, $G(x,\xi) = G(x,R)$ and $Y(\xi) = Y(R)$. The main issue here is that we don't know the distribution of random vector $R(\omega)$. To get around the difficulty, we use historical data (sample) and consider the following approximation problem:

$$\begin{array}{ll} \min\limits_{x} & -\dfrac{1}{N}\displaystyle\sum_{i=1}^{N}(x^T R^i - 0.01\|x\|_1) \\ \text{s.t.} & \dfrac{1}{N}\displaystyle\sum_{i=1}^{N}\big((Y_j - x^T R^i + 0.01\|x\|_1)_+ - (Y_j - Y_i)_+\big) \le 0, \quad j = 1,\ldots,N, \\ & x \in \mathcal{X}, \end{array} \tag{5.3}$$

where $\|x\|_1 = |x_1| + \ldots + |x_n|$, $\{R^1,\ldots,R^N\}$ is a set of historical data and $\{Y_i\}_{i=1}^N$ is the set of the benchmark return.

Problem (5.3) is also known as a sample average approximation of problem (1.4). Taking into account the short selling policy, we set an upper bound 2 and a lower bound -1 for each x_i, consequently

$$\mathcal{X} = \{x : \sum_{i=1}^{n} x_i = 1, x_i \in [-1,2], \text{for } i = 1,\cdots,n\}.$$

Note that the problem (5.3) is different from the portfolio optimization problem investigated in [8, 26] in that the underlying function $G(x,R)$ is nonlinear w.r.t. x.

We have carried out numerical tests on the four algorithms for problem (5.3). We have collected 300 daily returns of 93 FTSE100 assets prior to March 2011. The first 200 of them are used to construct the portfolio strategy and set up a backtest in order to investigate the performance of the portfolio in-sample. The remaining 100 are used to carry out an out-of-sample test to examine the efficiency of the portfolio out-of-sample.

Table 5.2 displays the test results of the four algorithms in terms of CPU time and optimal return based on the first 200 data. The four algorithms generate almost identical optimal portfolio and return. The numerical efficiency, however, seems different in terms of CPU time: generalized cutting-plane method apparently requires

more computational time. This is because the subproblem of generalized cutting-plane method is a nonlinear program in this example and the MATLAB optimization solver "fmincon" cannot solve them efficiently particularly when the problem size is large.

TABLE 5.2
Numerical results for (5.3), Example 5.2.

	CPU Time	*Opt.Return*
Algorithm 3.1 (L_1)	1.3474	0.0057
Algorithm 3.1 (L_∞)	1.0084	0.0057
Generalized cutting-plane method	35.4953	0.0057
Algorithm 4.1	1.4439	0.0057

In order to investigate the performance of the selected portfolio, we have carried out a backtest and an out-of-sample test for each of the algorithms. Figures 5.1-5.2 depict the cumulative distribution functions (CDF) of the returns based on optimal portfolios obtained through Algorithm 3.1 (L_1) and Algorithm 3.1 (L_∞) respectively, in comparison with the benchmark return (FTSE100 Index). The performance of the portfolio strategies generated by the generalized cutting-plane method and Algorithm 4.1 are displayed in Figures 5.3-5.4.

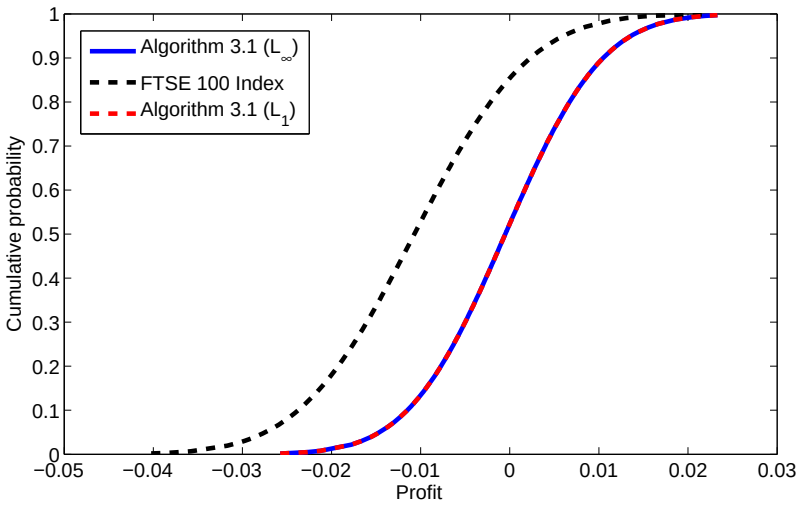

FIG. 5.1. *The cumulative distribution functions of Algorithm 3.1 (L_1) and Algorithm 3.1 (L_∞) in comparison with FTSE100 index for backtest.*

From Figures 5.1-5.4, it is easy to observe that the returns of optimal portfolio constructed through the 4 algorithms are almost identical and they dominate the benchmark return in first order (and hence in second order) in both backtest and out-of-sample test.

Table 5.3 displays the returns and risks based on the selected portfolios and the benchmark one both in-sample and out-of-sample. Here we use the Value-at-Risk (VaR) as the risk measure, which is one of the most commonly used measures of risk

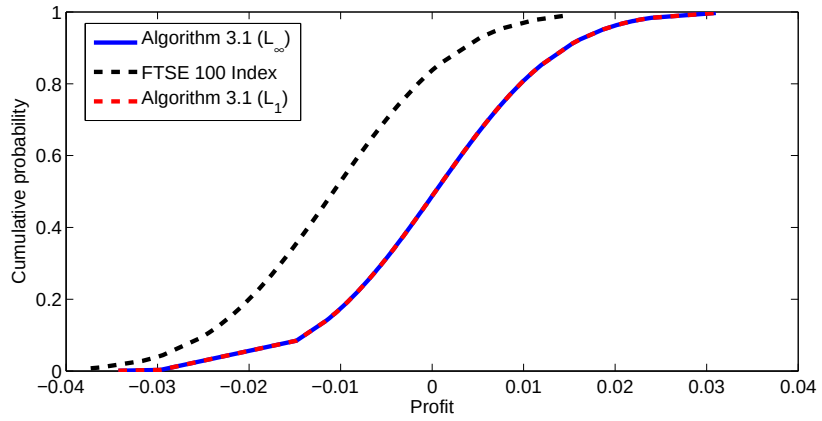

FIG. 5.2. *The cumulative distribution functions of Algorithm 3.1 (L_1) and Algorithm 3.1 (L_∞) in comparison with FTSE100 index for out-of-sample test.*

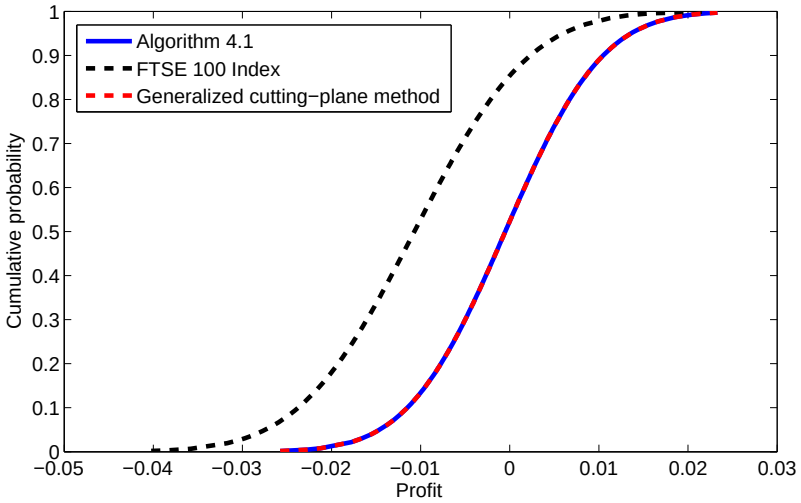

FIG. 5.3. *The cumulative distribution functions of the optimal portfolio returns of generalized cutting-plane method and Algorithm 4.1 in comparison with FTSE100 index for Backtest.*

in finance. It is defined as

$$\mathrm{VaR}_\alpha(-G(x,R)) := \min_{\eta\in\mathbb{R}}\{\eta : \mathrm{Prob}\{-G(x,R)\le\eta\}\ge\alpha\},$$

where $\alpha\in(0,1)$ and $-G(x,R)$ is the loss function. In this context, the formulation above means that with the probability less than $1-\alpha$, the loss $-G(x,R)$ will be greater than $\mathrm{VaR}_\alpha(-G(x,R))$ or equivalently, the return $G(x,R)$ will be less than $-\mathrm{VaR}_\alpha(-G(x,R))$. For a fixed α, a smaller $\mathrm{VaR}_\alpha(-G(x,R))$ means smaller risk. Three values of α are commonly considered: $0.90, 0.95, 0.99$. In our analysis, we

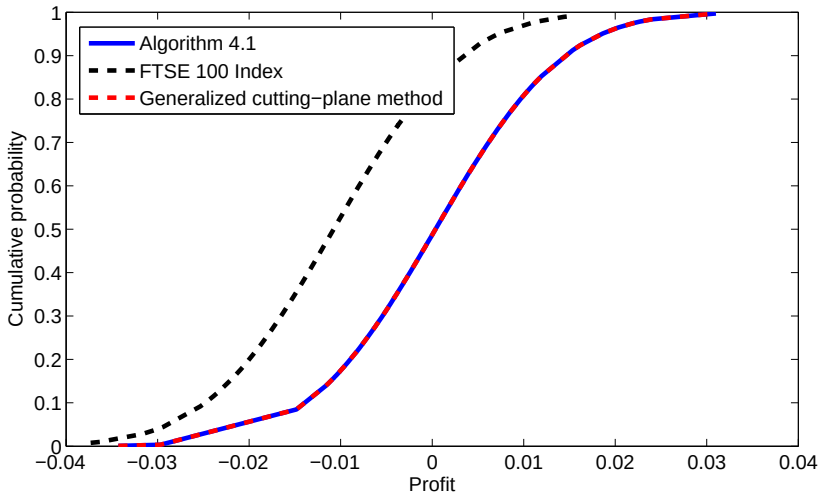

FIG. 5.4. *The cumulative distribution functions of the optimal portfolio returns of generalized cutting-plane method and Algorithm 4.1 in comparison with FTSE100 index for Backtest.*

consider $\alpha = 0.95$. From Table 5.3, we can see that the optimally selected portfolios generate higher returns with lower risks in comparison with those of the benchmark portfolio both in-sample and out-of-sample. We have also carried out some sensitivity

TABLE 5.3
Comparison of the selected portfolio to the benchmark portfolio, Example 5.2.

Data	Portfolio	Return	VaR
In-sample	Selected portfolio	0.0057	-0.0252
	Benchmark portfolio	0.0031	0.0008
Out-of-sample	Selected portfolio	0.0036	-0.0103
	Benchmark portfolio	-0.0071	0.0023

analysis of the four algorithms with respect to the change of problem size and number of scenarios. Figure 5.5 depicts CPU time of the four algorithms as the number of assets increase from 10 to 2500. It shows that Algorithm 3.1 (L_∞) and Algorithm 4.1 require considerably less CPU times as the number of assets increases. The underlying reason that Algorithm 3.1 (L_∞) outperforms Algorithm 3.1 (L_1) is that the former requires to calculate a subgradient of a single nonsmoth function while the latter requires to calculate a subgradient of the sum of N nonsmooth functions which takes more times as problem size increases.

Although generalized cutting-plane method is proposed to solve linear problems and in this nonlinear setting, it performs reasonably well with respect to large number of assets considered. Figure 5.6 displays similar phenomena. As the size of scenarios increases, generalized cutting-plane method have more nonlinear constraints while Algorithm 3.1 (L_1) takes more time to calculate a subgradient. There seems to be no significant impact on the other two algorithms. Since the difference between Algorithm 3.1 (L_∞) and Algorithm 4.1 is not clear in Figure 5.6, we enlarge them in

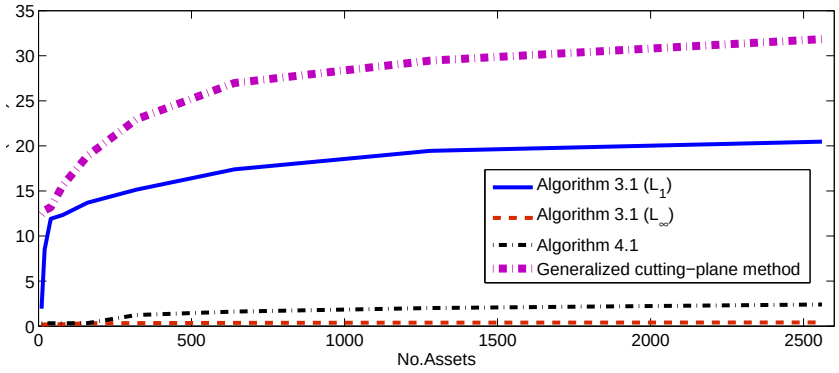

FIG. 5.5. *Computational time versus the number of assets for a fixed number of observations, Example 5.2.*

Figure 5.7.

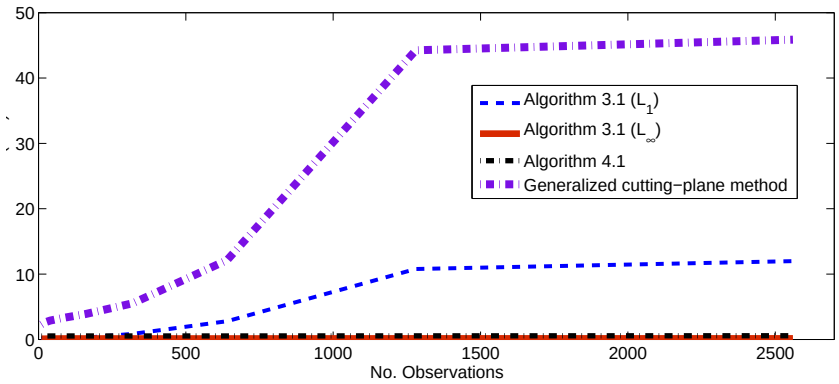

FIG. 5.6. *Computational time versus the number of observations for a fixed number of assets, Example 5.2.*

It is important to note, as one of the referees suggested, that the out-of-sample experiments implicitly assume the empirical samples approximate the true distribution of the return and there is a limitation of such an approach when sample size is small. It might be interesting to discuss the reliability of such approximation in terms of sample size but this is beyond the scope of this paper. We refer interested readers to the asymptotic analysis of the sample average approximation for second order dominance problems in [10] and [23].

The last example we have tested is a supply chain problem recently considered by Dentcheva and Martinez [3, Section 4].

EXAMPLE 5.3. *(Dentcheva and Martinez [3])* A company has a set F of factories that produce and supply perishable product to a set O of stores. Assume that demand at each store is stochastic and the goods are supplied before the demand is observed. If the demand is not met, the customers buy the product elsewhere and the sales are

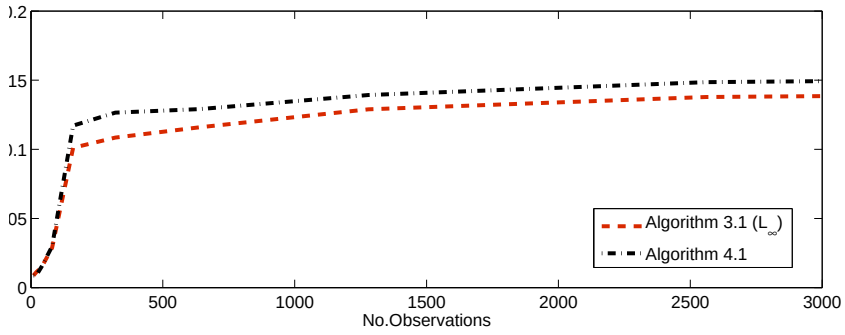

FIG. 5.7. *Computational time versus the number of observations for a fixed number of assets, Example 5.2.*

lost. If the stock of the store is larger than the demand, then the remaining products need to be disposed of. Assume that the disposal cost is a deterministic quantity and that each factory has a limited capacity to produce goods. Furthermore, we assume that a benchmark of the acceptable cost distribution is available. The objective is to determine a production and shipping plan for each factory in order to minimize the expected cost of the company. Denote by x_{ij} the quantity of goods delivered by factory i to store j, D_j the demand at store j, w_i the quantity produced by factory i, z_j the number of sales at store j, and y_j the amount disposed at store j. The shipping cost from factory i to store j is denoted by c_{ij}; the production cost of one unit of product at store i is a_i; the capacity of store i is C_i. The disposal cost on site j is d_j and the price store j sets for the product is b_j.

Dentcheva and Martinez proposed a two-stage stochastic program with stochastic ordering constraint model for this problem, see [3, Section 4] for details. Here, we formulate the problem as a one-stage stochastic problem with SSD constraints as follows:

$$\begin{array}{ll} \min & \displaystyle\sum_{i=1}^{F}\sum_{j=1}^{O} c_{ij}x_{ij} + \sum_{i=1}^{F} a_i w_i + \sum_{s=1}^{S} p_s Q(x, D^s) \\ \text{s.t.} & -Q(x, D) \succeq_2 -Y, \\ & \displaystyle w_i = \sum_{j=1}^{O} x_{ij}, i = 1, \ldots, F, \\ & 0 \le w_i \le C_i, i = 1, \ldots, F, \\ & x_{ij} \ge 0, i = 1, \ldots, F; j = 1, \ldots, O. \end{array} \tag{5.4}$$

Here $D := (D_1, \ldots, D_O)$ where each D_j has S scenarios,

$$Q(x, D^s) := \sum_{j=1}^{O} (d_j(x_j - D_j^s)_+ - b_j(D_j^s - (D_j^s - x_j)_+)),$$

$x_j = \sum_{i=1}^{F} x_{ij}$ and Y is a benchmark with S scenarios.

We assume that each component D_j of the demand D satisfies the γ-distribution with parameters $(2, 3)$, $j = 1, \ldots, J$. Each data set is generated through i.i.d. sampling with size S and $p_s := \frac{1}{S}$, for $s = 1, \ldots, S$. The benchmark is constructed from

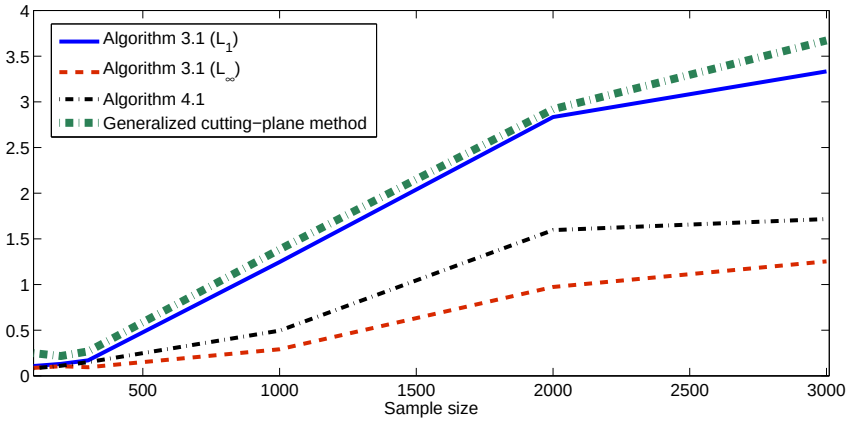

FIG. 5.8. *Computational time versus the number of observations for* $F = 10$ *and* $O = 10$*, Example 5.3.*

solving the problem without SSD constraints. We have carried out a number of numerical tests on the four algorithms for problem (5.4). Our tests are concentrated on CPU time against increment of sample size and problem size. There are two cases which may lead to the change of problem size: increase of the number of factories F and increase of the number of stores O. Figure 5.8 depicts CPU time of the four algorithms as the size increases from 10 to 3000, whereas Figures 5.9 and 5.10 depict the sensitivity of CPU time against the change of F and O respectively.

We have made a few observations from the numerical tests. First, the four algorithms perform well as sample size increases particularly when the size goes beyond 2000, see Figure 5.8. This is primarily because the sample size does not increase the size of problem (5.4) albeit it increases the number of terms in both the objective and constraint functions. Second, Algorithm 3.1 (L_∞) performs better than than other algorithms in most cases. An underlying reason is that it constructs a single level function instead of adding two or more cutting planes as in the Algorithm 4.1 and the generalized cutting-plane method. Similar comments apply to L_1 penalization method (3.6) although it is not as efficient as Algorithm 3.1 (L_∞). Third, increase of the number of stores O has more significant impact on on both the number of iterations and the CPU time in all of the four algorithms than does the increase of the number of factories F, see Figures 5.9 and 5.10. This is primarily due to the fact that increasing O results in more nonsmooth terms in the constraint functions while increasing F does not have such consequence, see the composition of $Q(x, D^s)$. Finally, in comparison with Dentcheva and Martinez's test results, our algorithms are less sensitive to the increase of sample size because we don't introduce new variables to deal with plus functions, on the other hand, our algorithm are more sensitive to the increase of O whereas their algorithms deal with such a problem through introduction of a new variable per scenario to get around the nonsmoothness of the plus function.

Acknowledgments. The authors would like to thank Casba Fábián for his comments on an earlier version of the paper which helped us clarify the cutting plane algorithms in [8]. They are also grateful to the two anonymous referees and the associate editor, Darinka Dentcheva, for insightful comments which helped us improve the

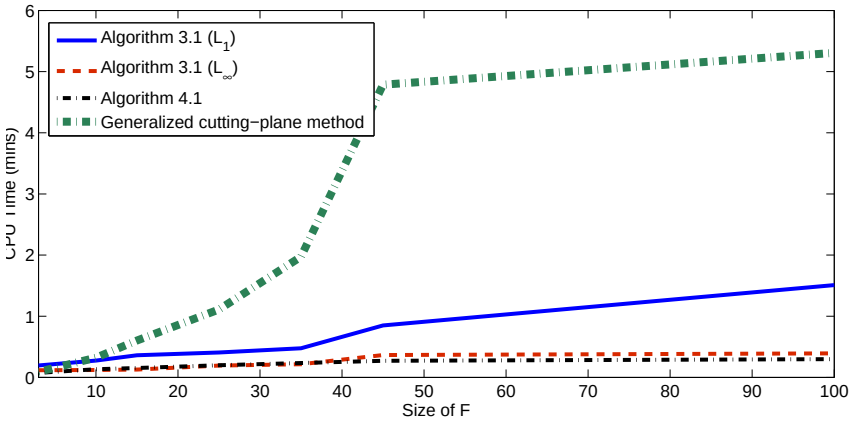

FIG. 5.9. *Computational time versus the number of factories F with $O = 10$ and $S = 300$, Example 5.3.*

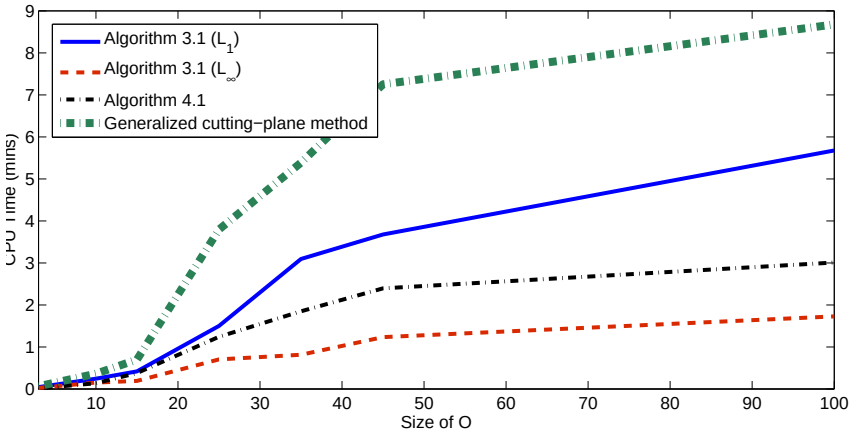

FIG. 5.10. *Computational time versus the number of stores O, with $F = 10$ and $S = 300$, Example 5.3.*

presentation of the paper. Finally they thank Darinka Dentcheva and Gabriela Martinez for providing us with reference [3] and subsequent discussions about Example 5.3.

REFERENCES

[1] Z. BODIE, A. KANE AND A. MARCUS, *Investments*, 6th. ed., McGraw-Hill, Now York, 2005.
[2] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
[3] D. DENTCHEVA AND G. MARTINEZ, Two-stage stochastic optimization problems with stochastic ordering constraints on the recourse, *European J. Oper. Res.*, 219 (2012), pp. 1–8.
[4] D. DENTCHEVA AND A. RUSZCZYŃSKI, Optimization with stochastic dominance constraints, *SIAM J. Optim.*, 14 (2003), pp. 548–566.
[5] ———, Portfolio optimization with stochastic dominance constraints, *J. Banking Financ*, 30

(2006), pp. 433–451.
[6] ———, Optimality and duality theory for stochastic optimization with nonlinear dominance constraints, *Math. Program.*, 99 (2004), pp. 329–350.
[7] Y. Ermoliev, Stochastic quasi-gradient methods and their application to system optimization, *Stochastics,* 9 (1983), pp. 1–36.
[8] C. Fábián, G. Mitra and D. Roman, Processing second-order stochastic dominance models using cutting-plane representations. *Math. Program.*, 130 (2011), pp. 33–57.
[9] T. Homem-de-Mello and S. Mehrotra, A cutting surface method for uncertain linear programs with polyhedral stochastic dominance constraints, *SIAM J. Optim.*, 20 (2009), pp. 1250-1273.
[10] J. Hu, T. Homen-De-Mello and S. Mehrotra, Sample average approximation of stochastic dominance constrained programs, *Math. Program.*, 133 (2012), pp. 171–201.
[11] W. Klein Haneveld and M. van der Vlerk, Integrated chance constraints: reduced forms and an algorithm, *Comput. Manag. Sci.,* 3 (2006), pp. 245–269.
[12] J. E. Kelley, The cutting-plane method for solving convex programs, *SIAM J. Appl. Math.*, 8 (1960), pp. 703-712.
[13] C. Lemarechal, A. Nemirovskii and Y. Nesterov, New variants of bundle methods, *Math. Program.*, 69 (1995), pp. 111–147.
[14] Y. Liu and H. Xu, *Stability and senstivity analysis of stochastic programs with second order dominance constrains*, *Mathematical Programming, Series A*, online publication 2012.
[15] R. Meskarian, J. Fliege and H. Xu, *Stochastic programming with multivariate second order stochastic dominance with applications in portfolio optimization*, Preprint, School of Mathematics, University of Southampton, June 2012.
[16] R. Meskarian, H. Xu and J. Fliege, Numerical methods for stochastic programs with second order dominance constraints with applications to portfolio optimization, *European J. Oper. Res.*, 216 (2011), pp. 376–385.
[17] A. Müller and M. Scarsini, Eds., *Stochastic Orders and Decision under Risk, Institute of Mathematical Statistics*, Hayward, CA, 1991.
[18] W. Ogryczak and A. Ruszczyński, From stochastic dominance to mean-risk models: Semideviations as risk measures, *European J. Oper. Res.*, 116 (1999), pp. 33–50.
[19] S. M. Robinson, An application of error bounds for convex programming in a linear space, *SIAM J. Control Optim.*, 13 (1975), pp. 271–273.
[20] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1982.
[21] D. Roman and G. Mitra, Portfolio selection models: A review and new directions, *Wilmott Journal*, 1 (2009), pp. 69–85.
[22] G. Rudolf and A. Ruszczyński, Optimization problems with second order stochastic dominance constraints: duality, compact formulations, and cut generation methods, *SIAM J. Optim.*, 19 (2008), pp. 1326–1343.
[23] H. Sun and H. Xu, Convergence analysis of stationary points in sample average approximation of stochastic programs with second order stochastic dominance constraints. To appear in *Math. Program.* Ser. B, 2012.
[24] W. Sun and Y. Yuan, *Optimization Theory and Methods*, Springer, New York, 2006.
[25] H. Xu, Level function method for quasiconvex programming, *J. Optim. Theory Appl.*, 108 (2001), pp. 407–437.
[26] H. Xu and D. Zhang, Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications, *Math. Program.*, 119 (2009), pp. 371–401.