Multivariate robust stochastic dominance and resulting risk-averse optimization

Zhiping Chen · Yu Mei · Jia Liu

Received: date / Accepted: date

Abstract By utilizing the min-biaffine scalarization function, we introduce a multivariate robust second-order stochastic dominance concept to flexibly compare random vectors. We discuss the basic properties of the dominance relation, and relate the multivariate robust second-order stochastic dominance to a functional which is continuous and subdifferentiable everywhere. We study the stochastic optimization problem with multivariate robust second-order stochastic dominance constraints, reformulate the constraints using the introduced functional and develop the necessary and sufficient conditions of optimality in the convex case. After specifying an ambiguity set based on moment information, we approximate the ambiguity set by a series of sets consisting of discrete distributions. Furthermore, we design a convex approximation to the proposed stochastic optimization problem and establish its qualitative stability under Kantorovich metric and pseudo metric. All these results lay a theoretical foundation for the modelling and solution of complex stochastic decision-making problems with multivariate robust second-order stochastic dominance constraints.

Keywords Multivariate stochastic order · Robust preference · Second-order stochastic dominance constraint · Stochastic optimization · Probability discretization · Stability analysis

Mathematics Subject Classification (2000) 49K45 · 49M25

Zhiping Chen, Corresponding author
Department of Computing Science, School of Mathematics and Statistics, Xi’an Jiaotong University
Xi’an, Shaanxi, P. R. China
zchen@xjtu.edu.cn

Yu Mei
Department of Computing Science, School of Mathematics and Statistics, Xi’an Jiaotong University,
Xi’an, Shaanxi, P. R. China
meiyu414@stu.xjtu.edu.cn

Jia Liu
Department of Computing Science, School of Mathematics and Statistics, Xi’an Jiaotong University,
Xi’an, Shaanxi, P. R. China
jialiu@xjtu.edu.cn
1 Introduction

The concept of stochastic dominance originated from decision theory and economics is important for comparing two random variables. Different definitions of stochastic dominance have been introduced in the literature (see [1–3]). One popular notion is second-order stochastic dominance, defined through the comparison of expectations of utility functions, which is related to risk-averse optimization in financial management.

In their seminal work [4], Dentcheva and Ruszczyński first considered the optimization problem with second-order and higher-order stochastic dominance constraints, they derived the optimality conditions and transformed the stochastic optimization problem into a linear programming problem. Furthermore, they developed duality relations and solved the dual problem by utilizing the piecewise linear structure of the dual functional in [5]. Luedtke proposed new linear formulations for second-order stochastic dominance constraints in [6], which revealed the connection between first- and second-order stochastic dominance constraints from a different perspective. A series of studies [7–9] have discussed the case where the random variables are induced by mixed-integer linear recourse.

By now, many cut generation methods have been proposed to solve stochastic optimization problems with second-order stochastic dominance constraints, see, for example, [10–12]. Attention has also been paid on the efficiency of applications of such kind of problems (see, e.g., [13]).

Recently, some researchers have considered the multivariate extension of stochastic dominance (e.g., [14]). As for the second-order stochastic dominance, there are mainly two ways to introduce the multivariate concept. The first method is to compare the expectations of multi-dimensional utility functions (see, e.g., [1,15]), while the second one is to follow the univariate concept after scalarization (see, e.g., [16]). We adopt the second technique in this paper. Through linear scalarization, the necessary and sufficient conditions of optimality and duality relations were developed in [16] for optimization problems with multivariate second-order stochastic dominance constraints. Soon after, Homem-de-Mello et al. in [17, 18] relaxed the above concept by considering a more general set of scalarization vectors. The current attention has been mainly paid on how to efficiently solve optimization problems with multivariate second-order stochastic dominance constraints, as discussed in a series of papers [17–21]. However, all these works are based on the linear scalarization function, which is used to transform multivariate stochastic dominance constraints into the univariate constraints. In 2016, Noyan and Rudolf introduced in [22] a more general kind of scalarization functions, named min-biaffine functions, based on which they extended the existing concept of multivariate second-order stochastic dominance.
With the development of distributionally robust optimization, Dentcheva and Ruszczyński proposed in [23] the notion of robust second-order stochastic dominance. They investigated the optimization problem with this kind of constraints and derived the corresponding conditions of optimality under different cases. As for the solution method of this class of problems, we only found one paper, [24]. In this paper, Guo et al. considered an ambiguity set based on moment information, investigated the underlying probability discretization scheme and finally proposed a convex approximation solution scheme for stochastic programs with robust second-order dominance constraints. The research about stochastic optimizations with robust stochastic dominance constraints just began, and there are still many issues to discuss.

As an integration and generalization of current versions of stochastic dominance, we consider the multivariate version of robust second-order stochastic dominance. This paper extends existing results from four aspects. Firstly, we take into account the uncertainty of probability distributions in multivariate second-order stochastic dominance. Secondly, we utilize the min-affine scalarization function to transform multivariate robust dominance constraints into univariate ones, which is more general and flexible for practical applications. Thirdly, we consider the stochastic optimization problem with multivariate robust second-order stochastic dominance constraints and analyze their optimality conditions. Finally, we examine the approximation scheme for the introduced stochastic optimization problem and obtain the corresponding stability results. All these results lay a solid foundation for the modelling and solution of complex stochastic optimizations with multivariate robust second-order stochastic dominance constraints.

The rest of this paper is organized as follows. In section 2, we briefly recall some preliminary results about the second-order stochastic dominance and min-biaffine scalarization functions. In section 3, we introduce the multivariate robust second-order stochastic dominance and discuss its important properties. Section 4 is devoted to the stochastic optimization problem with multivariate robust second-order stochastic dominance constraints. We establish the necessary and sufficient optimality conditions for such optimization problems in the convex case. In section 5, we discuss the approximation methods for the proposed stochastic optimization problem and carry out the stability analysis when the ambiguity set is specified by moment information. Section 6 gives the conclusions.

2 Preliminaries

We first introduce some notations used in the rest of this paper.
For vectors \( v, \tilde{v} \in \mathbb{R}^m \), we use \( v \succeq^{sep} \tilde{v} \) to represent the componentwise comparison. Both \( v_i \) and \([v_i]\) denote the \( i \)th component of \( v \). We use \( (\cdot)_+ = \max\{0, \cdot\} \) to represent the positive part. The space of regular countably additive measures on a compact set \( D \subset \mathbb{R}^n \) with finite variation is denoted by \( \mathcal{M}(D) \), while \( \mathcal{M}_+(D) \) represents all such measures which are nonnegative.

We suppose that probability measure of random vectors is \( P_0 \). \( L^p(\Omega, \mathcal{F}, P_0; \mathbb{R}^m) \) (shortly \( L^m \), and we omit the superscript when \( m = 1 \)) denotes the Cartesian product space \( L^m = \prod_{i=1}^m L^p(\mathbb{R}) \) with norm defined by
\[
||X||_p := \left( \sum_{i=1}^m \int_\Omega |X_i(\omega)|^p dP_0(\omega) \right)^{1/p}.
\]

We need to make sure that multivariate random vectors are defined in an adequate space, namely a linear and separable space. Actually, since the Cartesian product space of separable Banach spaces is still a separable Banach space with the norm \( ||X||_p \) (Theorem 1.23 in [25]), we have

**Lemma 2.1** \( L^m \) is a separable Banach space.

As a preparation for our later study, we recall some basic definitions and notations for second-order stochastic dominance and min-biaffine scalarization functions.

Given a random variable \( X \), we denote its cumulative distribution function by \( F(X; \alpha) = P_0(X \leq \alpha), \alpha \in \mathbb{R} \).

We say that a random variable \( X \) dominantes in the first order a random variable \( Y \) if \( F(X; \eta) \leq F(Y; \eta), \forall \eta \in \mathbb{R} \). We denote \( F_2(X; \cdot) \) as \( F_2(X; \eta) = \int_{-\infty}^{\eta} F(X; \alpha) d\alpha \). Moreover, we have
\[
F_2(X; \eta) = \mathbb{E}_{P_0}[(\eta - X)_+].
\] (1)

We say that a random variable \( X \) dominantes in the second order a random variable \( Y \), written as \( X \succeq^{(2)} Y \), if \( F_2(X; \eta) \leq F_2(Y; \eta), \forall \eta \in \mathbb{R} \). The second order stochastic dominance has an equivalent representation in terms of utility functions: \( \mathbb{E}_{P_0}[u(X)] \geq \mathbb{E}_{P_0}[u(Y)] \) for all \( u \in \mathcal{U} \), here \( \mathcal{U} \) denotes the set of all concave and nondecreasing functions \( u: \mathbb{R} \to \mathbb{R} \) satisfying \( \lim_{t \to -\infty} u(t)/t < \infty \), which is called the linear growth condition. We call \( \mathcal{U} \) a generator of \( \succeq^{(2)} \).

Next we examine the basic properties of the min-biaffine scalarization function. Noyan and Rudolf introduced in [22] the following min-biaffine scalarization function
\[
\theta(c, x) = \min_{1 \leq t \leq T} A_t(c, x),
\]
here $T$ is a positive integer, $A_t: \mathbb{IR}^m \times \mathbb{IR}^m \to \mathbb{IR}$ are affine in both variables for all $t = 1, \ldots, T$. For presentation convenience, we denote $	heta(c, x) = \min_{1 \leq t \leq T} \{a_T^T(c)x + b_t(c)\}$ when $c$ is fixed.

It is said that the random vector $X$ dominates $Y$ in the second order with respect to the scalarization function $\theta$ and a set $C \subset \mathbb{IR}^m$, denoted as $X \succeq_{\theta,C}^2 Y$, if

$$
\theta(c, X) \succeq_{(2)} \theta(c, Y), \ \forall c \in C.
$$

Such a class of functions include most scalarization functions that are commonly used in the multi-objective optimization. For example, the linear scalarization function that Dentcheva et al. adopted in [16] is a special case of the min-biaffine scalarization function with $T = 1$, $A_1(c, x) = c^T x$ and $C = \mathbb{IR}^m_+$. When $\theta$ is continuous with respect to $c$ for all $x \in \mathbb{IR}^m$ and $C$ is nonempty and convex, we have that $X \succeq_{\theta,C}^2 Y$ is equivalent to $X \succeq_{\tilde{\theta}, \tilde{C}}^2 Y$, here $\tilde{C} = \{c \in \cl\cone(C) : ||c||_1 \leq 1\}$, where $\cl\cone(C)$ denotes the closure of the conical hull of the set $C$ (see [22]). Since $\tilde{C}$ is a compact and convex set, it is easier to handle. In the remainder of this paper, we always assume that all scalarization vectors in $C$ (and thus those in $\tilde{C}$) are non-negative. This assumption is reasonable because it is rather strict to require that both $c$ and $-c$ are in $C$ and, more importantly, few random vectors can be compared under this situation.

As a foundation for the rest of this paper, the following lemma establishes the $p$th integrability of $\theta(c, X)$.

**Lemma 2.2** For fixed $c \in C$, $X \in \mathcal{L}_p^m$ implies $\theta(c, X) \in \mathcal{L}_p$.

**Proof** It is sufficient to prove the conclusion for $\theta(c, x) = a_T^T(c)x + b_t(c)$. This can be proved by Minkowski’s inequality and the convexity of $(\cdot)^p$, and we omit the details here. □

**Assumption 2.1** For any fixed $c \in C$, $\theta(c, \cdot)$ is nondecreasing in the sense that $x \succeq_{\text{sep}} y$ entails $\theta(c, x) \geq \theta(c, y)$.

This assumption is intuitive and natural. It automatically holds for many problems. For instance, in portfolio optimizations, given two profit vectors $x$ and $y$, if each component of $x$ is greater than that of $y$, then it is quite natural to assume that the scalarized profit $\theta(c, x)$ is larger than $\theta(c, y)$. To our knowledge, almost all existing min-biaffine scalarization functions satisfy Assumption 2.1. Under this condition, we have

**Lemma 2.3** Given Assumption 2.1, $a_t(c) \in \mathbb{IR}^m$ is a vector with non-negative components for any $c \in C$ and $t = 1, \ldots, T$.
Proof We prove by contradiction. Assume that there exist a vector \( \hat{c} \in C \), an integer \( \hat{t} \in \{1, \ldots, T\} \) and some integer \( \hat{t} \in \{1, \ldots, n\} \) such that \( |a_{\hat{t}}(\hat{c})|_1 < 0 \). Consider the \( m \)-dimensional unit vector \( e_1 \) and the \( m \)-dimensional zero vector \( 0 \). We have \( e_1 \preceq_{sp} 0 \). However

\[
\theta(\hat{c}, e_1) = \min_{1 \leq t \leq T} \{ a_t^T(\hat{c}) e_1 + b_t(\hat{c}) \} \leq a_T^T(\hat{c}) e_1 + b_T(\hat{c}) = |a_1(\hat{c})|_1 + b_1(\hat{c}) < b_1(\hat{c}) = \theta(\hat{c}, 0),
\]

which contradicts Assumption 2.1. \( \square \)

Lemma 2.3 implies that \( \theta(c, x) \) is non-decreasing with respect to \( x \) for any fixed \( c \). The min-biaffine scalarization function \( \theta \) also has the following property, which is a kind of Lipschitz continuity with respect to random vectors.

**Lemma 2.4** For any fixed \( c \in C \), there exists a constant \( C_0(c) = \max_{1 \leq t \leq T} ||a_t(c)||_1 \) such that for any \( X, Y \in \mathcal{L}^m_p \), we have

\[
||\theta(c, X) - \theta(c, Y)||_p \leq C_0(c)||X - Y||_p.
\]

**Proof** For any \( \omega \in \Omega \), assume \( t^* \in \arg\min_{1 \leq t \leq T} \{a_t^T(c)Y(\omega) + b_t(c)\} \). If \( \theta(c, X(\omega)) \geq \theta(c, Y(\omega)) \), then we get

\[
\left| \theta(c, X(\omega)) - \theta(c, Y(\omega)) \right|^p = \left( \min_{1 \leq t \leq T} \{a_t^T(c)X(\omega) + b_t(c)\} - \min_{1 \leq t \leq T} \{a_t^T(c)Y(\omega) + b_t(c)\} \right)^p 
\leq \left( a_{t^*}^T(c)X(\omega) + b_{t^*}(c) - \left[ a_{t^*}^T(c)Y(\omega) + b_{t^*}(c) \right] \right)^p = \left( a_{t^*}^T(c) \left[ X(\omega) - Y(\omega) \right] \right)^p 
\leq \left( \sum_{i=1}^m \left[ a_{t^*}(c) \right]_i \left| X_i(\omega) - Y_i(\omega) \right| \right)^p = \left( \left[ a_{t^*}(c) \right]_1 \left| X_1(\omega) - Y_1(\omega) \right| \right)^p 
\leq [C_0(c)]^p \sum_{i=1}^m \left[ a_{t^*}(c) \right]_i \left| X_i(\omega) - Y_i(\omega) \right|^p \quad \text{(by the convexity of \( (\cdot)^p \))} 
\leq [C_0(c)]^p \sum_{i=1}^m \left| X_i(\omega) - Y_i(\omega) \right|^p \quad \text{(by \( \left[ a_{t^*}(c) \right]_1 \leq 1 \))}
\]

Notice that the above upper bound is independent of \( t^* \). Similarly, if \( \theta(c, X(\omega)) < \theta(c, Y(\omega)) \), we have

\[
\left| \theta(c, X(\omega)) - \theta(c, Y(\omega)) \right|^p = \left( \theta(c, Y(\omega)) - \theta(c, X(\omega)) \right)^p \leq [C_0(c)]^p \sum_{i=1}^m \left| X_i(\omega) - Y_i(\omega) \right|^p.
\]

Hence,

\[
||\theta(c, X) - \theta(c, Y)||_p = \left( \int_{\Omega} \left| \theta(c, X(\omega)) - \theta(c, Y(\omega)) \right|^p dP_0(\omega) \right)^{1/p}
\]
\[ \leq \left[ C_0(c) \right] \left( \int_{\Omega} \left| \sum_{i=1}^{m} X_i(\omega) - Y_i(\omega) \right|^p dP_0(\omega) \right)^{1/p} = \left[ C_0(c) \right] \| X - Y \|_p. \]

Since the min-biaffine function \( \theta(c, x) \) is concave and continuous with respect to both \( c \) and \( x \), it is superdifferentiable at any point \(( c, x )\).

### 3 Multivariate robust stochastic dominance

In many real applications, it is difficult to specify the true measure \( P_0 \) precisely, but we know that \( P_0 \) is included in a set of probability measures. In detail, we consider \( X, Y \) in a measurable space \(( \Omega, \mathcal{F} )\) with \( \mathcal{F} \) the Borel \( \sigma \)-algebra on \( \Omega \). Let \( \mathcal{M} \) be the set of all probability measures on \(( \Omega, \mathcal{F} )\). We consider a subset \( Q \subset \mathcal{M} \) which contains the true measure \( P_0 \). We assume that the absolute values of each component of both \( X \) and \( Y \) raised to the \( p \)th power have finite integral with respect to any measure \( P \in Q \).

Now it comes to recalling the concept of robust second-order stochastic dominance, first introduced in [23].

**Definition 3.1** A random variable \( X \in \mathbb{L}_p \) dominates robustly a random variable \( Y \in \mathbb{L}_p \) in the second-order over a set of probability measures \( Q \) if

\[ E_P[u(X)] \geq E_P[u(Y)], \quad \forall u \in \mathcal{U}, \forall P \in Q. \tag{3} \]

Here \( \mathcal{U} \) is the set of concave and nondecreasing utility functions defined above. We denote the robust dominance relation by \( X \succeq Q \ Y \).

Like many papers about robust optimization, see, for example, [23,26], we assume that \( Q \) satisfies the following conditions.

**Assumption 3.1** \( Q \) is convex, closed, and bounded.

Assumption 3.1 implies that,

\[ B = \sup_{P_1 \in Q} \sup_{P_2 \in Q} \left\| \frac{dP_1}{dP_2} \right\|_q < \infty, \]

where \( q \) satisfies \( 1/p + 1/q = 1 \).

Chosen a min-biaffine scalarization function \( \theta: \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R} \) and a nonempty, convex set \( C \subset \mathbb{R}_+^m \), we can now introduce the following concept of multivariate robust second-order stochastic dominance.
Definition 3.2 A random vector $X \in \mathcal{L}_p^m$ dominates robustly a random vector $Y \in \mathcal{L}_p^m$ in the second-order over a set of probability measures $\mathcal{Q}$ if

$$E_P[u(\theta(c, X))] \geq E_P[u(\theta(c, Y))], \quad \forall c \in C, \forall u \in \mathcal{U}, \forall P \in \mathcal{Q}. \quad (4)$$

We denote the multivariate robust second-order stochastic dominance relation by $X \succeq^{\theta, C, \mathcal{Q}}_{(2)} Y$.

Remark 3.1 Definition 3.2 generalizes various current stochastic dominance definitions and, especially, includes the following typical cases:

(i) If $\mathcal{Q}$ is a singleton set, $\theta(c, x) = c^T x$ and $C = \mathbb{R}_m^m$, the multivariate robust second-order stochastic dominance relation reduces to the linearly multi-dimension second-order stochastic dominance relation in [16];

(ii) If $m = 1$ and $\theta(c, x) = x$, it reduces to the robust second-order stochastic dominance relation defined in Definition 3.1;

(iii) If $m = 1$, $\mathcal{Q}$ is a singleton set and $\theta(c, x) = x$, it reduces to the classical second-order stochastic dominance.

Definition 3.2 is very flexible in the sense that the set $C$, the ambiguity set $\mathcal{Q}$ and the scalarization function $\theta$ can all be specified by decision makers. The flexible choice of $C$, $\mathcal{Q}$ and $\theta$ can efficiently reflect different risk aversions for different decision makers.

The following observations can be deduced from Definition 3.2.

Proposition 3.1 For $X, Y \in \mathcal{L}_p^m$, the following conditions are equivalent:

(i) $X \succeq^{\theta, C, \mathcal{Q}}_{(2)} Y$;

(ii) $\theta(c, X) \succeq^{\mathcal{Q}}_{(2)} \theta(c, Y), \forall c \in C$;

(iii) $E_P[\varphi(X)] \geq E_P[\varphi(Y)], \forall \varphi \in \Phi, \forall P \in \mathcal{Q}$, where

$$\Phi = \left\{ \int_C [Q(c)][\theta(c, x)] \mu(dc) : \mu \in \mathcal{M}_+(C), \right\};$$

$$Q: C \rightarrow \mathcal{U}, \text{ such that } (c, x) \rightarrow [Q(c)][\theta(c, x)] \text{ is Lebesgue measurable on } C \times \mathbb{R}_m^m \};$$

(iv) $E_P[(\eta - \theta(c, X))_+] \leq E_P[(\eta - \theta(c, Y))_+], \forall \eta \in \mathbb{R}, \forall c \in C, \forall P \in \mathcal{Q}$;

(v) $\text{CVaR}_\alpha(X) \geq \text{CVaR}_\alpha(Y), \forall c \in C, \forall P \in \mathcal{Q}, \forall \alpha \in (0, 1]$. 
Proof From what we have discussed in Sections 2 and 3, (i), (ii) and (iv) are equivalent. The equivalence between (i) and (iii) follows similarly as that in Proposition 2 in [16] after proper adaption. The equivalence between (iv) and (v) is based on the relationship between second-order dominance and CVaR (see, e.g., [5]).

\[ \square \]

Remark 3.2 \( C \) can be replaced by \( \tilde{C} \) without any influence according to the discussion in Section 2.

In what follows, we treat \( Y \) as the reference random vector from \( \mathcal{L}_m^p \). We know from Proposition 3.1 that \( X \succeq_{(2)} \tilde{C}, \mathcal{Q} \) \( Y \) is equivalent to

\[
\sup_{P \in \mathcal{Q}} E_P[(\eta - \theta(c, X))_+ - (\eta - \theta(c, Y))_+] \leq 0, \forall (c, \eta) \in \tilde{C} \times \mathbb{R}.
\]  

(5)

We next introduce a functional \( \sigma : \mathcal{L}_p^m \rightarrow \mathbb{R} \) defined as

\[
\sigma(V) = \sup_{P \in \mathcal{Q}} E_P[V],
\]  

(6)

which has the following properties as discussed in [27].

Lemma 3.1 If the set \( \mathcal{Q} \) is convex, closed and bounded, then \( \sigma(\cdot) \) is convex and subdifferentiable everywhere. Moreover for any \( V \in \mathcal{L}_p \) we have \( \partial \sigma(V) = \{P \in \mathcal{Q} : E_P[V] = \sigma(V)\} \), and \( \sigma(\cdot) \) is Lipschitz continuous on \( \mathcal{L}_p \) with modulus \( B \).

It is known from Lemma 2.2 that \( \theta(c, X) \) and \( \theta(c, Y) \) are \( p \)-th integrable, thus \( (\eta - \theta(c, X))_+ - (\eta - \theta(c, Y))_+ \in \mathcal{L}_p^m \). We define \( \rho_{c,\eta} : \mathcal{L}_p^m \rightarrow \mathbb{R} \) as

\[
\rho_{c,\eta}(X) = \sigma[(\eta - \theta(c, X))_+ - (\eta - \theta(c, Y))_+] \\
= \sup_{P \in \mathcal{Q}} E_P[(\eta - \theta(c, X))_+ - (\eta - \theta(c, Y))_+], (c, \eta) \in \tilde{C} \times \mathbb{R}.
\]  

(7)

Proposition 3.2 Given Assumptions 2.1 and 3.1, \( \rho_{c,\eta}(\cdot) \) has the following properties:

(i) it is convex;

(ii) it is nonincreasing in the sense that

\[
X_1(\omega) \succeq_{sep} X_2(\omega), \forall \omega \in \Omega \Rightarrow \rho_{c,\eta}(X_1) \leq \rho_{c,\eta}(X_2);
\]
(iii) it is Lipschitz continuous with modulus $B \cdot C_0(c)$.

Proof
(i) By Lemma 3.1, the functional $V \mapsto \sup_{P \in \mathcal{Q}} E_P[V]$ is convex. Moreover, $\sup_{P \in \mathcal{Q}} E_P[V]$ is nondecreasing, i.e.,

$$V_1(\omega) \geq V_2(\omega), \forall \omega \in \Omega \Rightarrow \sup_{P \in \mathcal{Q}} E_P[V_1] \geq \sup_{P \in \mathcal{Q}} E_P[V_2].$$

Together with the fact that $X \mapsto (\eta - \theta(c, X))_+$ is convex, we conclude that $\rho_{c, \eta}$ is convex.

(ii) Suppose that $X_1(\omega) \succeq_{sep} X_2(\omega), \forall \omega \in \Omega$, we have from Assumption 2.1 that

$$(\eta - \theta(c, X_1)(\omega))_+ - (\eta - \theta(c, X_2)(\omega))_+ \leq (\eta - \theta(c, Y)(\omega))_+, \forall \omega \in \Omega, \forall (c, \eta) \in \tilde{C} \times \mathbb{R},$$

which yields $\rho_{c, \eta}(X_1) \leq \rho_{c, \eta}(X_2)$.

(iii) From Lemma 2.4, Proposition 3.3 in [23] and Hölder’s inequality, we obtain

$$|\rho_{c, \eta}(X_1) - \rho_{c, \eta}(X_2)| \leq B \||\eta - \theta(c, X_1)|_+ - |\eta - \theta(c, X_2)|_+\|_p \leq B \|\theta(c, X_1) - \theta(c, X_2)\|_p \leq B \cdot C_0(c) \|X_1 - X_2\|_p,$$

which establishes the conclusion (iii).

Now we consider the subdifferential of $\rho_{c, \eta}(\cdot)$, we need the following lemma deriving the subdifferential of a composite functional.

Lemma 3.2 [28] Suppose that $\mathcal{X}$ is a Banach space, $\mathcal{Z}$ is a separable Banach space and $F: \mathcal{Z} \to \mathcal{X}$ is a mapping. We write $f(z, \omega)$ or $f_\omega(z)$ for $[F(z)](\omega)$. Assume that $F$ is convex (i.e., $f(\cdot, \omega)$ is convex for every $\omega \in \Omega$) and continuous at $\tilde{z}$, the function $\delta: \mathcal{X} \to \mathbb{R}$ is convex and non-decreasing (i.e., $\delta(Y) \geq \delta(X)$ if $Y(\omega) \geq X(\omega)$ for all $\omega \in \Omega$), and is finite valued and continuous at $\tilde{X} = F(\tilde{z})$. Then the composite function $\psi(\cdot) := \delta(F(\cdot))$ is subdifferentiable at $\tilde{z}$ and

$$\partial \psi(\tilde{z}) = \text{cl} \left( \bigcup_{\mu \in \partial \delta(\tilde{X})} \int_\Omega \partial f_\omega(\tilde{z}) d\mu(\omega) \right).$$
For \((c, \eta) \in \bar{C} \times \mathbb{R}\), we define the multifunctions \(D_{c, \eta} : \mathcal{L}_p^m \rightarrow \mathbb{R}\) as:

\[
D_{c, \eta}(X, \omega) = \begin{cases} 
\{ -a_i(c) \}, & \text{if } \theta_c(X(\omega)) < \eta \text{ and } i = \text{argmin}_i \{ a_i^T(c) X(\omega) + b_i(c) \}, \\
\text{conv}\{ -a_i(c), -a_{i_2}(c) \}, & \text{if } \theta_c(X(\omega)) < \eta \text{ and } \{ i_1, i_2 \} = \text{argmin}_i \{ a_i^T(c) X(\omega) + b_i(c) \} (i_1 \neq i_2), \\
\vdots, & \\
\text{conv}\{ -a_1(c), \cdots, -a_{i_T}(c) \}, & \text{if } \theta_c(X(\omega)) < \eta \text{ and } \{ i_1, \cdots, i_{T-1} \} = \text{argmin}_i \{ a_i^T(c) X(\omega) + b_i(c) \} \quad (9) \\
n\text{conv}\{ 0, -a_i(c) \}, & \text{if } \theta_c(X(\omega)) = \eta \text{ and } i = \text{argmin}_i \{ a_i^T(c) X(\omega) + b_i(c) \}, \\
\text{conv}\{ 0, -a_1(c), -a_{i_2}(c) \}, & \text{if } \theta_c(X(\omega)) = \eta \text{ and } \{ i_1, i_2 \} = \text{argmin}_i \{ a_i^T(c) X(\omega) + b_i(c) \} (i_1 \neq i_2), \\
\vdots, & \\
\text{conv}\{ 0, -a_1(c), \cdots, -a_{i_{T-1}}(c) \}, & \text{if } \theta_c(X(\omega)) = \eta \text{ and } \{ i_1, \cdots, i_{T-1} \} = \text{argmin}_i \{ a_i^T(c) X(\omega) + b_i(c) \} \quad (9) \\
\{ 0 \}, & \text{if } \theta_c(X(\omega)) > \eta.
\end{cases}
\]

We know from Lemma 2.3 that the elements in \(D_{c, \eta}\) are all non-positive vectors. We further define

\[
\mathcal{A}_{c, \eta}(X) = \partial \sigma[ (\eta - \theta_c(X))_+ - (\eta - \theta_c(Y))_+] , \quad X \in \mathcal{L}_p^m.
\]

**Proposition 3.3** For any \((c, \eta) \in \bar{C} \times \mathbb{R}\), the functional \(\rho_{c, \eta}(\cdot)\) is continuous and subdifferentiable on \(\mathcal{L}_p^m\), and its subdifferential at a point \(X \in \mathcal{L}_p^m\) is

...
\[
\partial \rho_{c,\eta}(X) = \left\{ Q \in \mathcal{L}^m_q : \exists P_{c,\eta} \in \mathcal{A}_{c,\eta}(X), \exists \lambda(\omega) \in D_{c,\eta}(X,\omega) \text{ such that } Q = \int_{\Omega} \lambda(\omega) dP_{c,\eta}(\omega) \right\}
\]
\[
= D_{c,\eta}(X,\cdot) \circ \mathcal{A}_{c,\eta}(X).
\]

(10)

Proof Define the functional \( \Psi_{c,\eta} : \mathcal{L}^m_p \to \mathcal{L}_p \) by

\[
\Psi_{c,\eta}(X) = (\eta - \theta_c(X))_+ - (\eta - \theta_c(Y))_+.
\]

Then \( \rho_{c,\eta}(\cdot) \) is a composite function, \( \rho_{c,\eta}(X) = \sigma(\Psi_{c,\eta}(X)) \). We write \( [\Psi_{c,\eta}(X)](\omega) \) as \( \Psi_{\omega}(X) \), for short. We notice that

\[
\partial \Psi_{\omega}(X) = D_{c,\eta}(X,\omega), \quad \partial \sigma(\Psi_{c,\eta}(X)) = \mathcal{A}_{c,\eta}(X).
\]

We set \( X = \mathcal{L}^p, Z = \mathcal{L}^m_p, F = \Psi_{c,\eta}, \delta = \sigma \) and \( \psi = \rho_{c,\eta} \) in Lemma 3.2, it is known from Lemma 2.1 that \( \mathcal{L}^m_p \) is a separable Banach space. It is also easy to verify that \( \Psi_{\omega}(X) \) is convex for every \( \omega \in \Omega \) and continuous everywhere. By equation (6) and Lemma 3.1, \( \sigma \) is convex, non-decreasing and Lipschitz continuous, and thus is finite and continuous everywhere. Then we know from Lemma 3.2 that the composite function \( \rho_{c,\eta} \) is subdifferentiable everywhere and for any \( X \in \mathcal{L}^m_p \),

\[
\partial \rho_{c,\eta}(X) = \text{cl} \left( \bigcup_{\mu \in \partial \sigma(\Psi_{c,\eta}(X))} \int_{\Omega} \partial \Psi_{\omega}(X) d\mu(\omega) \right)
\]
\[
= \text{cl} \left\{ Q \in \mathcal{L}^m_q : \exists P_{c,\eta} \in \mathcal{A}_{c,\eta}(X), \exists \lambda(\omega) \in D_{c,\eta}(X,\omega), Q = \int_{\Omega} \lambda(\omega) dP_{c,\eta}(\omega) \right\}.
\]

(11)

One can refer to [28, 29] for better understanding the integral of a set or a vector. Then it is sufficient to prove that \( \partial \rho_{c,\eta}(X) \) in (10) is closed. Assume that \( Q^k \to \hat{Q} \) in the weak* topology, and we will show that \( \hat{Q} \) is an element of this set. For each \( Q^k \), according to (11), there exist \( P^k \in \mathcal{A}_{c,\eta}(X) \) and \( \lambda^k(\omega) \in D_{c,\eta}(X,\omega) \) such that

\[
Q^k = \int_{\Omega} \lambda^k(\omega) dP^k(\omega).
\]

Due to Lemma 3.1, \( \mathcal{A}_{c,\eta}(X) \) is a convex, weakly* closed and bounded set. As a result, it is weakly* compact. Therefore, we can extract from the sequence \( \{P^k\} \) a weakly* convergent subsequence \( \{P^k\} \in K, \) and its weakly* limit \( \hat{P} \) is an element of \( \mathcal{A}_{c,\eta}(X) \).

Next, we make a partition of \( \Omega \) and discuss how \( Q^k \) and \( \hat{Q} \) take value in different situations.

Case 1: \( A \subset \{ \omega \in \Omega : \theta_c(X(\omega)) < \eta \} \) and \( i = \text{argmin}_t \{a_t^2(c)X(\omega) + b_t(c)\} \), then we have
\[ Q^k(A) = \int_A -a_i(c)dP^k(\omega) = -a_i(c)P^k(A). \]

Consequently, \( \hat{Q}(A) = -a_i(c)\hat{P}(A). \)

**Case 2:** \( B \subset \{ \omega \in \Omega : \theta_c(X(\omega)) = \eta \} \) and \( \{ 1, \cdots, T \} = \arg\min_t \{ a_T^T(c)X(\omega) + b_t(c) \} \), then we have

\[ Q^k(B) \in \int_B \text{conv}\{0, -a_1(c), \cdots, -a_T(c)\}dP^k(\omega) = \text{conv}\{0, -a_1(c), \cdots, -a_T(c)\}P^k(B). \]

This ensures, together with the fact that \( \text{conv}\{0, -a_1(c), \cdots, -a_T(c)\} \) is a compact set, that

\[ \hat{Q}(B) \in \text{conv}\{0, -a_1(c), \cdots, -a_T(c)\}\hat{P}(B). \]

**Case 3:** \( C \subset \{ \omega \in \Omega : \theta_c(X(\omega)) > \eta \} \), we have \( Q^k(C) = 0 \) and \( \hat{Q}(C) = 0. \)

The other cases can be discussed in the way similar to that in **Case 2**.

The above results mean that \( \hat{Q} \) is a \( \hat{P} \)-continuous vector measure (i.e., \( \lim_{\hat{P}(E) \to 0} \hat{Q}(E) = 0 \)), see, for example, [30]. Furthermore, \( \hat{Q} \) is a measure with bounded variation, i.e.,

\[ ||\hat{Q}||(\Omega) = \sup_{\pi} \sum_{E \in \pi} ||\hat{Q}(E)||_p \leq \sup_{\pi} \sum_{E \in \pi} \max_{1 \leq t \leq T} \{ ||a_t(c)||_p \} \cdot \hat{P}(E) \leq \max_{t} \{ ||a_t(c)||_p \} < \infty, \]

where the supremum is taken over all finite partitions \( \pi \) of \( \Omega \). From the vector extensions of Radon-Nikodým Theorem (see, e.g., Chap.III in [30]), there exists a Bochner integrable mapping \( g \in \mathcal{L}^m_1 \) (i.e., there exists a sequence of simple functions \( g_n \) such that \( \lim_{n \to \infty} \int_{\Omega} ||g_n - g||_p d\hat{P} = 0 \) such that \( \hat{Q}(E) = \int_{E} gd\hat{P} \) for all \( E \in \mathcal{F} \). From what we have discussed in the last paragraph, we also have \( g(\omega) \in D_{c,\eta}(X, \omega), \omega \in \Omega \). Therefore, the limit point \( \hat{Q} \) is in \( \partial_{\rho,c,\eta}(X) \).

Proposition 3.3 differs from Lemma 1 in [23] in that we are dealing with random vectors rather than scalars.

**4 Optimization problem with multivariate robust second-order stochastic dominance constraints**

Having examined the basic properties of the introduced multivariate robust second-order stochastic dominance, we consider in this section its application to the risk-averse optimization problem. We assume that \( Z_0 \) is a nonempty, convex and closed subset of a Banach space \( \mathcal{X} \). Choosing \( G: \mathcal{X} \to \mathcal{L}^m_p, H: \mathcal{X} \to \mathcal{L}^p, Y \in \mathcal{L}^m_p \), and a
functional $\phi : \mathcal{L}_p \rightarrow \mathbb{R}$, we consider the following optimization problem with a multivariate robust second-order stochastic dominance constraint:

$$
\begin{align*}
\min_{z \in Z_0} & \quad \phi(H(z)) \\
\text{s.t.} & \quad G(z) \succeq_{(2)}^{\theta,\mathcal{C},\mathcal{Q}} Y.
\end{align*}
$$

(12) (13)

The optimization model (12)-(13) contains almost all the existing optimization models with second-order stochastic dominance constraints due to the flexibility of our new dominance definition. From what we have obtained in the last section, the multivariate robust second-order stochastic dominance constraint (13) can be reformulated as:

$$
\rho_{c,\eta}(G(z)) \leq 0, \quad \forall (c, \eta) \in \tilde{C} \times \mathbb{R}.
$$

Let $\mathcal{M} : \tilde{C} \Rightarrow \mathbb{R}$ be a multifunction with a nonempty compact graph. In what follows, we only consider $(c, \eta)$ in the set graph $\mathcal{M}$, i.e., we consider a relaxation problem of problem (12)-(13):

$$
\begin{align*}
\min_{z \in Z_0} & \quad \phi(H(z)) \\
\text{s.t.} & \quad \rho_{c,\eta}(G(z)) \leq 0, \quad \forall (c, \eta) \in \text{graph } \mathcal{M} \subset \tilde{C} \times \mathbb{R}.
\end{align*}
$$

(14) (15)

The reason for this relaxation is to satisfy the Slater constraint qualification. If for every $c \in \tilde{C}$, $\mathcal{M}(c)$ includes all possible realizations of $\theta(c, Y(\omega))$, problems (12)-(13) and (14)-(15) are equivalent. Our objective is to develop the optimality conditions of problem (14)-(15) under the following conditions:

**Assumption 4.1**

1. For almost all $\omega \in \Omega$, $z \mapsto [H(z)](\omega)$ and $z \mapsto [G(z)](\omega)$ are continuous and concave mappings, here $G$ is concave in the sense that:

$$
G(\lambda z_1 + (1 - \lambda) z_2)(\omega) \succeq_{\text{sep}} \lambda G(z_1)(\omega) + (1 - \lambda) G(z_2)(\omega)
$$

for any $\lambda \in [0, 1]$;

2. $\phi(\cdot)$ is a continuous, nonincreasing and convex functional.

Given Assumption 4.1, the optimization problem (14)-(15) is a convex semi-infinite optimization problem with composite structure.

**Definition 4.1** Problem (14)-(15) satisfies the uniform multivariate robust dominance condition if there exists a point $\tilde{z} \in Z_0$ such that
\[
\max_{P \in \mathbb{Q}} \max_{(c, \eta) \in \text{graph} \mathcal{M}} \mathbb{E}_P[(\eta - \theta(c, G(\hat{z})))_+ - (\eta - \theta(c, Y))_+] < 0.
\]

With the above preparations, we can establish the optimality conditions for problem (14)-(15).

**Theorem 4.1** Assume that Assumption 4.1 holds and the uniform multivariate robust dominance condition is satisfied. If \( \hat{z} \) is an optimal solution of problem (14)-(15), then there exist measures \( \hat{S} \in \partial \phi[H(\hat{z})], P_{c, \eta} \in A_{c, \eta}(G(\hat{z})) \), a measurable selection \( \lambda_{c, \eta} \in D_{c, \eta}(G(\hat{z}), \omega), (c, \eta) \in \text{graph} \mathcal{M}, \omega \in \Omega \), and a measure \( \hat{\nu} \in \mathcal{M}_+\text{(graph}\mathcal{M}) \), such that \( \hat{z} \) is an optimal solution of the problem

\[
\min_{z \in Z_0} \left\{ \int_{\Omega} H(z)\hat{S}(d\omega) + \int_{\text{graph} \mathcal{M}} \int_{\Omega} \lambda_{c, \eta}(\omega)G(z)P_{c, \eta}(d\omega)d\hat{\nu} \right\}
\]

and the following complementary condition is satisfied

\[
\int_{\text{graph} \mathcal{M}} \mathbb{E}_{P_{c, \eta}}[(\eta - \theta(c, G(\hat{z})))_+]d\hat{\nu} = \int_{\text{graph} \mathcal{M}} \mathbb{E}_{P_{c, \eta}}[(\eta - \theta(c, Y))_+]d\hat{\nu}.
\]

Conversely, if for some \( \hat{S} \in \partial \phi[H(\hat{z})], P_{c, \eta} \in A_{c, \eta}(G(\hat{z})), \lambda_{c, \eta}(\omega) \in D_{c, \eta}(G(\hat{z}, \omega) \text{ and } \hat{\nu} \in \mathcal{M}_+\text{(graph}\mathcal{M}) \), the optimal solution of problem (16) satisfies (17) and (15), then \( \hat{z} \) is an optimal solution of problem (14)-(15).

**Proof** Define

\[
E = \{(z, X, V) \in Z_0 \times L^m_p \times L^p_p : G(z)(\omega) \succeq_{\text{sep}} X(\omega), H(z)(\omega) \succeq V(\omega), \omega \in \Omega\}.
\]

We consider the following optimization problem:

\[
\min \phi(V)
\]

s.t. \( \rho_{c, \eta}(X) \leq 0, \forall (c, \eta) \in \text{graph} \mathcal{M}, (z, X, V) \in E. \)

(18)

We should keep the following claim in mind, which establishes the relationship between problems (18) and (14)-(15). If \( \hat{z} \) is an optimal solution of problem (14)-(15), then \((\hat{z}, G(\hat{z}), H(\hat{z}))\) is an optimal solution of problem (18).

Conversely, for every optimal solution \((\hat{z}, \hat{X}, \hat{V})\) of problem (18), \((\hat{z}, G(\hat{z}), H(\hat{z}))\) is also an optimal solution of problem (18), which entails that \( \hat{z} \) is an optimal solution of problem (14)-(15).
The Lagrangian function of problem (18) is:

\[
L(z, X, V, \mu) = \phi(V) + \int_{\text{graph}M} \rho_{c, \eta}(X) d\mu.
\]

(19)

Notice that the uniform multivariate robust dominance condition is a form of the Slater condition. If \( \hat{z} \) is an optimal solution of problem (14)-(15), then we know that \( (\hat{z}, G(\hat{z}), H(\hat{z})) \) is an optimal solution of problem (18). Applying the necessary conditions of optimality in Banach spaces (see, e.g., Theorem 3.4 in [31]), we conclude that there exists a measure \( \hat{\nu} \in \mathcal{M}_+ (\text{graph}M) \) such that

\[
L(\hat{z}, G(\hat{z}), H(\hat{z}), \hat{\nu}) = \min_{(z, X, V) \in E} L(z, X, V, \hat{\nu}),
\]

(20)

and the following complementary condition holds:

\[
\int_{\text{graph}M} \rho_{c, \eta}(G(\hat{z})) d\hat{\nu} = 0.
\]

(21)

As the Lagrangian function in (19) is continuous, convex and hence everywhere subdifferentiable with respect to \((X, V)\), the minimization problem in (20) can be equivalently stated as follows: there exists a subgradient \((\hat{Q}, \hat{S}) \in \partial_{(X, V)} L(\hat{z}, G(\hat{z}), H(\hat{z}), \hat{\nu})\) such that for all \((z, X, V) \in E\),

\[
< \hat{S}, V - H(\hat{z}) > + < \hat{Q}, X - G(\hat{z}) > \geq 0.
\]

(22)

To compute \( \hat{Q} \), we need the interchangeability of partial differential operator and integral. Since the space \( L^p_m \), for \( p \in [1, \infty) \), is separable, we have from Theorem 1 in [23] that

\[
\partial \int_{\text{graph}M} \rho_{c, \eta}(X) d\hat{\nu} = \int_{\text{graph}M} \partial \rho_{c, \eta}(X) d\hat{\nu},
\]

where the last integral is understood as the collection of vectors with each component being the weak* integrals of all weakly* measurable selections of the multifunction \((c, \eta) \mapsto \partial \rho_{c, \eta}(X)\). Applying the representation given in Proposition 3.3, we conclude that
\[
\int_{\text{graph}\Omega} \partial \rho_{c,\eta}(X) d\nu
\]
\[
= \left\{ \int_{\text{graph}\Omega} Q_{c,\eta} d\hat{\nu} : \exists P_{c,\eta} \in \mathcal{A}_{c,\eta}(X), \exists \lambda_{c,\eta}(\omega) \in D_{c,\eta}(X,\omega), Q_{c,\eta} = \int_{\Omega} \lambda_{c,\eta}(\omega) dP_{c,\eta}(\omega) \right\},
\]
(23)

where \( \int_{\text{graph}\Omega} Q_{c,\eta} d\hat{\nu} \) is a vector with each component being a weakly* integral, that is
\[
\{ \int_{\text{graph}\Omega} Q_{c,\eta} d\hat{\nu}, X \} = \int_{\text{graph}\Omega} \{ Q_{c,\eta}, X \} d\hat{\nu}, \forall X \in \mathcal{L}_p^m.
\]

Thus, substituting \( G(\hat{z}) \) for \( X \) in (23), we can obtain the form of \( \hat{Q} \).

Let \( V = H(z) \) and \( X = G(z) \) in (22), we obtain
\[
\langle \hat{S}, H(z) - H(\hat{z}) \rangle + \langle \hat{Q}, G(z) - G(\hat{z}) \rangle \geq 0, \forall z \in Z_0.
\]
(24)

Conversely, as \( P_{c,\eta} \) is a probability measure, it is then known from Lemma 2.3 that every component of \( Q_{c,\eta} \) in (23) is a non-positive measure, which implies that \( \hat{Q} \) is a non-positive vector. Since \( \phi(\cdot) \) is nonincreasing, its subdifferential is non-positive, which infers that \( \hat{S} \) is non-positive. Therefore, by definition of \( E \), (24) implies (22).

By now, we have established an equivalence relationship between (22) and (24).

Utilizing (23) and (24), there exist measures \( P_{c,\eta} \in \mathcal{A}_{c,\eta}(G(\hat{z})) \) and measurable selections \( \lambda_{c,\eta}(\omega) \in D_{c,\eta}(G(\hat{z}), \omega) \) such that
\[
0 \leq \langle \hat{S}, H(z) - H(\hat{z}) \rangle + \int_{\text{graph}\Omega} (Q_{c,\eta}, G(z) - G(\hat{z})) d\hat{\nu}
\]
\[
= \int_{\Omega} [H(z) - H(\hat{z})] \hat{S}(d\omega) + \int_{\text{graph}\Omega} \int_{\Omega} \lambda_{c,\eta}(\omega)(G(z) - G(\hat{z})) P_{c,\eta}(d\omega) d\nu \quad \forall z \in Z_0.
\]
(25)

It follows from (25) that \( \hat{z} \) is an optimal solution of problem (16). Furthermore, using a selection of measures \( P_{c,\eta} \in \mathcal{A}_{c,\eta}(G(\hat{z})) \), we can rewrite the left-hand side of (21) as
\[
\int_{\text{graph}\Omega} \rho_{c,\eta}(G(\hat{z})) d\hat{\nu} = \sup_{P \in \mathcal{Q}} \{ \mathbb{E}_P[(\eta - \theta(c, G(\hat{z})))_+] - \mathbb{E}_P[(\eta - \theta(c, Y))_+] \} d\nu
\]
\[
= \int_{\text{graph}\Omega} \{ \mathbb{E}_{P_{c,\eta}}[(\eta - \theta(c, G(\hat{z})))_+] - \mathbb{E}_{P_{c,\eta}}[(\eta - \theta(c, Y))_+] \} d\nu, \quad \forall z \in Z_0.
\]
Conversely, noticing that all the above equalities still hold, the last conclusion of this theorem follows immediately from the sufficient conditions of optimality for minimizing the Lagrangian function in (19).

Theorem 4.1 extends the optimality conditions established in [23] to the multivariate case with min-biaffine scalarization functions. When \( m = 1, C = \mathbb{R}_+ \) and \( \theta(c,x) = x, \forall c \in C \), Theorem 4.1 reduces to Theorem 1 in [23]. To the best of our knowledge, most results on the multivariate stochastic dominance constraints [17, 18] are based on linear scalarization functions. Here, we establish the optimality condition with min-biaffine scalarization functions, which contains the results with linear scalarization functions as special cases.

5 Discretization and stability analysis

Except for the optimality conditions of problem (14)-(15), we also should investigate its approximation and the related stability properties. These results would lay the foundation for solving problem (14)-(15) numerically. To this end, we first need to specify a proper ambiguity set \( \mathcal{Q} \) of distributions which contains the true probability distribution. In general, \( G(z) \) depends on the decision variable \( z \) and is a random vector in our model. In other words, problem (14)-(15) is a decision dependent stochastic optimization problem, which is extremely hard to solve. We follow the routine in the distributionally robust optimization literature [32] and the robust second-order stochastic optimization problem in [24]. That is, we solve problem (14)-(15) under the condition that the random vector \( \xi \) is independent of the decision variable \( z \). To ensure its tractability and practical application, this treatment is rather reasonable and suitable for many real problems. For example, in portfolio optimization, the randomness of portfolio returns originates from a common factor: the random profit vector. Specifically, we consider a special case of problem (12)-(13) by letting \( G(z) = G(z, \xi) \) where the randomness comes only from \( \xi \):

\[
\min_{z \in Z_0} \phi(H(z)) \\
\text{s.t. } \mathbb{E}_P \left[ \left( \eta - \theta(c, G(z, \xi)) \right)_+ - \left( \eta - \theta(c, Y(\xi)) \right)_+ \right] \leq 0, \forall (c, \eta) \in \text{graph} M, \forall P \in \mathcal{Q},
\]

where \( \xi : \Xi \rightarrow \mathbb{R}^l \) is a random vector, \( \Xi \) is the support set of \( \Omega \). Like that in papers such as [32], we assume that \( \Xi \) is compact. Suppose \( Z_0 \subset \mathbb{R}^n \) is nonempty, convex and compact, \( Y : \mathbb{R}^l \rightarrow \mathbb{R}^m \) is continuous, and \( G : \mathbb{R}^n \times \mathbb{R}^l \rightarrow \mathbb{R}^m \) is concave with respect to \( z \) and continuous with respect to \( (z, \xi) \). We will use \( \mathcal{P} \) to denote the set of all probability measures on \( (\Xi, \mathcal{B}) \), where \( \mathcal{B} \) is the Borel sigma algebra on \( \Xi \). \( \xi \) can denote the random vector or its realization, depending on the context.
Without loss of generality, we assume that \( Q \) is defined via moments. That is, 
\[
Q = \{ P \in \mathcal{P} : \mathbb{E}_P[f(\xi)] \leq 0 \},
\]
where \( f : \Xi \to \mathbb{R}^n \) is a continuous vector-valued functional. Define the discrete approximation to the set of probability distributions as
\[
Q_N := \left\{ \sum_{i=1}^N p_i \delta_{\xi_i} : \sum_{i=1}^N p_i f(\xi_i) \leq 0, \sum_{i=1}^N p_i = 1, p_i \geq 0, i = 1, \ldots, N \right\}.
\]
Obviously \( Q_N \subset Q \). We can now construct an approximation to problem (26) as follows:
\[
\min_{z \in Z_0} \phi(H(z)) \quad \text{s.t.} \quad \sup_{(c,\eta) \in \text{graph} M} \sup_{P \in Q_N} \mathbb{E}_P[h(z,c,\eta,\xi)] \leq 0.
\]

To facilitate exposition, we denote
\[
v_N(z) := \sup_{(c,\eta) \in \text{graph} M} \sup_{P \in Q_N} \mathbb{E}_P[h(z,c,\eta,\xi)],
\]
\[
v(z) := \sup_{(c,\eta) \in \text{graph} M} \sup_{P \in Q} \mathbb{E}_P[h(z,c,\eta,\xi)].
\]

Obviously, \( v_N(z) \leq v(z) \) for \( Q_N \subset Q \). We denote the feasible regions of problems (26) and (27) respectively by 
\( \mathcal{F} := \{ z \in Z_0 : v(z) \leq 0 \} \) and \( \mathcal{F}_N := \{ z \in Z_0 : v_N(z) \leq 0 \} \). Denote the optimal value and the optimal solution set of problem (26) by \( \vartheta := \min \{ \phi(H(z)) : z \in \mathcal{F} \} \) and \( S := \{ z \in \mathcal{F} : \vartheta = \phi(H(z)) \} \). Likewise, let the optimal value and the optimal solution set of problem (27) be 
\( \vartheta_N := \min \{ \phi(H(z)) : z \in \mathcal{F}_N \} \) and \( S_N := \{ z \in \mathcal{F}_N : \vartheta_N = \phi(H(z)) \} \). The aim of this section is to qualitatively analyze the convergence of \( v_N \) to \( \vartheta \), \( \mathcal{F}_N \) to \( \mathcal{F} \) and \( S_N \) to \( S \) when \( N \) tends to infinity.

Like that in papers such as [24,33], we introduce the following natural assumptions.

**Assumption 5.1** (i) There exists a probability measure \( P^* \in \mathcal{P} \) such that \( \mathbb{E}_{P^*}[f(\xi)] < 0 \);

(ii) The sequence \( \{\xi_i\}_{i \in N} \subset \Xi \) satisfies that for any \( \epsilon > 0 \) and \( \xi \in \Xi \), there exists an index \( N' \in \{1, \ldots, N\} \) such that \( ||\xi - \xi^{N'}|| \leq \epsilon \);

(iii) For each \( \xi \in \Xi \), every component of \( G(z,\xi) \), i.e., \( G_i(z,\xi), i = 1, \ldots, m \), is Lipschitz continuous with Lipschitz modulus bounded by \( \kappa(\xi) \), i.e., \( |G_i(z_1,\xi) - G_i(z_2,\xi)| \leq \kappa_i(\xi)||z_1 - z_2||_2 \), and \( \kappa := \sup_{\xi \in \Xi} \sum_{i=1}^m \kappa_i(\xi) \) is finite;
converges to $Q$.

We know from Proof (iv) and the Hausdorff distance under Kantorovich metric $H_K$ is a pseudo-metric.

The similar inequality holds when $\theta(c_1,x) > \theta(c_2,x)$. The proof is complete. \hfill \Box

Before stating our main results about the qualitative stability, we need two metrics: Kantorovich metric and pseudo-metric. The Kantorovich metric of $A_1$ from $A_2$ is denoted by $\mathcal{D}_K(A_1, A_2)$ (refer to [24] for specific definition) and the Hausdorff distance under Kantorovich metric is $H_K(A_1, A_2) := \max\{\mathcal{D}_K(A_1, A_2), \mathcal{D}_K(A_2, A_1)\}$.

We now turn to the pseudo-metric. Denote for simplicity $h(z,c) := (\eta - \theta(c,G(z)))_+$.

Define the set of functions $\mathcal{G} := \{g(\cdot) := h(z,c,\eta,\xi) : z \in Z_0, (c,\eta) \in \text{graph} M\}$. The pseudo-metric is defined as $\mathcal{D}(P, Q) := \sup_{g \in \mathcal{G}}|E_P[g]| - E_Q[g]|$. We define the pseudo-metric between two sets by $\mathcal{D}(A_1, A_2) := \sup\mathcal{P} \sup_{P \in A_1, Q \in A_2} \mathcal{D}(P, Q)$ and $\mathcal{H}(A_1, A_2) := \max\{\mathcal{D}(A_1, A_2), \mathcal{D}(A_2, A_1)\}$.

**Proposition 5.1** Under Assumptions 2.1-5.1, $\mathcal{H}(Q_N, Q)$ tends to zero as $N$ tends to infinity.

*Proof* We know from $Q_N \subset Q$ that $\mathcal{D}(Q_N, Q) = 0$. Then it is sufficient to prove that $\mathcal{D}(Q, Q_N) = 0$. Since $Q_N$ converges to $Q$ under the Kantorovich metric, for any $P \in Q$, there exists a sequence $\{P_N\}$, $P_N \in Q_N$, such that $P_N$ converges to $P$ under the Kantorovich metric. This means that $P_N$ converges to $P$ weakly.

Next, we prove that $P_N$ converges to $P$ under the pseudo-metric. Assume the contrary that there exist a positive number $\delta > 0$ and a sequence $\{z_N, c_N, \eta_N\}$ such that

$$\left|E_{P_N}[h(z_N, c_N, \eta_N, \xi)] - E_P[h(z_N, c_N, \eta_N, \xi)]\right| \geq \delta.$$  \hfill (30)
Because $Z_0 \times \text{graph}\mathcal{Q}$ is compact, without loss of generality, we assume that $(z_N, c_N, \eta_N)$ converges to a point $(z, c, \eta)$, which is still in $Z_0 \times \text{graph}\mathcal{Q}$. By the triangle inequality, we obtain

\[
|\mathbb{E}_{P_N}[h(z_N, c_N, \eta_N, \xi)] - \mathbb{E}_P[h(z, c, \eta, \xi)]| \leq |\mathbb{E}_{P_N}[h(z_N, c_N, \eta_N, \xi)] - \mathbb{E}_{P_N}[h(z, c, \eta, \xi)]| + |\mathbb{E}_{P_N}[h(z, c, \eta, \xi)] - \mathbb{E}_P[h(z, c, \eta, \xi)]|,
\]

For the first term at the right-hand side, we have

\[
|\mathbb{E}_{P_N}[h(z_N, c_N, \eta_N, \xi)] - \mathbb{E}_{P_N}[h(z, c, \eta, \xi)]| \leq \mathbb{E}_{P_N}[(\eta_N - \theta(c_N, G(z_N, \xi)))_+ - (\eta - \theta(c, G(z_N, \xi)))_+] = 2|\eta_N - \eta| + \mathbb{E}_{P_N}[(\theta(c_N, G(z_N, \xi)) - \theta(c, G(z_N, \xi)))] + \mathbb{E}_{P_N}[(\theta(c, G(z_N, \xi)) - \theta(c, G(z_N, \xi)))]
\]

The last but one inequality applies Lemma 2.4. Similarly, we can show that the third term at the right-hand side of (31) also goes to zero. Since $P_N$ converges to $P$ weakly, and $h$ is continuous and bounded with respect to $\xi$, we can easily prove that the second term at the right-hand side of (31) tends to zero, too. All these results contradict the inequality (30).

We will establish some important properties of $v_N$ and $v$. By using the same proof as that for Proposition 3.1 in [24], we have

**Lemma 5.2** $v_N(z)$ converges uniformly to $v(z)$ over $Z_0$ as $N$ tends to infinity.

Furthermore, we establish the Lipschitz continuity of $v$.

**Proposition 5.2** $v(\cdot)$ is Lipschitz continuous on $Z_0$ with the Lipschitz modulus being $C_0\kappa$.

**Proof** For chosen $z \in Z_0$, we denote $V := \{\sup_{(c, \eta) \in \text{graph}\mathcal{Q}} \mathbb{E}_P[h(z, c, \eta, \xi)] : P \in \mathcal{Q}\}$ and

\[
V_N := \{\sup_{(c, \eta) \in \text{graph}\mathcal{Q}} \mathbb{E}_P[h(z, c, \eta, \xi)] : P \in \mathcal{Q}_N\}.\]

Since $\Xi$ is a compact set, both $V$ and $V_N$ are bounded subsets in IR. In what follows, we prove that $V$ is closed. Suppose that $\{v_k\} \subset V$ is a sequence such that $v_k \to v^*$ and $P_k \in \mathcal{Q}$ is such that $\sup_{(c, \eta) \in \text{graph}\mathcal{Q}} \mathbb{E}_{P_k}[h(z, c, \eta, \xi)] = v_k$. We will show $v^* \in V$. Since $\mathcal{Q}$ is weakly
compact, we can assume without loss of generality that $P_k$ converges to $P \in Q$ weakly. Next, we will prove that
\[
\lim_{k \to \infty} \sup_{(c, \eta) \in \text{graph} M} \left| \mathbb{E}_{P_k} [h(z, c, \eta, \xi)] - \mathbb{E}_P [h(z, c, \eta, \xi)] \right| = 0.
\]
We prove by contradiction. Assume that there exist a positive number $\delta > 0$, an infinity set of indexes $K$ and a sequence \{(c_k, \eta_k)\}_{k \in K} \subset \text{graph} M$ such that
\[
\mathbb{E}_{P_k} [h(z, c_k, \eta_k, \xi)] - \mathbb{E}_P [h(z, c_k, \eta_k, \xi)] \geq \delta, \quad \forall k \in K.
\]
Without loss of generality, we assume that $(c_k, \eta_k)$ converges to a point $(c^*, \eta^*) \in \text{graph} M$ as $\text{graph} M$ is compact.

By the triangle inequality,
\[
\left| \mathbb{E}_{P_k} [h(z, c_k, \eta_k, \xi)] - \mathbb{E}_P [h(z, c_k, \eta_k, \xi)] \right| \leq \left| \mathbb{E}_{P_k} [h(z, c_k, \eta_k, \xi)] - \mathbb{E}_{P_k} [h(z, c, \eta, \xi)] \right| + \left| \mathbb{E}_P [h(z, c_k, \eta_k, \xi)] - \mathbb{E}_P [h(z, c, \eta, \xi)] \right|.
\]

For the first term at the right-hand side, we have
\[
\left| \mathbb{E}_{P_k} [h(z, c_k, \eta_k, \xi)] - \mathbb{E}_{P_k} [h(z, c, \eta, \xi)] \right| \\
\leq 2|\eta_k - \eta| + \mathbb{E}_{P_k} [\theta(c_k, G(z, \xi)) - \theta(c, G(z, \xi))] + \mathbb{E}_{P_k} [\theta(c_k, Y(\xi)) - \theta(c, Y(\xi))] \\
\leq 2|\eta_k - \eta| + 2A_0 \|c_k - c\|_2 \to 0
\]

We can similarly show that the third term at the right-hand side of (33) also tends to zero. The second term at the right-hand side of (33) goes to zero too by utilizing the facts that $P_N$ converges to $P$ weakly and that $h$ is bounded and continuous with respect to $\xi$. All these results lead to a contradiction to (32). Therefore, we obtain that
\[
v^* = \lim_{k \to \infty} v_k = \lim_{k \to \infty} \sup_{(c, \eta) \in \text{graph} M} \mathbb{E}_{P_k} [h(z, c, \eta, \xi)] = \sup_{(c, \eta) \in \text{graph} M} \mathbb{E}_P [h(z, c, \eta, \xi)] \in V,
\]
which is equivalent to the fact that $V$ is closed.

For fixed $y \in Z_0$, let $P^* \in \{P \in Q : v(y) = \sup_{(c, \eta) \in \text{graph} M} \mathbb{E}_P [h(y, c, \eta, \xi)]\}$. Because $V$ is compact, $P^*$ exists. Then we have
\[
v(z) \geq \sup_{(c, \eta) \in \text{graph} M} \mathbb{E}_P [h(z, c, \eta, \xi)] \\
\geq \sup_{(c, \eta) \in \text{graph} M} \mathbb{E}_P [h(y, c, \eta, \xi)] - \sup_{(c, \eta) \in \text{graph} M} \left| \mathbb{E}_P [h(z, c, \eta, \xi)] - \mathbb{E}_P [h(y, c, \eta, \xi)] \right|
\]
\[
\geq \sup_{(c, \eta) \in \text{graph} M} \mathbb{E}_{\mathcal{P}^*}[h(y, c, \eta, \xi)] - \sup_{c \in C} C_0(c) \sum_{i=1}^{m} \int_{\Xi} \left| G_i(z, \xi) - G_i(y, \xi) \right| dP^*(\xi) \quad \text{(by Lemma 2.4)}
\]

\[
\geq v(y) - C_0 \kappa ||z - y||_2.
\]

Exchanging the roles of \(y\) and \(z\) completes the proof. \(\square\)

With all the above developed essential properties and results, we can now establish the following qualitative stability results for problem (26).

**Theorem 5.1** Suppose that Assumptions 2.1-5.1 hold and \(F\) is nonempty. Then

(i) \(\lim_{N \to \infty} \mathcal{H}(F_N, F) = 0\);
(ii) \(\lim_{N \to \infty} \vartheta_N = \vartheta\);
(iii) \(\lim_{N \to \infty} \mathcal{D}(S_N, S) = 0\).

The proof is similar to that of Theorem 3.2 in [24], which is thus omitted here.

**Remark 5.1** Our theoretical results extend those results in [24] because the optimization problem induced by our new dominance concept is more generic, which makes the problem-dependent pseudo-metric, optimal value and optimal solutions are much more complex than those in [24].

Theorem 5.1 tells us that problem (27) approximates problem (26) well when \(N\) is large. Therefore, it is reasonable for us to directly deal with the approximation problem (27), which is a convex programming problem with the expectation taken under discrete distribution.

**6 Conclusion**

After analyzing the basic properties of min-biaffine scalarization functions, we used these functions to introduce a new preference relation to compare random vectors, which is called the multivariate robust second-order stochastic dominance. This dominance relation extends many existing concepts of second-order stochastic dominance, including multivariate stochastic dominance and univariate robust stochastic dominance. We reformulated the introduced dominance relation as a functional which is convex, nonincreasing, Lipschitz continuous and subdifferentiable.

We studied the stochastic optimization problem with the proposed multivariate robust second-order stochastic dominance constraints, and developed the necessary and sufficient conditions of optimality in the convex case. Through specifying an ambiguity set and employing a series of discrete distribution sets to approximate the ambiguity set, we not only showed how to approximate the new stochastic optimization problem, but also established the
correspondingly qualitative stability. To this end, we have transformed the initial problem into a convex problem with distributionally robust constraints, where expectations are taken under discrete distributions.

Based on the obtained results, an interesting and natural topic for future research is: how to design concrete numerical algorithms to efficiently solve the stochastic optimization problems with multivariate robust second-order stochastic dominance constraints.

Acknowledgements This research was supported by the National Natural Science Foundation of China (Grant Numbers 71371152 and 11571270).

References