Pacific Journal of Optimization
Vol. 18, No. 4, pp. 737–754, 2022

SAMPLE AVERAGE APPROXIMATION AND PENALTY METHOD FOR A CLASS OF STOCHASTIC MULTIOBJECTIVE BILEVEL PROGRAMS*

Meiju Luo and Gui-Hua Lin†

Abstract: This paper considers a class of stochastic multiobjective bilevel programs in which both the upper-level and the lower-level programs are multiobjective. By introducing auxiliary variables to the optimality conditions of the lower-level program, we transform the bilevel program into a stochastic multiobjective program with scalar variational inequality constraints. We employ some penalty function and some sample average approximation techniques to present an approximation method. Then, we give a comprehensive convergence analysis for the proposed method.

Key words: *stochastic multiobjective bilevel program, sample average approximation, penalty function, convergence*

Mathematics Subject Classification: *90C33, 90C15*

---

1 Introduction

The origin of bilevel programs can be traced back to the work [27] which used to model the market economy. Nowadays, bilevel programs have become an important branch in the optimization field and attracted more and more attention. Many applications of bilevel programs can be found in network design, transport system planning, management, economics, and so on [28, 14, 22, 29]. However, due to their hierarchical structures, bilevel programs are very difficult to solve and even the linear cases are known to be NP-hard [4]. For recent developments on numerical methods for bilevel programs, we refer the reader to [3, 10, 7, 23, 32] and the references therein.

Up to now, most researches on bilevel programs mainly focus on the single objective cases, while the multiobjective bilevel programs have important applications in practice [1, 20, 33, 16, 2]. In this paper, we consider the following stochastic multiobjective bilevel program (SMBLP):

$$\begin{aligned} \min \quad & \mathbf{E}[F(x,y,\xi)] \\ \text{s.t.} \quad & x \in X,\ y \in S(x), \end{aligned} \tag{1.1}$$

*This work was supported in part by Natural Science Foundation Guiding Project of Liaoning Provincial Department of Science and Technology (No. 2019-ZD-0194), Natural Science Foundation General Project of Liaoning Provincial Department of Science and Technology (No. 2021-MS-153), Scientific Research Funding Project of Liaoning Provincial Education Department (No. LJC201902), and National Natural Science Foundation of China (No. 12071280).

†Corresponding author.

where $S(x)$ denotes the solution set of the lower-level program

$$\min_{y} \quad \mathbf{E}[f(x,y,\xi)] \qquad \text{s.t.} \quad y \in Y. \tag{1.2}$$

Here, $\mathbf{E}$ denotes the mathematical expectation operator related to the stochastic variable ξ defined on the probability space $(\Omega, \mathbf{F}, P)$ with a support set Ξ, X is a nonempty closed subset of R^n, Y is a nonempty closed convex subset of R^m, $F : R^{n+m} \times \Omega \to R^p$ and $f : R^{n+m} \times \Omega \to R^q$ are vector-valued random functions. Suppose that each objective function f_i $(i = 1, \ldots, q)$ is convex with respect to the lower-level variable. Note that, although the lower-level constraint set Y does not depend on x, it is still widely used in practice [22].

The multiobjective bilevel programs have several cases. The first is that the upper-level is single objective but the lower-level is multiobjective. This kind of problems was originally considered in Hausdorff topological spaces in [5], in which they were called semivectorial bilevel optimization problems and the authors suggested an exterior penalty approach to handle them. Afterwards, Dempe et al. [9] derived some first-order necessary optimality conditions for this type of problems. The second is that the upper-level is multiobjective but the lower-level is single objective. By using the Karush-Kuhn-Tucker (KKT) conditions of the lower-level programs, this kind of problems may be formulated into multiobjective programs with complementarity constraints and, along this way, some numerical methods and optimality conditions were respectively presented [31, 30, 8]. The third is that both the upper-level and the lower-level programs are multiobjective; see, e.g., [21, 26] for theoretical results and algorithms for this type of bilevel programs.

There are also a few works to study the stochastic bilevel programs. In particular, Kosuch et al. [18] and Kovacevic et al. [19] considered the applications of the stochastic single objective bilevel programs in practical pricing, service provision, and swing option pricing. Chen et al. [6] and Lin et al. [20] discussed optimality conditions and numerical methods for the stochastic multiobjective bilevel programs, in which the upper-level programs are multiobjective but the lower-level programs are single objective, and their applications in network design and healthcare management.

In summarize, the SMBLP (1.1)–(1.2) considered in this paper is more general and, to the best of our knowledge, it has not been studied so far. Since a convex multiobjective program is equivalent to a vector variational inequality, (1.1)–(1.2) is generally transformed into a stochastic multiobjective program with vector variational inequality constraints. We consider a different way in this paper. That is, we use some kind of necessary and sufficient conditions for the lower-level multiobjective program to transform (1.1)–(1.2) into a stochastic multiobjective program with scalar variational inequality constraints and then employ a penalty function and some sample average approximation (SAA) technique to present an SAA-based penalty method.

The rest of this paper is organized as follows. In Section 2, we first introduce how to reformulate the SMBLP to a multiobjective program with scalar variational inequality constraint and then present an SAA-based penalty approximation method by using some penalty function and some SAA technique. In Section 3, we investigate the limiting behavior of weak Pareto optimal solutions of the SAA-based penalty approximation problems and, in Section 4, we consider the convergence properties of Pareto stationary points of the approximation problems. Finally, in Section 5, we make some concluding remarks.

Throughout, we adopt the following standard notations. For a given nonempty set $D \subseteq R^n$, $\text{int}(D)$ denotes its interior. For a given real-valued function $\psi : R^n \times R^m \to R$, $\nabla_x \psi(x,y)$

and $\nabla_x^2\psi(x,y)$ denote its gradient and Hessian matrix respectively at x, $\nabla_{xy}\psi(x,y)$ and $\nabla_{xy}^2\psi(x,y)$ denote its gradient and Hessian matrix respectively at (x,y). For a given vector-valued function $\Psi : R^n \times R^t \to R^m$, $\nabla_x\Psi(x,y) \in R^{m\times n}$ denotes its Jacobian matrix at x, $\nabla_{xy}\Psi(x,y) \in R^{m\times(n+t)}$ denotes its Jacobian matrix at (x,y). Moreover, $\|\cdot\|$ stands for the 2-norm and $\|\cdot\|_F$ denotes the Frobenius norm, whereas $\mathrm{Proj}_Y(x)$ stands for the projection of a point x onto the closed convex set Y. For $a \in R^n$, $b \in R^n$, $a \prec b$ means $a_i < b_i$ for each $i = 1,\ldots,n$.

2 SAA-Based Penalty Approximation Method

By [17], $\bar{y}$ is a weak Pareto optimal solution of (1.2) if and only if it solves the vector variational inequalities of finding $\bar{y} \in Y$ such that

$$\nabla_y \mathbf{E}[f(x,\bar{y},\xi)](z-\bar{y}) \notin -\mathrm{int}R_+^q, \quad \forall z \in Y,$$

which is equivalent to

$$\big((z-\bar{y})^T\nabla_y\mathbf{E}[f_1(x,\bar{y},\xi)],\ldots,(z-\bar{y})^T\nabla_y\mathbf{E}[f_q(x,\bar{y},\xi)]\big)^T \notin -\mathrm{int}R_+^q, \quad \forall z \in Y. \tag{2.1}$$

For simplicity, we denote by $\mathbf{f}_j(x,y,\xi) := \nabla_y f_j(x,y,\xi)$ $(j=1,\ldots,q)$. By Theorem 16.8 of [24] and Assumption (A2) given in the next section, we have $\nabla_y\mathbf{E}[f_j(x,y,\xi)] = \mathbf{E}[\mathbf{f}_j(x,y,\xi)]$ $(j=1,\ldots,q)$. By Theorem 2.1 in [34], the above vector variational inequalities (2.1) reduce to the scalar variational inequalities of finding $(\bar{y},\bar{\lambda}) \in Y \times \Lambda$ such that

$$(z-\bar{y})^T\sum_{j=1}^q \bar{\lambda}_j\mathbf{E}[\mathbf{f}_j(x,\bar{y},\xi)] \geq 0, \quad \forall z \in Y, \tag{2.2}$$

where $\Lambda := \{\lambda \in R^q | \lambda \geq 0, \sum_{j=1}^q \lambda_j = 1\}$. Thus, since (1.2) is assumed to be a convex program, the SMBLP (1.1)–(1.2) can be transformed into the stochastic multiobjective program with scalar variational inequality constraints

$$\begin{aligned} \min \quad & \mathbf{E}[F(x,y,\xi)] \\ \text{s.t.} \quad & (z-y)^T\sum_{j=1}^q \lambda_j\mathbf{E}[\mathbf{f}_j(x,y,\xi)] \geq 0, \quad \forall z \in Y, \\ & (x,y,\lambda) \in X \times Y \times \Lambda. \end{aligned} \tag{2.3}$$

To deal with the scalar variational inequality constraints in (2.3), similarly as in [12], we define the regularized gap function $h : X \times Y \times \Lambda \to R$ as

$$h(x,y,\lambda) := \max_{z\in Y}\Big\{(y-z)^T\sum_{j=1}^q \lambda_j\mathbf{E}[\mathbf{f}_j(x,y,\xi)] - \frac{\alpha}{2}\|y-z\|^2\Big\}, \tag{2.4}$$

where $\alpha > 0$ is a given parameter. It follows from [12] that, for any $(y,\lambda) \in Y \times \Lambda$,

$$h(x,y,\lambda) = \big(y-H(x,y,\lambda)\big)^T\sum_{j=1}^q \lambda_j\mathbf{E}[\mathbf{f}_j(x,y,\xi)] - \frac{\alpha}{2}\big\|y-H(x,y,\lambda)\big\|^2,$$

where $H(x,y,\lambda) := \mathrm{Proj}_Y\big(y-\alpha^{-1}\sum_{j=1}^q \lambda_j\mathbf{E}[\mathbf{f}_j(x,y,\xi)]\big)$. Moreover, by Theorem 3.1 in [12], we have $h(x,y,\lambda) \geq 0$ for each $(x,y,\lambda) \in X \times Y \times \Lambda$ and $h(x,y,\lambda) = 0$ if and only if (x,y,λ)

solves (2.2). Therefore, the SMBLP (1.1)–(1.2) is further equivalent to

$$\begin{array}{ll} \min & \mathbf{E}[F(x,y,\xi)] \\ \text{s.t.} & h(x,y,\lambda)=0, \\ & (x,y,\lambda)\in X\times Y\times\Lambda. \end{array} \tag{2.5}$$

From now on, we denote by $\mathcal{F}$ the feasible region of (2.5).

We now present our approximation method for (2.5). Although the variational inequality constraint in (2.3) is transformed into the equality constraint in (2.5), it is still the essential difficulty in solving (2.5). Here, we employ a penalty function to get the following approximation problem of (2.5):

$$\begin{array}{ll} \min & \Theta(x,y,\lambda):=\mathbf{E}[F(x,y,\xi)]+\rho h^2(x,y,\lambda)\mathbf{e} \\ \text{s.t.} & (x,y,\lambda)\in X\times Y\times\Lambda, \end{array} \tag{2.6}$$

where $\rho>0$ is a penalty factor and $\mathbf{e}:=(1,1,\ldots,1)^T\in R^p$.

Note that for a given function $g:R^n\to R^n$, set

$$G(x)=\max_{z\in Y}\{g(x)^T(x-z)\}, \tag{2.7}$$

by [13] page 3, when the maximizer on the right-hand side of (2.7) is unique, the gradient of G at x is actually given by

$$\nabla_x G(x)=g(x)-\nabla_x g(x)^T(z-x). \tag{2.8}$$

In addtion, by Theorem 3.2 in [12], if $\mathbf{E}[\mathbf{f}_j(x,y,\xi)]$ $(j=1,\ldots,q)$ is continuously differentiable, $h(x,y,\lambda)$ is continuously differentiable. Since the maximizer of formulation (2.4) is unique, we use the formulation (2.8) to obtain the gradient as follows:

$$\nabla_{xy\lambda}h(x,y,\lambda) = \begin{pmatrix} \sum_{j=1}^q \lambda_j\nabla_x\mathbf{E}[\mathbf{f}_j(x,y,\xi)]^T\Big(y-H(x,y,\lambda)\Big) \\ \sum_{j=1}^q \lambda_j\mathbf{E}[\mathbf{f}_j(x,y,\xi)]-\big(\sum_{j=1}^q \lambda_j\nabla_y\mathbf{E}[\mathbf{f}_j(x,y,\xi)]-\alpha I\big)\big(H(x,y,\lambda)-y\big) \\ \mathbf{E}[\mathbf{f}(x,y,\xi)]^T\big(y-H(x,y,\lambda)\big) \end{pmatrix}, \tag{2.9}$$

where I denotes the unit matrix in appropriate dimension, $\mathbf{E}[\mathbf{f}(x,y,\xi)]=\big[\mathbf{E}[\mathbf{f}_1(x,y,\xi)],\ldots,\mathbf{E}[\mathbf{f}_q(x,y,\xi)]\big]$. Recalling that each $f_i(x,y,\xi)$ $(i=1,\ldots,q)$ is convex with respect to y, we have the following result.

Theorem 2.1. *$(\bar{x},\bar{y})$ is a weak Pareto optimal solution of (1.1)–(1.2) if and only if there exists $\bar{\lambda}\in\Lambda$ such that $(\bar{x},\bar{y},\bar{\lambda})$ is a weak Pareto optimal solution of (2.5).*

Proof. Suppose that $(\bar{x},\bar{y})$ is a weak Pareto optimal solution of (1.1)–(1.2). It follows that $\bar{y}$ is a weak Pareto optimal solution of (1.2). Since (1.2) is a convex program, there exists $\bar{\lambda}\in\Lambda$ satisfying (2.2), from which we have $h(\bar{x},\bar{y},\bar{\lambda})=0$ and hence $(\bar{x},\bar{y},\bar{\lambda})\in\mathcal{F}$. Since $(\bar{x},\bar{y})$ is a weak Pareto optimal solution of (1.1)–(1.2), there does not exist $(x,y)\in X\times Y$ such that $\mathbf{E}[F(x,y,\xi)]\prec\mathbf{E}[F(\bar{x},\bar{y},\xi)]$. Therefore, $(\bar{x},\bar{y},\bar{\lambda})$ is a weak Pareto optimal solution of (2.5).

Conversely, suppose that $(\bar{x},\bar{y},\bar{\lambda})$ is a weak Pareto optimal solution of (2.5). We have $h(\bar{x},\bar{y},\bar{\lambda})=0$ and there does not exist $(x,y,\lambda)\in\mathcal{F}$ such that $\mathbf{E}[F(x,y,\xi)]\prec\mathbf{E}[F(\bar{x},\bar{y},\xi)]$.

Hence, $(\bar{y}, \bar{\lambda})$ is a solution of (2.2). In addition, since (1.2) is a convex program, $(\bar{x}, \bar{y})$ is feasible to (1.1). If $(\bar{x}, \bar{y})$ is not a weak Pareto optimal solution of (1.1)–(1.2), there must exist $(\tilde{x}, \tilde{y})$ satisfying the constraints of (1.1)–(1.2) and

$$\mathbf{E}[F(\tilde{x}, \tilde{y}, \xi)] \prec \mathbf{E}[F(\bar{x}, \bar{y}, \xi)]. \tag{2.10}$$

Since $\tilde{y}$ is a weak Pareto optimal solution of (1.2) and (1.2) is a convex program, there exists $\tilde{\lambda} \in \Lambda$ satisfying (2.2). So, we have $h(\tilde{x}, \tilde{y}, \tilde{\lambda}) = 0$ and hence $(\tilde{x}, \tilde{y}, \tilde{\lambda}) \in \mathcal{F}$. Combing with (2.10), this contradicts the fact that $(\bar{x}, \bar{y}, \bar{\lambda})$ is a weak Pareto optimal solution of (2.5). Therefore, $(\bar{x}, \bar{y})$ must be a weak Pareto optimal solution of (1.1)–(1.2). □

Note that the objective function in (2.6) contains a mathematical expectation, which is usually difficult to compute in practice. We employ the SAA techniques to deal with expectations. Recall that, given an integrable function $\psi : \Omega \to R$, the Monte Carlo sampling estimate for $\mathbf{E}[\psi(\xi)]$ is obtained by taking independently and identically distributed random samples $\Omega_k := \{\xi^1, \ldots, \xi^{N_k}\}$ from Ω, where $N_k \to +\infty$ as $k \to +\infty$, and letting $\mathbf{E}[\psi(\xi)] \approx \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \psi(\xi^i)$. The strong law of large numbers guarantees that this procedure converges with probability one (abbreviated by w.p.1) [25], that is,

$$\lim_{k \to +\infty} \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \psi(\xi^i) = \mathbf{E}[\psi(\xi)], \quad \text{w.p.1.} \tag{2.11}$$

Thus, by taking some independently and identically distributed samples $\xi^1, \ldots, \xi^{N_k}$ from Ω_k, we get the following SAA-based penalty approximation problem of (2.6):

$$\begin{aligned} \min \quad & \Theta^k(x, y, \lambda) := \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} F(x, y, \xi^i) + \rho^k \big(h^k(x, y, \lambda)\big)^2 \mathbf{e} \\ \text{s.t.} \quad & (x, y, \lambda) \in X \times Y \times \Lambda, \end{aligned} \tag{2.12}$$

where

$$h^k(x, y, \lambda) := \big(y - H^k(x, y, \lambda)\big)^T \sum_{j=1}^{q} \lambda_j \Big(\frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \mathbf{f}_j(x, y, \xi^i)\Big) - \frac{\alpha}{2} \big\| y - H^k(x, y, \lambda) \big\|^2 \tag{2.13}$$

with $H^k(x, y, \lambda) := \mathrm{Proj}_Y \Big(y - \alpha^{-1} \sum_{j=1}^{q} \lambda_j \big(\frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \mathbf{f}_j(x, y, \xi^i) \big) \Big)$ and $\rho^k > 0$ being a penalty factor. We discuss the limiting behavior of this approximation approach in the subsequent sections.

3 Convergence Analysis of Weak Pareto Optimal Solutions

In this section, we investigate the convergence properties of the weak pareto optimal solutions of the approximation problem (2.12). We first make some assumptions.

(A1) For every $\xi \in \Omega$, each $F_i(x, y, \xi)$ $(i = 1, \ldots, p)$ and each $f_i(x, y, \xi)$ $(i = 1, \ldots, q)$ are twice continuously differentiable with respect to (x, y) on $X \times Y$.

(A2) There exist $\varrho > 0$ and an integrable function $\mathcal{K}(\xi)$ such that $\mathbf{E}[\mathcal{K}(\xi)] < +\infty$, $\mathbf{E}[\mathcal{K}^2(\xi)] < +\infty$, and

$$\begin{aligned}
&\|\nabla^2_{xy} f_i(x,y,\xi) - \nabla^2_{xy} f_i(\bar{x},\bar{y},\xi)\|_F \le \mathcal{K}(\xi)\|(x,y)-(\bar{x},\bar{y})\|^{\varrho},\\
&\max\Big\{\sum_{j=1}^{p}|F_j(x,y,\xi)| + \sum_{j=1}^{p}\|\nabla_{xy}F_j(x,y,\xi)\| + \sum_{j=1}^{p}\|\nabla^2_{xy}F_j(x,y,\xi)\|_F,\\
&\quad \sum_{j=1}^{q}|f_j(x,y,\xi)| + \sum_{j=1}^{q}\|\nabla_{xy}f_j(x,y,\xi)\| + \sum_{j=1}^{q}\|\nabla_{xy}\mathbf{f}_j(x,y,\xi)\|_F\Big\} \le \mathcal{K}(\xi)
\end{aligned}$$

for any $(x,y) \in X \times Y$, $(\bar{x},\bar{y}) \in X \times Y$ and almost every $\xi \in \Omega$.

(A3) The penalty factor ρ^k is taken to satisfy $\lim_{k\to+\infty}\rho^k = +\infty$ and

$$\lim_{k\to+\infty}\rho^k\big(h^k(x,y,\lambda) - h(x,y,\lambda)\big) = 0, \quad \forall (x,y,\lambda) \in X \times Y \times \Lambda, \tag{3.1}$$

in along with

$$\lim_{k\to+\infty}\rho^k\|(x^k,y^k,\lambda^k) - (\bar{x},\bar{y},\bar{\lambda})\| = 0 \quad \text{if} \quad \lim_{k\to+\infty}(x^k,y^k,\lambda^k) = (\bar{x},\bar{y},\bar{\lambda}). \tag{3.2}$$

Here, h and h^k are given in (2.4) and (2.13) respectively.

Remark 3.1. The convergence in (2.11) is of order $O(k^{-\frac{1}{2}})$ with probability one [15], which implies that the sequence

$$\Big\{\sqrt{k}\big(h^k(x,y,\lambda) - h(x,y,\lambda)\big)\Big\} \tag{3.3}$$

is convergence with probability one as $k \to +\infty$. Therefore, we may set $\rho^k = k^{\nu}$ with $\nu \in (0, \frac{1}{2})$. It is easy to see that (3.1) holds by (3.3). On the other hand, under the condition that $(\bar{x},\bar{y},\bar{\lambda})$ is an accumulation point of (x^k,y^k,λ^k), there at least exists a sub-column of (x^k,y^k,λ^k) satisfying (3.2).

Lemma 3.2. *Let* (A1) – (A2) *hold and* $\lim_{k\to+\infty}(x^k,y^k,\lambda^k) = (\bar{x},\bar{y},\bar{\lambda})$. *Then, we have* $\lim_{k\to+\infty} h^k(x^k,y^k,\lambda^k) = h(\bar{x},\bar{y},\bar{\lambda})$ *with probability one.*

Proof. By the mean value theorem and (A2), we have

$$\begin{aligned}
&\big\|\mathbf{f}_j(x^k,y^k,\xi^i) - \mathbf{f}_j(\bar{x},\bar{y},\xi^i)\big\|\\
=\ &\Big\|\int_0^1 \Big(\nabla_{xy}\mathbf{f}_j\big(tx^k + (1-t)\bar{x}, ty^k + (1-t)\bar{y}, \xi^i\big)\Big)^T \big((x^k,y^k) - (\bar{x},\bar{y})\big)dt\Big\|\\
\le\ &\mathcal{K}(\xi^i)\big\|(x^k,y^k) - (\bar{x},\bar{y})\big\|, \quad \forall j = 1,\dots,q.
\end{aligned} \tag{3.4}$$

We then have from the nonexpansive property of Proj_Y, (A2) and (3.4) that

$$
\begin{aligned}
&\big\|H^k(x^k,y^k,\lambda^k)-H^k(\bar{x},\bar{y},\bar{\lambda})\big\|\\
&\le \bigg\|\Big(x^k-\alpha^{-1}\sum_{j=1}^{q}\lambda_j^k\big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(x^k,y^k,\xi^i)\big)\Big)-\Big(\bar{x}-\alpha^{-1}\sum_{j=1}^{q}\bar{\lambda}_j\big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(\bar{x},\bar{y},\xi^i)\big)\Big)\bigg\|\\
&\le \|x^k-\bar{x}\|+\alpha^{-1}\Big[\sum_{j=1}^{q}\lambda_j^k\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\big\|\mathbf{f}_j(x^k,y^k,\xi^i)-\mathbf{f}_j(\bar{x},\bar{y},\xi^i)\big\|\\
&\quad+\sum_{j=1}^{q}|\lambda_j^k-\bar{\lambda}_j|\cdot\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\big\|\mathbf{f}_j(\bar{x},\bar{y},\xi^i)\big\|\Big]\\
&\le \|x^k-\bar{x}\|\\
&\quad+\alpha^{-1}\Big[\sum_{j=1}^{q}\lambda_j^k\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathcal{K}(\xi^i)\big\|(x^k,y^k)-(\bar{x},\bar{y})\big\|+\sum_{j=1}^{q}|\lambda_j^k-\bar{\lambda}_j|\cdot\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathcal{K}(\xi^i)\Big]\\
&\to 0 \;\text{ as }\; k\to+\infty \quad w.p.1.
\end{aligned}
\tag{3.5}
$$

Since $h^k(x,y,\lambda)\ge 0$ for any $(x,y,\lambda)\in X\times Y\times\Lambda$, we have

$$
\big\|y-H^k(x,y,\lambda)\big\|\le\frac{2}{\alpha}\sum_{j=1}^{q}\lambda_j\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\big\|\mathbf{f}_j(x,y,\xi^i)\big\|\le\frac{2}{\alpha}\cdot\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathcal{K}(\xi^i). \tag{3.6}
$$

Hence, by the inequality $\|a\|-\|b\|\le\|a-b\|$ and (3.6), we have

$$
\big\|H^k(x,y,\lambda)\big\|\le\frac{2}{\alpha}\cdot\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathcal{K}(\xi^i)+\|y\|. \tag{3.7}
$$

It follows from (A2) and (3.5)–(3.6) that

$$
\begin{aligned}
&\Big|\big\|y^k-H^k(x^k,y^k,\lambda^k)\big\|^2-\big\|\bar{y}-H^k(\bar{x},\bar{y},\bar{\lambda})\big\|^2\Big|\\
&\le\Big|\big\|y^k-H^k(x^k,y^k,\lambda^k)\big\|+\big\|\bar{y}-H^k(\bar{x},\bar{y},\bar{\lambda})\big\|\Big|\\
&\qquad\qquad\qquad\cdot\Big|\|y^k-\bar{y}\|+\big\|H^k(x^k,y^k,\lambda^k)-H^k(\bar{x},\bar{y},\bar{\lambda})\big\|\Big|\\
&\le\frac{4}{\alpha}\cdot\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathcal{K}(\xi^i)\cdot\Big|\|y^k-\bar{y}\|+\big\|H^k(x^k,y^k,\lambda^k)-H^k(\bar{x},\bar{y},\bar{\lambda})\big\|\Big|\\
&\to 0 \;\text{ as }\; k\to+\infty \quad w.p.1.
\end{aligned}
\tag{3.8}
$$

Moreover, by (A2) and (3.4), there holds

$$\begin{aligned}
&\Big\|\sum_{j=1}^{q}\lambda_j^k\big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(x^k,y^k,\xi^i)\big)-\sum_{j=1}^{q}\bar{\lambda}_j\big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(\bar{x},\bar{y},\xi^i)\big)\Big\|\\
\leq\quad&\sum_{j=1}^{q}|\lambda_j^k-\bar{\lambda}_j|\cdot\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\big\|\mathbf{f}_j(x^k,y^k,\xi^i)\big\|\\
&\qquad\qquad+\sum_{j=1}^{q}\bar{\lambda}_j\cdot\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\big\|\mathbf{f}_j(x^k,y^k,\xi^i)-\mathbf{f}_j(\bar{x},\bar{y},\xi^i)\big\|\\
\leq\quad&\sum_{j=1}^{q}|\lambda_j^k-\bar{\lambda}_j|\cdot\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathcal{K}(\xi^i)+\sum_{j=1}^{q}\bar{\lambda}_j\cdot\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathcal{K}(\xi^i)\big\|(x^k,y^k)-(\bar{x},\bar{y})\big\|\\
\to\quad&0\ \text{ as }\ k\to+\infty\quad w.p.1.
\end{aligned}\tag{3.9}$$

It then follows from (A2) and (3.9) that

$$\begin{aligned}
&\Big|{y^k}^T\sum_{j=1}^{q}\lambda_j^k\big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(x^k,y^k,\xi^i)\big)-\bar{y}^T\sum_{j=1}^{q}\bar{\lambda}_j\big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(\bar{x},\bar{y},\xi^i)\big)\Big|\\
\leq\quad&\|y^k-\bar{y}\|\cdot\sum_{j=1}^{q}\lambda_j^k\Big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\big\|\mathbf{f}_j(x^k,y^k,\xi^i)\big\|\Big)\\
&\quad+\|\bar{y}\|\cdot\Big\|\sum_{j=1}^{q}\lambda_j^k\big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(x^k,y^k,\xi^i)\big)-\sum_{j=1}^{q}\bar{\lambda}_j\big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(\bar{x},\bar{y},\xi^i)\big)\Big\|\\
\to\quad&0\ \text{ as }\ k\to+\infty\quad w.p.1.
\end{aligned}\tag{3.10}$$

Therefore, by (A2), (3.5), (3.7) and (3.9), we have

$$\begin{aligned}
&\Big|\big(H^k(x^k,y^k,\lambda^k)\big)^T\sum_{j=1}^{q}\lambda_j^k\Big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(x^k,y^k,\xi^i)\Big)-\big(H^k(\bar{x},\bar{y},\bar{\lambda})\big)^T\\
&\qquad\qquad\qquad\qquad\sum_{j=1}^{q}\bar{\lambda}_j\Big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(\bar{x},\bar{y},\xi^i)\Big)\Big|\\
&\leq\big\|H^k(x^k,y^k,\lambda^k)-H^k(\bar{x},\bar{y},\bar{\lambda})\big\|\Big\|\sum_{j=1}^{q}\lambda_j^k\Big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(x^k,y^k,\xi^i)\Big)\Big\|\\
&\quad+\Big\|\Big(H^k(\bar{x},\bar{y},\bar{\lambda})\Big)\Big\|\Big\|\sum_{j=1}^{q}\lambda_j^k\Big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(x^k,y^k,\xi^i)\Big)-\sum_{j=1}^{q}\bar{\lambda}_j\Big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(\bar{x},\bar{y},\xi^i)\Big)\Big\|\\
&\leq\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathcal{K}(\xi^i)\big\|H^k(x^k,y^k,\lambda^k)-H^k(\bar{x},\bar{y},\bar{\lambda})\big\|+\Big[\frac{2}{\alpha}\cdot\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathcal{K}(\xi^i)+\|\bar{y}\|\Big]\\
&\quad\cdot\Big\|\sum_{j=1}^{q}\lambda_j^k\Big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(x^k,y^k,\xi^i)\Big)-\sum_{j=1}^{q}\bar{\lambda}_j\Big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(\bar{x},\bar{y},\xi^i)\Big)\Big\|\\
&\to 0\ \text{ as }\ k\to+\infty\quad w.p.1.
\end{aligned}\tag{3.11}$$

In consequence, by (2.11), (3.8) and (3.10)–(3.11), there holds

$$\begin{aligned}
&\big|h^k(x^k,y^k,\lambda^k)-h(\bar{x},\bar{y},\bar{\lambda})\big| \\
\leq\ & \Big|\big(y^k-H^k(x^k,y^k,\lambda^k)\big)^T\sum_{j=1}^{q}\lambda_j^k\big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(x^k,y^k,\xi^i)\big)-\big(\bar{y}-H^k(\bar{x},\bar{y},\bar{\lambda})\big)^T \\
&\cdot\sum_{j=1}^{q}\bar{\lambda}_j\big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(\bar{x},\bar{y},\xi^i)\big)\Big|+\frac{\alpha}{2}\Big|\big\|y^k-H^k(x^k,y^k,\lambda^k)\big\|^2-\big\|\bar{y}-H^k(\bar{x},\bar{y},\bar{\lambda})\big\|^2\Big| \\
&+\Big|\big(\bar{y}-H^k(\bar{x},\bar{y},\bar{\lambda})\big)^T\sum_{j=1}^{q}\bar{\lambda}_j\big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(\bar{x},\bar{y},\xi^i)\big)-\big(\bar{y}-H(\bar{x},\bar{y},\bar{\lambda})\big)^T \\
&\cdot\sum_{j=1}^{q}\bar{\lambda}_j\mathbf{E}[\mathbf{f}_j(\bar{x},\bar{y},\xi)]\Big|+\frac{\alpha}{2}\Big|\big\|\bar{y}-H^k(\bar{x},\bar{y},\bar{\lambda})\big\|^2-\big\|\bar{y}-H(\bar{x},\bar{y},\bar{\lambda})\big\|^2\Big| \\
\to\ & 0 \ \text{ as } \ k\to+\infty \quad w.p.1.
\end{aligned}$$

Therefore, the conclusion holds. □

Next, we show our main convergence result. We denote by $\bar{\mathcal{O}}$ and $\mathcal{O}_k$ the sets of weak Pareto optimal solutions of problems (2.5) and (2.12) respectively.

Theorem 3.3. *Suppose that* (A1) – (A3) *hold,* $(x^k,y^k,\lambda^k)\in\mathcal{O}_k$ *for each* k*, and* $(\bar{x},\bar{y},\bar{\lambda})$ *is an accumulation point of* $\{(x^k,y^k,\lambda^k)\}$*. Then, we have* $(\bar{x},\bar{y},\bar{\lambda})\in\bar{\mathcal{O}}$ *with probability one.*

Proof. Without loss of generality, we assume that $\lim_{k\to\infty}(x^k,y^k,\lambda^k)=(\bar{x},\bar{y},\bar{\lambda})$. It follows from the mean value theorem and (A2) that, for each (x^k,y^k,ξ^i), there exists $(x^{k_i},y^{k_i})=\gamma^{k_i}(x^k,y^k)+(1-\gamma^{k_i})(\bar{x},\bar{y})$ with $\gamma^{k_i}\in(0,1)$ such that

$$\begin{aligned}
\big|F_j(x^k,y^k,\xi^i)-F_j(\bar{x},\bar{y},\xi^i)\big| &= \Big|\nabla_{xy}F_j(x^{k_i},y^{k_i},\xi^i)^T\big((x^k,y^k)-(\bar{x},\bar{y})\big)\Big| \\
&\leq \mathcal{K}(\xi^i)\big\|(x^k,y^k)-(\bar{x},\bar{y})\big\|, \quad \forall j=1,\ldots,p. \qquad (3.12)
\end{aligned}$$

Thus, we have from (A2), (2.11) and (3.12) that

$$\begin{aligned}
&\Big|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}F_j(x^k,y^k,\xi^i)-\mathbf{E}[F_j(\bar{x},\bar{y},\xi)]\Big| \\
\leq\ & \frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\big|F_j(x^k,y^k,\xi^i)-F_j(\bar{x},\bar{y},\xi^i)\big|+\Big|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}F_j(\bar{x},\bar{y},\xi^i)-\mathbf{E}[F_j(\bar{x},\bar{y},\xi)]\Big| \\
\leq\ & \frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathcal{K}(\xi^i)\big\|(x^k,y^k)-(\bar{x},\bar{y})\big\|+\Big|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}F_j(\bar{x},\bar{y},\xi^i)-\mathbf{E}[F_j(\bar{x},\bar{y},\xi)]\Big| \\
\to\ & 0 \ \text{ as } \ k\to+\infty \quad w.p.1, \quad \forall j=1,\ldots,p. \qquad (3.13)
\end{aligned}$$

Since $(x^k,y^k,\lambda^k)\in\mathcal{O}_k$, there does not exist $(x,y,\lambda)\in X\times Y\times\Lambda$ such that

$$\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}F_j(x,y,\xi^i)+\rho^k\big(h^k(x,y,\lambda)\big)^2<\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}F_j(x^k,y^k,\xi^i)+\rho^k\big(h^k(x^k,y^k,\lambda^k)\big)^2$$

for each $j=1,\ldots,p$. Let $(\tilde{x},\tilde{y},\tilde{\lambda})\in\mathcal{F}\subseteq X\times Y\times\Lambda$ be arbitrarily fixed. Then, there at least exists one j_k such that

$$\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}F_{j_k}(\tilde{x},\tilde{y},\xi^i)+\rho^k\big(h^k(\tilde{x},\tilde{y},\tilde{\lambda})\big)^2\geq\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}F_{j_k}(x^k,y^k,\xi^i)+\rho^k\big(h^k(x^k,y^k,\lambda^k)\big)^2.$$

Taking a further subsequence if necessary, we may assume that there is an index j_0 such that

$$\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_{j_0}(\tilde{x},\tilde{y},\xi^i)+\rho^k\big(h^k(\tilde{x},\tilde{y},\tilde{\lambda})\big)^2 \geq \frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_{j_0}(x^k,y^k,\xi^i)+\rho^k\big(h^k(x^k,y^k,\lambda^k)\big)^2. \tag{3.14}$$

Hence, by (2.11) and (3.13), we have

$$\begin{aligned} & \rho^k\big(h^k(x^k,y^k,\lambda^k)\big)^2-\rho^k\big(h^k(\tilde{x},\tilde{y},\tilde{\lambda})\big)^2 \\ \leq\quad & \frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_{j_0}(\tilde{x},\tilde{y},\xi^i)-\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_{j_0}(x^k,y^k,\xi^i) \\ \xrightarrow{k\to+\infty}\quad & \mathbf{E}[F_{j_0}(\tilde{x},\tilde{y},\xi)]-\mathbf{E}[F_{j_0}(\bar{x},\bar{y},\xi)]. \quad w.p.1. \end{aligned}$$

This indicates that the item

$$\rho^k\Big(\big(h^k(x^k,y^k,\lambda^k)\big)^2-\big(h^k(\tilde{x},\tilde{y},\tilde{\lambda})\big)^2\Big) \tag{3.15}$$

is almost surely bounded.

On the other hand, by (2.11), there holds

$$h^k(\tilde{x},\tilde{y},\tilde{\lambda})\to h(\tilde{x},\tilde{y},\tilde{\lambda}),\quad w.p.1 \quad \text{as} \quad k\to+\infty. \tag{3.16}$$

Since (3.15) is almost surely bounded and $h(x,y,\lambda)\geq 0$, taking a limit in (3.15), we have from (3.16) and Lemma 3.2 that, with probability one,

$$h(\bar{x},\bar{y},\bar{\lambda})=h(\tilde{x},\tilde{y},\tilde{\lambda})=0,\quad \forall(\tilde{x},\tilde{y},\tilde{\lambda})\in\mathcal{F}.$$

This means that $(\bar{x},\bar{y},\bar{\lambda})\in\mathcal{F}$.

We next show that $(\bar{x},\bar{y},\bar{\lambda})$ is almost surely a weak Pareto optimal solution of problem (2.5). In fact, by $(\tilde{x},\tilde{y},\tilde{\lambda})\in\mathcal{F}$, (2.11), (3.16) and (A3), we have

$$\begin{aligned} \rho^k\big(h^k(\tilde{x},\tilde{y},\tilde{\lambda})\big)^2 &= \rho^k\Big(\big(h^k(\tilde{x},\tilde{y},\tilde{\lambda})\big)^2-\big(h(\tilde{x},\tilde{y},\tilde{\lambda})\big)^2\Big) \\ &= \rho^k\big(h^k(\tilde{x},\tilde{y},\tilde{\lambda})+h(\tilde{x},\tilde{y},\tilde{\lambda})\big)\big(h^k(\tilde{x},\tilde{y},\tilde{\lambda})-h(\tilde{x},\tilde{y},\tilde{\lambda})\big) \\ &\to 0 \quad w.p.1 \ \text{ as } \ k\to+\infty. \end{aligned} \tag{3.17}$$

In addition, we have from (3.14) that

$$\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_{j_0}(\tilde{x},\tilde{y},\xi^i)+\rho^k\big(h^k(\tilde{x},\tilde{y},\tilde{\lambda})\big)^2\geq\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_{j_0}(x^k,y^k,\xi^i). \tag{3.18}$$

Letting $k\to+\infty$ in (3.18) and taking (2.11), (3.13), (3.17) into account, we obtain

$$\mathbf{E}[F_{j_0}(\tilde{x},\tilde{y},\xi)]\geq\mathbf{E}[F_{j_0}(\bar{x},\bar{y},\xi)] \quad w.p.1,\quad \forall(\tilde{x},\tilde{y},\tilde{\lambda})\in\mathcal{F}.$$

This means that $(\bar{x},\bar{y},\bar{\lambda})$ is a weak Pareto optimal solution of (2.5) with probability one, that is, $(\bar{x},\bar{y},\bar{\lambda})\in\bar{\mathcal{O}}$ with probability one. □

4 Convergence Analysis of Pareto Stationary Points

Since both (2.5) and (2.12) are generally nonconvex, in this section, we focus on the convergence of Pareto stationary points of the approximation problem (2.12). To this end, we suppose $X := \{x \in R^n \,|\, g_i(x) \leq 0, i = 1, \ldots, r\}$ and $Y := \{y \in R^m \,|\, c_i(y) \leq 0, i = 1, \ldots, s\}$, where $g_i(x)$ $(i = 1, \ldots, r)$ are continuously differentiable functions and $c_i(y)(i = 1, \ldots, s)$ are continuously differentiable convex functions. Set $I_g(x) := \{i |\, g_i(x) = 0, i = 1, \ldots, r\}$, $I_c(y) := \{i |\, c_i(y) = 0, i = 1, \ldots, s\}$, and $I(\lambda) := \{i |\, \lambda_i = 0, i = 1, \ldots, q\}$.

Definition 4.1. $(\bar{x}, \bar{y}, \bar{\lambda})$ is called a Pareto stationary point of (2.5) if there exist multipliers $\bar{\sigma} \in R^p_+$, $\bar{\delta} \in R$, $\bar{u} \in R^r_+$, $\bar{v} \in R^s_+$, $\bar{w} \in R^q_+$, $\bar{l} \in R$ such that

$$\begin{pmatrix} \sum_{i=1}^{p} \left(\bar{\sigma}_i \nabla_x \mathbf{E}[F_i(\bar{x}, \bar{y}, \xi)]\right) \\ \sum_{i=1}^{p} \left(\bar{\sigma}_i \nabla_y \mathbf{E}[F_i(\bar{x}, \bar{y}, \xi)]\right) \\ 0 \end{pmatrix} + \bar{\delta} \begin{pmatrix} \nabla_x h(\bar{x}, \bar{y}, \bar{\lambda}) \\ \nabla_y h(\bar{x}, \bar{y}, \bar{\lambda}) \\ \nabla_\lambda h(\bar{x}, \bar{y}, \bar{\lambda}) \end{pmatrix} + \begin{pmatrix} \sum_{i=1}^{r} \bar{u}_i \nabla_x g_i(\bar{x}) \\ \sum_{i=1}^{s} \bar{v}_i \nabla_y c_i(\bar{y}) \\ \bar{l}\mathbf{e}_0 - \bar{w} \end{pmatrix} = 0, \quad (4.1)$$

$$\bar{u}^T g(\bar{x}) = 0, \quad g(\bar{x}) \leq 0, \quad (4.2)$$

$$\bar{v}^T c(\bar{y}) = 0, \quad c(\bar{y}) \leq 0, \quad (4.3)$$

$$\bar{w}^T \bar{\lambda} = 0, \quad \bar{\lambda} \geq 0, \quad \sum_{i=1}^{q} \bar{\lambda}_i = 1, \quad \sum_{i=1}^{p} \bar{\sigma}_i = 1, \quad (4.4)$$

where $\mathbf{e}_0$ is a q-dimensional unit vector. In what follows, the set of all multipliers associated with $(\bar{x}, \bar{y}, \bar{\lambda})$ is denoted by $D(\bar{x}, \bar{y}, \bar{\lambda})$.

Definition 4.2. (x^k, y^k, λ^k) is called a Pareto stationary point of (2.12) if there exist multipliers $\sigma^k \in R^p_+$, $u^k \in R^r_+$, $v^k \in R^s_+$, $w^k \in R^q_+$, $l^k \in R$ such that

$$\begin{pmatrix} \sum_{i=1}^{p} \sigma_i^k \left(\frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \nabla_x F_i(x^k, y^k, \xi^i) + 2\rho^k h^k(x^k, y^k, \lambda^k) \nabla_x h^k(x^k, y^k, \lambda^k)\right) \\ \sum_{i=1}^{p} \sigma_i^k \left(\frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \nabla_y F_i(x^k, y^k, \xi^i) + 2\rho^k h^k(x^k, y^k, \lambda^k) \nabla_y h^k(x^k, y^k, \lambda^k)\right) \\ \sum_{i=1}^{p} \sigma_i^k \left(2\rho^k h^k(x^k, y^k, \lambda^k) \nabla_\lambda h^k(x^k, y^k, \lambda^k)\right) \end{pmatrix}$$

$$+ \begin{pmatrix} \sum_{i=1}^{r} u_i^k \nabla_x g_i(x^k) \\ \sum_{i=1}^{s} v_i^k \nabla_y c_i(y^k) \\ l^k \mathbf{e}_0 - w^k \end{pmatrix} = 0, \quad (4.5)$$

$$(u^k)^T g(x^k) = 0, \quad g(x^k) \leq 0, \quad (4.6)$$

$$(v^k)^T c(y^k) = 0, \quad c(y^k) \leq 0, \quad (4.7)$$

$$(w^k)^T \lambda_i^k = 0, \quad \lambda^k \geq 0, \quad \sum_{i=1}^{q} \lambda_i^k = 1, \quad \sum_{i=1}^{p} \sigma_i^k = 1. \quad (4.8)$$

Definition 4.3. Let $(\bar{x}, \bar{y}, \bar{\lambda})$ be a feasible point of (2.5). The Basic Regularity Condition (BRC) is said to be satisfied at $(\bar{x}, \bar{y}, \bar{\lambda})$ if there exist $t \in \{1, \ldots, p\}$ and multipliers $\bar{\sigma}_i \geq 0$

($i \in \{1,\dots,p\}$ but $i \neq t$), $\bar{\delta} \in R$, $\bar{u}_i \geq 0$ ($i \in I_g(\bar{x})$), $\bar{u}_i = 0$ ($i \neq I_g(\bar{x})$), $\bar{v}_i \geq 0$ ($i \in I_c(\bar{y})$), $\bar{v}_i = 0$ ($i \neq I_c(\bar{y})$), $\bar{w}_i \geq 0$ ($i \in I(\bar{\lambda})$), $\bar{w}_i = 0$ ($i \neq I(\bar{\lambda})$), $\bar{l} \in R$ such that

$$\begin{pmatrix} \sum\limits_{i\in\{1,\dots,p\},i\neq t} \bar{\sigma}_i \nabla_x \mathbf{E}[F_i(\bar{x},\bar{y},\xi)] \\ \sum\limits_{i\in\{1,\dots,p\},i\neq t} \bar{\sigma}_i \nabla_y \mathbf{E}[F_i(\bar{x},\bar{y},\xi)] \\ 0 \end{pmatrix} + \bar{\delta} \begin{pmatrix} \nabla_x h(\bar{x},\bar{y},\bar{\lambda}) \\ \nabla_y h(\bar{x},\bar{y},\bar{\lambda}) \\ \nabla_\lambda h(\bar{x},\bar{y},\bar{\lambda}) \end{pmatrix} + \begin{pmatrix} \sum\limits_{i=1}^{r} \bar{u}_i \nabla_x g_i(\bar{x}) \\ \sum\limits_{i=1}^{s} \bar{v}_i \nabla_y c_i(\bar{y}) \\ \bar{l}\mathbf{e}_0 - \bar{w} \end{pmatrix} = 0$$

implies all multipliers to be zero.

Remark 4.4. The BRC property was first given in [11]. For the multiobjective problem

$$\begin{array}{ll} \min & (f_1(x),\dots,f_m(x)) \\ \text{s.t.} & g_i(x) \leq 0,\ i=1,\dots,k, \\ & h_r(x) = 0,\ r=1,\dots,p, \end{array} \tag{4.9}$$

if $\bar{x}$ is a Pareto optimal solution of (4.9) and the BRC holds at $\bar{x}$, the multiplier subset is nonempty bounded and, moreover, there is at least one scalar subproblem in the form

$$\begin{array}{ll} \min & f_q(x) \\ \text{s.t.} & f_j(x) \leq f_j(\bar{x}),\ j=1,\dots,m, j\neq q, \\ & g_i(x) \leq 0,\ i=1,\dots,k, \\ & h_r(x) = 0,\ r=1,\dots,p \end{array}$$

such that the Mangasarian-Fromovitz constraint qualification holds at $\bar{x}$ and the converse is also true, that is, if there is a scalar subproblem in the above form satisfying the Mangasarian-Fromovitz constraint qualification at $\bar{x}$, the BRC holds at $\bar{x}$ for (4.9).

Lemma 4.5. *Let* (A1) $-$ (A2) *hold and* $\lim_{k\to+\infty}(x^k,y^k,\lambda^k) = (\bar{x},\bar{y},\bar{\lambda})$. *Then, we have*

$$\lim_{k\to+\infty} \nabla_{xy\lambda} h^k(x^k,y^k,\lambda^k) = \nabla_{xy\lambda} h(\bar{x},\bar{y},\bar{\lambda})$$

with probability one.

Proof. By (A1) and (A2), we have that, for each $j=1,\dots,q$,

$$\left\| \nabla_{xy}\mathbf{f}_j(x^k,y^k,\xi^i) - \nabla_{xy}\mathbf{f}_j(\bar{x},\bar{y},\xi^i) \right\|_F \leq \mathcal{K}(\xi^i) \left\| (x^k,y^k) - (\bar{x},\bar{y}) \right\|^{\varrho}. \tag{4.10}$$

It then follows from (2.11) and (3.5) that

$$\begin{aligned} \left\| H^k(x^k,y^k,\lambda^k) - H(\bar{x},\bar{y},\bar{\lambda}) \right\| &\leq \left\| H^k(x^k,y^k,\lambda^k) - H^k(\bar{x},\bar{y},\bar{\lambda}) \right\| + \left\| H^k(\bar{x},\bar{y},\bar{\lambda}) - H(\bar{x},\bar{y},\bar{\lambda}) \right\| \\ &\to 0 \quad \text{as} \quad k\to+\infty \quad w.p.1. \end{aligned} \tag{4.11}$$

Moreover, we have from (A2), (2.11) and (4.10) that

$$\begin{aligned} &\Big\| \sum_{j=1}^{q} \lambda_j^k \Big(\frac{1}{N_k} \sum_{\xi^i\in\Omega_k} \nabla_x \mathbf{f}_j(x^k,y^k,\xi^i) \Big) - \sum_{j=1}^{q} \bar{\lambda}_j \nabla_x \mathbf{E}[\mathbf{f}_j(\bar{x},\bar{y},\xi)] \Big\|_F \\ &\leq \sum_{j=1}^{q} |\lambda_j^k - \bar{\lambda}_j| \cdot \frac{1}{N_k} \sum_{\xi^i\in\Omega_k} \mathcal{K}(\xi^i) + \sum_{j=1}^{q} \bar{\lambda}_j \cdot \frac{1}{N_k} \sum_{\xi^i\in\Omega_k} \mathcal{K}(\xi^i) \|x^k - \bar{x}\|^{\varrho} \\ &\quad + \sum_{j=1}^{q} \bar{\lambda}_j \Big\| \frac{1}{N_k} \sum_{\xi^i\in\Omega_k} \nabla_x \mathbf{f}_j(\bar{x},\bar{y},\xi^i) - \nabla_x \mathbf{E}[\mathbf{f}_j(\bar{x},\bar{y},\xi)] \Big\|_F \\ &\to 0 \quad \text{as} \quad k\to+\infty \quad w.p.1. \end{aligned} \tag{4.12}$$

Hence, from (A2), (2.9), (3.6), (4.11) and (4.12), we have

$$\begin{aligned}
&\big\|\nabla_x h^k(x^k,y^k,\lambda^k)-\nabla_x h(\bar{x},\bar{y},\bar{\lambda})\big\| \\
&\le \Big\|\sum_{j=1}^{q}\lambda_j^k\big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\nabla_x\mathbf{f}_j(x^k,y^k,\xi^i)\big)-\sum_{j=1}^{q}\bar{\lambda}_j\nabla_x\mathbf{E}[\mathbf{f}_j(\bar{x},\bar{y},\xi)]\Big\|_F\cdot\big\|y^k-H^k(x^k,y^k,\lambda^k)\big\| \\
&\quad+\big\|\sum_{j=1}^{q}\bar{\lambda}_j\mathbf{E}[\mathbf{f}_j(\bar{x},\bar{y},\xi)]\big\|_F\cdot\Big[\|y^k-\bar{y}\|+\big\|H^k(x^k,y^k,\lambda^k)-H(\bar{x},\bar{y},\bar{\lambda})\big\|\Big] \\
&\to 0 \ \text{ as } \ k\to+\infty \quad w.p.1. \qquad (4.13)
\end{aligned}$$

By (2.11) and (3.4), we have

$$\begin{aligned}
&\Big\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(x^k,y^k,\xi^i)-\mathbf{E}[\mathbf{f}_j(\bar{x},\bar{y},\xi)]\Big\| \\
\le\ &\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\big\|\mathbf{f}_j(x^k,y^k,\xi^i)-\mathbf{f}_j(\bar{x},\bar{y},\xi^i)\big\|+\Big\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(\bar{x},\bar{y},\xi^i)-\mathbf{E}[\mathbf{f}_j(\bar{x},\bar{y},\xi)]\Big\| \\
\to\ &0 \ \text{ as } \ k\to+\infty \quad w.p.1. \qquad (4.14)
\end{aligned}$$

It then follows from (A2), (2.9), (4.11), (4.13) and (4.14) that

$$\begin{aligned}
&\big\|\nabla_y h^k(x^k,y^k,\lambda^k)-\nabla_y h(\bar{x},\bar{y},\bar{\lambda})\big\| \\
&\le\sum_{j=1}^{q}|\lambda_j^k-\bar{\lambda}_j|\cdot\Big\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(x^k,y^k,\xi^i)\Big\|+\sum_{j=1}^{q}\bar{\lambda}_j\Big\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}_j(x^k,y^k,\xi^i)-\mathbf{E}[\mathbf{f}_j(\bar{x},\bar{y},\xi)]\Big\| \\
&\quad+\Big\|\sum_{j=1}^{q}\lambda_j^k\big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\nabla_y\mathbf{f}_j(x^k,y^k,\xi^i)\big)\cdot\big(H^k(x^k,y^k,\lambda^k)-y^k\big)-\sum_{j=1}^{q}\bar{\lambda}_j\nabla_y\mathbf{E}[\mathbf{f}_j(\bar{x},\bar{y},\xi)] \\
&\cdot\big(H(\bar{x},\bar{y},\bar{\lambda})-\bar{y}\big)\Big\|_F+\alpha\Big[\|y^k-\bar{y}\|+\|H^k(x^k,y^k,\lambda^k)-H(\bar{x},\bar{y},\bar{\lambda})\|\Big] \\
&\to 0 \ \text{ as } \ k\to+\infty \quad w.p.1.
\end{aligned}$$

In consequence, by (2.9), (3.6), (4.11) and (4.14), there holds

$$\begin{aligned}
&\big\|\nabla_\lambda h^k(x^k,y^k,\lambda^k)-\nabla_\lambda h(\bar{x},\bar{y},\bar{\lambda})\big\| \\
&\le\Big\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\mathbf{f}(x^k,y^k,\xi^i)-\mathbf{E}[\mathbf{f}(\bar{x},\bar{y},\xi)]\Big\|_F\cdot\big\|y^k-H^k(x^k,y^k,\lambda^k)\big\| \\
&\quad+\big\|\mathbf{E}[\mathbf{f}(\bar{x},\bar{y},\xi)]\big\|_F\cdot\Big[\big\|y^k-\bar{y}\|+\|H^k(x^k,y^k,\lambda^k)-H(\bar{x},\bar{y},\bar{\lambda})\|\Big] \\
&\to 0 \ \text{ as } \ k\to+\infty \quad w.p.1.
\end{aligned}$$

Therefore, the conclusion holds. □

From now on, we denote by $\bar{\mathcal{S}}$ and $\mathcal{S}_k$ the sets of Pareto stationary points of (2.5) and (2.12) respectively.

Theorem 4.6. *Let* (A1) – (A3) *hold and the constant C satisfy $\big\|\Theta^k(x^k,y^k,\lambda^k)\big\|\le C$ for each k. Let $(x^k,y^k,\lambda^k)\in\mathcal{S}_k$ for each k, $(\bar{x},\bar{y},\bar{\lambda})$ be an accumulation point of $\big\{(x^k,y^k,\lambda^k)\big\}$, and the BRC hold at $(\bar{x},\bar{y},\bar{\lambda})$. Then, we have $(\bar{x},\bar{y},\bar{\lambda})\in\bar{\mathcal{S}}$ with probability one.*

Proof. Without loss of generality, we assume that $\lim_{k\to\infty}(x^k, y^k, \lambda^k) = (\bar{x}, \bar{y}, \bar{\lambda})$. Let $\sigma^k \in R^p_+$, $u^k \in R^r_+$, $v^k \in R^s_+$, $w^k \in R^q_+$, $l^k \in R$ be the corresponding multiplier vectors in (4.2)–(4.8). By the mean value theorem and (A2), we have

$$\begin{aligned} & \big\|\nabla_{xy}F_j(x^k, y^k, \xi^i) - \nabla_{xy}F_j(\bar{x}, \bar{y}, \xi^i)\big\| \\ = \;& \Big\| \int_0^1 \big(\nabla^2_{xy}F_j(tx^k + (1-t)\bar{x}, ty^k + (1-t)\bar{y}, \xi^i)\big)^T \big((x^k, y^k) - (\bar{x}, \bar{y})\big)dt \Big\| \\ \le \;& \mathcal{K}(\xi^i)\big\|(x^k, y^k) - (\bar{x}, \bar{y})\big\|, \quad \forall j = 1, \ldots, p. \end{aligned} \tag{4.15}$$

Then, we have from (A2), (2.11) and (4.15) that

$$\begin{aligned} & \Big\| \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \nabla_{xy}F_j(x^k, y^k, \xi^i) - \nabla_{xy}\mathbf{E}[F_j(\bar{x}, \bar{y}, \xi)] \Big\| \\ \le \;& \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \big\|\nabla_{xy}F_j(x^k, y^k, \xi^i) - \nabla_{xy}F_j(\bar{x}, \bar{y}, \xi^i)\big\| \\ & \qquad + \Big\| \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \nabla_{xy}F_j(\bar{x}, \bar{y}, \xi^i) - \nabla_{xy}\mathbf{E}[F_j(\bar{x}, \bar{y}, \xi)] \Big\| \\ \le \;& \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \mathcal{K}(\xi^i)\big\|(x^k, y^k) - (\bar{x}, \bar{y})\big\| \\ & \qquad + \Big\| \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \nabla_{xy}F_j(\bar{x}, \bar{y}, \xi^i) - \nabla_{xy}\mathbf{E}[F_j(\bar{x}, \bar{y}, \xi)] \Big\| \\ \to \;& 0 \ \text{ as } \ k \to +\infty \quad w.p.1, \quad \forall j = 1, \ldots, p. \end{aligned} \tag{4.16}$$

In addition, by the condition $\big\|\Theta^k(x^k, y^k, \lambda^k)\big\| \le C$ and (2.12), we have

$$\big(h^k(x^k, y^k, \lambda^k)\big)^2 \le (\rho^k)^{-1}\Big(C + \big\|\frac{1}{N_k} \sum_{\xi^i \in \Omega_k} F(x^k, y^k, \xi^i)\big\|\Big), \quad \forall k. \tag{4.17}$$

Note that $\lim_{k\to+\infty} \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} F(x^k, y^k, \xi^i) = \mathbf{E}[F(\bar{x}, \bar{y}, \xi)]$ holds with probability one by (3.13). Letting $k \to +\infty$ in (4.17), by Lemma 3.2, we have

$$h(\bar{x}, \bar{y}, \bar{\lambda}) = 0 \quad w.p.1. \tag{4.18}$$

Note that

$$\begin{aligned} & \big|h^k(x^k, y^k, \lambda^k) - h^k(\bar{x}, \bar{y}, \bar{\lambda})\big| \\ & \le \|y^k - \bar{y}\| + \Big| \big(H^k(x^k, y^k, \lambda^k)\big)^T \sum_{j=1}^q \lambda^k_j \big(\frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \mathbf{f}_j(x^k, y^k, \xi^i)\big) \\ & \quad - \big(H^k(\bar{x}, \bar{y}, \bar{\lambda})\big)^T \sum_{j=1}^q \bar{\lambda}_j \big(\frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \mathbf{f}_j(\bar{x}, \bar{y}, \xi^i)\big) \Big| \\ & \quad + \frac{\alpha}{2}\Big[\big\|y^k - H^k(x^k, y^k, \lambda^k)\big\|^2 - \big\|\bar{y} - H^k(\bar{x}, \bar{y}, \bar{\lambda})\big\|^2\Big]. \end{aligned}$$

This together with (3.8), (3.11) and (A3) implies $\lim_{k\to+\infty} \rho^k\big(h^k(x^k, y^k, \lambda^k) - h^k(\bar{x}, \bar{y}, \bar{\lambda})\big) =$

0 with probability one. It then follows from (4.18) and (A3) that

$$\begin{aligned} \left|\rho^k h^k(x^k,y^k,\lambda^k)\right| &= \left|\rho^k\big(h^k(x^k,y^k,\lambda^k)-h(\bar{x},\bar{y},\bar{\lambda})\big)\right| \\ &\le \rho^k\left|h^k(x^k,y^k,\lambda^k)-h^k(\bar{x},\bar{y},\bar{\lambda})\right|+\rho^k\left|h^k(\bar{x},\bar{y},\bar{\lambda})-h(\bar{x},\bar{y},\bar{\lambda})\right| \\ &\to 0 \quad \text{as} \quad k\to+\infty \quad w.p.1. \end{aligned} \tag{4.19}$$

(i) We first show that the sequences $\{\sigma^k\}$, $\{u^k\}$, $\{v^k\}$, $\{w^k\}$, $\{l^k\}$ are all bounded with probability one. To this end, we set

$$\tau_k := \sqrt{\|\sigma^k\|^2+\|u^k\|^2+\|v^k\|^2+\|w^k\|^2+\|l^k\|^2} \tag{4.20}$$

and

$$\hat{\sigma}^k=\frac{\sigma^k}{\tau^k},\quad \hat{u}^k=\frac{u^k}{\tau^k},\quad \hat{v}^k=\frac{v^k}{\tau^k},\quad \hat{w}^k=\frac{w^k}{\tau^k},\quad \hat{l}^k=\frac{l^k}{\tau^k}. \tag{4.21}$$

Taking a subsequence if necessary, we may suppose that all limits $\tilde{\sigma} := \lim_{k\to+\infty}\hat{\sigma}^k$, $\tilde{u} := \lim_{k\to+\infty}\hat{u}^k$, $\tilde{v} := \lim_{k\to+\infty}\hat{v}^k$, $\tilde{w} := \lim_{k\to+\infty}\hat{w}^k$, $\tilde{l} := \lim_{k\to+\infty}\hat{l}^k$ exist with probability one. It is easy to see from (4.20)–(4.21) that

$$\|\tilde{\sigma}\|^2+\|\tilde{u}\|^2+\|\tilde{v}\|^2+\|\tilde{w}\|^2+\|\tilde{l}\|^2=1. \tag{4.22}$$

Assume by contradiction that $\{\sigma^k\}$, $\{u^k\}$, $\{v^k\}$, $\{w^k\}$, $\{l^k\}$ are not all bounded almost surely. It follows that $\{\tau_k\}$ is almost surely unbounded. Without loss of generality, we suppose $\lim_{k\to+\infty}\tau_k=+\infty$ with probability one. By (4.8), we have $\tilde{\sigma}=0$ with probability one. Dividing (4.2) by τ_k and taking a limit, we have from (4.16), (4.19), and Lemma 4.5 that

$$\begin{pmatrix} \sum_{i=1}^{r}\tilde{u}_i\nabla_x g_i(\bar{x}) \\ \sum_{i=1}^{s}\tilde{v}_i\nabla_y c_i(\bar{y}) \\ \tilde{l}\mathbf{e}_0-\tilde{w} \end{pmatrix}=0,\quad w.p.1.$$

It then follows from the BRC that $\tilde{u}_i=0$ $(i=1,\ldots,r)$, $\tilde{v}_i=0$ $(i=1,\ldots,s)$, $\tilde{w}_i=0$ $(i=1,\ldots,q)$, and $\tilde{l}=0$ with probability one. This together with $\tilde{\sigma}=0$ contradicts (4.22). Therefore, $\{\sigma^k\}$, $\{u^k\}$, $\{v^k\}$, $\{w^k\}$, $\{l^k\}$ are all bounded with probability one.

(ii) We next show that $(\bar{x},\bar{y},\bar{\lambda})$ satisfies (4.1)–(4.4) with probability one. In fact, by (i), we may assume that the limits $\bar{\sigma}=\lim_{k\to+\infty}\sigma^k$, $\bar{u}=\lim_{k\to+\infty}u^k$, $\bar{v}=\lim_{k\to+\infty}v^k$, $\bar{w}=\lim_{k\to+\infty}w^k$, $\bar{l}=\lim_{k\to+\infty}l^k$ exist with probability one. Taking a limit in (4.2)–(4.8), we obtain by (4.16), (4.19) and Lemma 4.5 that (4.1)–(4.4) hold with probability one. That is, $(\bar{x},\bar{y},\bar{\lambda})\in\bar{\mathcal{S}}$ with probability one. □

5 Conclusions

For the SMBLP (1.1)–(1.2), since the lower-level program is assumed to be convex, we used the necessary and sufficient conditions for multiobjective problems and some techniques to transform (1.1)–(1.2) into the stochastic multiobjective problem (2.3) with scalar variational inequality constraints. By means of the regularized gap function (2.4), we transformed (2.3) into (2.5). Furthermore, by some penalty function and sample average approximation techniques, we used (2.12) to approximate (2.5). Convergence analysis for both the weak Pareto optimal solutions and the Pareto stationary points have been considered as well. Note that all the above theoretical results were based on the (weak) Pareto solution. As a future work, we will try to extend the results to cases of the cone efficient solutions or the proper efficient solutions. We will also investigate the applications of (1.1)–(1.2) in practice.

References

[1] M.J. Alves and C.H. Antunes, A semivectorial bilevel programming approach to optimize electricity dynamic time-of-use retail pricing, *Comput. Oper. Res.* 92 (2017) 130–144.

[2] A.H. Barahimi, A. Eydi and A. Aghaie, Multi-modal urban transit network design considering reliability: Multi-objective Bi-level optimization, *Reliab. Eng. Syst. Saf.* 216 (2021): 107922.

[3] J.F. Bard, *Practical Bilevel Optimization: Algorithms and Applications*, Kluwer Academic Publications, Dordrecht, The Netherlands, 1998.

[4] O. Ben-Ayed and C.E. Blair, Computational difficulties of bilevel linear programming, *Oper. Res.* 38 (1990) 556–560.

[5] H. Bonnel and J. Morgan, Semivectorial bilevel optimization problem: Penalty approach, *J. Optim. Theory Appl.* 131 (2006) 365–382.

[6] A. Chen, J. Kim, S. Lee and Y. Kim. Stochastic multi-objective models for network design problem, *Expert Syst. Appl.* 37 (2010) 1608–1619.

[7] B. Colson, P. Marcotte and G. Savard, An overview of bilevel programming, *Ann. Oper. Res.* 153 (2007) 235–256.

[8] S. Dempe and J. Dutta, Is bilevel programming a special case of a mathematical program with complementarity constraints, *Math. Program.* 131 (2012) 37–48.

[9] S. Dempe, N. Gadhi and A.B. Zemkoho, New optimality conditions for the semivectorial bilevel optimization problem, *J. Optim. Theory Appl.* 157 (2013) 54–74.

[10] S. Dempe, V. Kalashnikov, G.A. Pérez-Valdés and N. Kalashnykova, *Bilevel Programming Problems*, Springer, Berlin, 2015.

[11] J. Dutta and C.S. Lalitha, Bounded sets of KKT multipliers in vector optimization, *J. Glob. Optim.* 36 (2006), 425–437.

[12] M. Fukushima, Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems, *Math. Program.* 53 (1992) 99–110.

[13] M. Fukushima, Merit functions for variational inequality and complementarity problems, in: *Nonlinear Optimization and Applications*, G. Di Pillo, F. Giannessi (eds.) Plenum, New York, 1996, pp. 155–170.

[14] H.A. Gil, F.D. Galiana and E.L. Da Silva, Nodal price control: a mechanism for transmission network cost allocation, *IEEE Trans. Power Syst.* 21 (2006), 3–10.

[15] P. Hall and J.S. Marron, Local minima in cross-validation function, *J. R. Stat. Soc.* 53 (1991), 245–252.

[16] H. Karimi, S. Jadid and H. Saboori, Multi-objective bi-level optimisation to design real-time pricing for demand response programs in retail markets, *IET Gener. Transm. Distrib.* 13 (2019) 1287–1296.

[17] K.R. Kazmi, *On vector equilibrium problem*, Proceedings of the Indian Academy of Sciences 110 (2000) 213–223.

[18] S. Kosuch, P. L. Bodic, J. Leung and A. Lisser, *On a stochastic bilevel programming problem*, Networks 59 (2012) 107–116.

[19] R.M. Kovacevic and G.C. Pflug, Electricity swing option pricing by stochastic bilevel optimization: A survey and new approaches, *Eur. J. Oper. Res.* 237 (2014) 389–403.

[20] G.H. Lin, D.L. Zhang and Y.C. Liang, Stochastic multiobjective problems with complementarity constraints and applications in healthcare management, *Eur. J. Oper. Res.* 226 (2013) 461–470.

[21] Y.B. Lü and Z.P. Wan, A smoothing method for solving bilevel multiobjective programming problems, *J. Oper. Res. Soc. China* 2 (2014) 511–525.

[22] J. Mirrlees, The theory of moral hazard and unobservable behaviour: Part I, *Rev. Econ. Stud.* 66 (1999) 3–22.

[23] J. Nie, L. Wang, J.J. Ye and S. Zhong. *A Lagrange Multiplier Expression Method for Bilevel Polynomial Optimization*, (2021), to appear.

[24] B. Patrick, *Probability and Measure*, New York: Wiley-Interscience, 1995.

[25] S.M. Robinson, Analysis of sample-path optimization, *Math. Oper. Res.* 21 (1996) 513–528.

[26] A. Sinha, P. Malo and K. Deb, Approximated set-valued mapping approach for handling multiobjective bilevel problems, *Comput. Oper. Res.* 77 (2017) 194–209.

[27] H.V. Stackelberg, *Marketform and Gleichgemicht*, Vienna: Julius Press, 1934.

[28] H. Yang, X.N. Zhang and Q. Meng, Stackelberg games and multiple equilibrium behaviors on networks, *Transp. Res. Pt. B-Methodol.* 41 (2007) 841–861.

[29] Z. Yao, S.C.H. Leung and K.K. Lai, Manufacturer's revenue-sharing contract and retail competition, *Eur. J. Oper. Res.* 186 (2008) 637–651.

[30] J.J. Ye, Constraint qualifications and KKT conditions for bilevel programming problems, *Math. Oper. Res.* 31 (2006) 811–824.

[31] J.J. Ye, Necessary optimality conditions for multiobjective bilevel programs, *Math. Oper. Res.* 36 (2011) 165–184.

[32] J.J. Ye, X. Yuan, S. Zeng and J.Zhang, Difference of convex algorithms for bilevel programs with applications in hyper parameter selection, (2021), to appear.

[33] Y.F. Yin, Multiobjective bilevel optimization for transportation planning and management problems, *J. Adv. Transp.* 36 (2002) 93–105.

[34] Y. Zhao, J. Zhang, X.M. Yang and G.H. Lin, Expected residual minimization formulation for a class of stochastic vector variational inequalities, *J. Optim. Theory Appl.* 175 (2017) 545–566.

Manuscript received 16 September 2021
revised 17 November 2021
accepted for publication 17 May 2022

Meiju Luo
School of Mathematics and Statistics
Liaoning University, Shenyang 110036, China
E-mail address: meijuluolnu@126.com

Gui-Hua Lin
School of Management
Shanghai University, Shanghai 200444, China
E-mail address: guihualin@shu.edu.cn