

Monte Carlo and Quasi-Monte Carlo Sampling Methods for a Class of Stochastic Mathematical Programs with Equilibrium Constraints*

Gui-Hua Lin[†] Huifu Xu[‡] and Masao Fukushima[§]

December 10, 2005

Abstract. In this paper, we consider a class of stochastic mathematical programs with equilibrium constraints introduced by Birbil et al. (2004). Firstly, by means of a Monte Carlo method, we obtain a nonsmooth discrete approximation of the original problem. Then, we propose a smoothing method together with a penalty technique to get a standard nonlinear programming problem. Some convergence results are established. Moreover, since quasi-Monte Carlo methods are generally faster than Monte Carlo methods, we discuss a quasi-Monte Carlo sampling approach as well. Furthermore, we give an example in economics to illustrate the model and show some numerical results with this example.

Key words. Stochastic mathematical program with equilibrium constraints, Monte Carlo/quasi-Monte Carlo methods, penalization.

2000 Mathematics Subject Classification. 90C30, 90C33, 90C15.

1 Introduction

The purpose of this paper is to develop an efficient numerical method for solving the stochastic mathematical program with equilibrium constraints (SMPEC) formulated as follows:

$$\begin{aligned} \min \quad & \mathbb{E}[f(x, y, \omega)] \\ \text{s.t.} \quad & g(x, y) \leq 0, \quad h(x, y) = 0, \\ & 0 \leq y \perp \mathbb{E}[F(x, y, \omega)] \geq 0, \end{aligned} \tag{1.1}$$

where \mathbb{E} means expectation with respect to the random variable $\omega \in \Omega$, the functions $f : \mathfrak{R}^{n+m} \times \Omega \rightarrow \mathfrak{R}$, $g : \mathfrak{R}^{n+m} \rightarrow \mathfrak{R}^{s_1}$, $h : \mathfrak{R}^{n+m} \rightarrow \mathfrak{R}^{s_2}$, and $F : \mathfrak{R}^{n+m} \times \Omega \rightarrow \mathfrak{R}^m$ are all twice continuously

*The first author's work was supported in part by the Scientific Research Grant-in-Aid from Japan Society for the Promotion of Science and SRF for ROCS, SEM. The second author's work was supported in part by the United Kingdom Engineering and Physical Sciences Research Council grant. The third author's work was supported in part by the Scientific Research Grant-in-Aid from Japan Society for the Promotion of Science.

[†]Department of Applied Mathematics, Dalian University of Technology, Dalian 116024, China. Current address: Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan. E-mail: ghlin@amp.i.kyoto-u.ac.jp.

[‡]School of Mathematics, University of Southampton, Highfield Southampton, UK. E-mail: H.Xu@soton.ac.uk.

[§]Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan. E-mail: fuku@amp.i.kyoto-u.ac.jp.

differentiable, and the symbol \perp means the two vectors are perpendicular to each other. When the underlying sample space Ω has a finite number of samples, problem (1.1) reduces to an ordinary MPEC and there have been proposed a number of approaches [4]. Throughout the paper, we suppose that Ω has infinitely many samples.

Recently there has been quite active research on various formulations of SMPECs such as the *lower-level wait-and-see* model [8, 18–20] and the *here-and-now* model [1, 8–12]. Despite the high potential of practical applicability, however, the formulation (1.1) has rarely been studied except the recent work of Birbil, Gürkan and Listeş [1] who first treated the SMPEC of this form and presented a sample-path method for solving it along with rigorous convergence analysis.

In this paper, we propose a Monte Carlo sampling method combined with a penalty technique for solving problem (1.1). The proposed method is similar in spirit to that of Gürkan et al. [1]. However, there is a substantial difference between the two methods. Specifically, the approximation problems solved in our method are ordinary nonlinear programs, while those in the method of [1] are still MPECs. This fact indicates that the proposed method may significantly mitigate the computational difficulty in solving the SMPEC (1.1).

We establish convergence of global optimal solutions and stationary points of approximation problems generated by the proposed method. Moreover, since quasi-Monte Carlo methods are generally faster than Monte Carlo methods, we suggest a combined quasi-Monte Carlo sampling and penalty method. Finally, we show some preliminary numerical results with a stochastic version of Stackelberg-Nash-Cournot game.

The following notations will be used throughout the paper. For a given function $c : \mathfrak{R}^s \rightarrow \mathfrak{R}^{s'}$ and a vector $t \in \mathfrak{R}^s$, $\nabla c(t) \in \mathfrak{R}^{s \times s'}$ is the transposed Jacobian of c at t and $\mathcal{I}_c(t) := \{i : c_i(t) = 0\}$ stands for the active index set of c at t . For a matrix A , we let A_i denote a column vector whose elements consist of the i th row of A . In addition, I and O denote the identity matrix and the zero matrix of suitable dimension, respectively.

2 Preliminaries

In this section, we recall some basic concepts that are often employed in the literature on MPEC. Let (x^*, y^*) be a feasible point of problem (1.1) and denote

$$G(x, y) := y, \quad H(x, y) := \mathbb{E}[F(x, y, \omega)].$$

Definition 2.1 We say the *MPEC-linear independence constraint qualification* (MPEC-LICQ) holds at (x^*, y^*) if the set of vectors

$$\left\{ \nabla g_i(x^*, y^*), \nabla h_j(x^*, y^*), \nabla G_i(x^*, y^*), \nabla H_j(x^*, y^*) : \right. \\ \left. i \in \mathcal{I}_g(x^*, y^*), j \in \{1, \dots, s_2\}, i \in \mathcal{I}_G(x^*, y^*), j \in \mathcal{I}_H(x^*, y^*) \right\} \quad (2.1)$$

is linearly independent.

Definition 2.2 [16] Suppose that there exist Lagrangian multiplier vectors $\alpha^* \in \mathfrak{R}^{s_1}, \beta^* \in \mathfrak{R}^{s_2}$, and $\gamma^*, \delta^* \in \mathfrak{R}^m$ such that

$$\begin{aligned} \mathbb{E}[\nabla_{(x,y)} f(x^*, y^*, \omega)] + \nabla g(x^*, y^*)\alpha^* + \nabla h(x^*, y^*)\beta^* \\ - \begin{pmatrix} O \\ I \end{pmatrix} \gamma^* - \mathbb{E}[\nabla_{(x,y)} F(x^*, y^*, \omega)]\delta^* = 0, \end{aligned} \quad (2.2)$$

$$0 \leq \alpha^* \perp -g(x^*, y^*) \geq 0, \quad (2.3)$$

$$\gamma_i^* = 0, \quad i \notin \mathcal{I}_G(x^*, y^*), \quad (2.4)$$

$$\delta_i^* = 0, \quad i \notin \mathcal{I}_H(x^*, y^*). \quad (2.5)$$

- We call (x^*, y^*) a *Clarke or C-stationary* point of (1.1) if $\gamma_i^* \delta_i^* \geq 0$ holds for each $i \in \mathcal{I}_G(x^*, y^*) \cap \mathcal{I}_H(x^*, y^*)$.
- We call (x^*, y^*) a *Bouligand or B-stationary* point of (1.1) if there holds $\gamma_i^* \geq 0$ and $\delta_i^* \geq 0$ for each $i \in \mathcal{I}_G(x^*, y^*) \cap \mathcal{I}_H(x^*, y^*)$.

Definition 2.3 We say that the *lower-level strict complementarity (LLSC)* condition holds at (x^*, y^*) if $\mathcal{I}_G(x^*, y^*) \cap \mathcal{I}_H(x^*, y^*) = \emptyset$.

Note that, when the LLSC holds, there is no difference between the stationarity concepts given in Definition 2.2.

3 Monte Carlo Sampling and Penalty Method

For $\epsilon \geq 0$, we define $\phi_\epsilon : \mathfrak{R}^2 \rightarrow \mathfrak{R}$ by $\phi_\epsilon(a, b) := a + b - \sqrt{a^2 + b^2 + \epsilon^2}$. Then ϕ_0 is the well-known Fischer-Burmeister function, which is differentiable except at the origin. When $\epsilon > 0$, the function ϕ_ϵ is differentiable everywhere. Furthermore, we define $\Phi_\epsilon : \mathfrak{R}^{2m} \rightarrow \mathfrak{R}^m$ by

$$\Phi_\epsilon(y, w) := \begin{pmatrix} \phi_\epsilon(y_1, w_1) \\ \vdots \\ \phi_\epsilon(y_m, w_m) \end{pmatrix}.$$

It is obvious that problem (1.1) is equivalent to

$$\begin{aligned} \min \quad & \mathbb{E}[f(x, y, \omega)] \\ \text{s.t.} \quad & g(x, y) \leq 0, \quad h(x, y) = 0, \\ & \Phi_0(y, \mathbb{E}[F(x, y, \omega)]) = 0. \end{aligned} \quad (3.1)$$

Since both the objective function and the constraints involve expectations, problem (1.1) or (3.1) is more difficult to deal with than an ordinary MPEC. Moreover, the constraints in problem (1.1)

fail to satisfy a standard constraint qualification at any feasible point [3], while (3.1) is actually a nonsmooth program. We next employ a penalty technique and the Monte Carlo sampling method to get some appropriate approximations of the above problems.

For a function $\psi : \Omega \rightarrow \mathfrak{R}$, the Monte Carlo sampling estimate for $\mathbb{E}[\psi(\omega)]$ is obtained by taking independently and identically distributed random samples $\{\omega_1, \dots, \omega_k\}$ from Ω and letting $\mathbb{E}[\psi(\omega)] \approx \frac{1}{k} \sum_{\ell=1}^k \psi(\omega_\ell)$. The strong law of large numbers guarantees that this procedure converges with probability one (abbreviated by “w.p.1”), i.e.,

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{\ell=1}^k \psi(\omega_\ell) = \mathbb{E}[\psi(\omega)] := \int_{\Omega} \psi(\omega) d\zeta(\omega) \quad \text{w.p.1,} \quad (3.2)$$

where $\zeta(\omega)$ is the distribution function of ω . See [14] for more details.

Thus, by taking independently and identically distributed random samples $\{\omega_1, \dots, \omega_k\}$ from Ω , we obtain the following approximation of problem (3.1):

$$\begin{aligned} \min \quad & \frac{1}{k} \sum_{\ell=1}^k f(x, y, \omega_\ell) \\ \text{s.t.} \quad & g(x, y) \leq 0, \quad h(x, y) = 0, \\ & \Phi_0\left(y, \frac{1}{k} \sum_{\ell=1}^k F(x, y, \omega_\ell)\right) = 0. \end{aligned}$$

Note that the above problem is essentially an MPEC. Then, we introduce a smoothing parameter $\epsilon_k > 0$ and, in order to simplify the constraints, we employ a penalty technique to get the following smooth approximation:

$$\begin{aligned} \min \quad & \theta_k(x, y) := \frac{1}{k} \sum_{\ell=1}^k f(x, y, \omega_\ell) + \rho_k \left\| \Phi_{\epsilon_k}\left(y, \frac{1}{k} \sum_{\ell=1}^k F(x, y, \omega_\ell)\right) \right\|^2 \\ \text{s.t.} \quad & g(x, y) \leq 0, \quad h(x, y) = 0, \end{aligned} \quad (3.3)$$

where $\rho_k > 0$ is a penalty parameter. Problem (3.3) is no longer an MPEC and its constraints are independent of k .

In what follows, we let \mathcal{F} and \mathcal{X} denote the feasible regions of problems (1.1) and (3.3), respectively, and we suppose \mathcal{F} is nonempty. It is obvious that $\mathcal{F} \subseteq \mathcal{X}$.

4 Convergence Analysis

We investigate convergence properties of the Monte Carlo sampling and penalty method in this section. In the rest of this section, we suppose that F is affine with respect to (x, y) and is given by

$$F(x, y, \omega) := N(\omega)x + M(\omega)y + q(\omega),$$

where $N : \Omega \rightarrow \mathfrak{R}^{m \times n}$, $M : \Omega \rightarrow \mathfrak{R}^{m \times m}$, and $q : \Omega \rightarrow \mathfrak{R}^m$ are all continuous. In what follows, we denote

$$\bar{N} := \mathbb{E}[N(\omega)], \quad \bar{M} := \mathbb{E}[M(\omega)], \quad \bar{q} := \mathbb{E}[q(\omega)].$$

In order to obtain some convergence results for the proposed method, we suppose that the parameters ρ_k and ϵ_k satisfy the following conditions with probability one:

$$\lim_{k \rightarrow \infty} \rho_k = +\infty, \quad \limsup_{k \rightarrow \infty} \rho_k \epsilon_k < +\infty, \quad (4.1)$$

$$\left. \begin{aligned} \lim_{k \rightarrow \infty} \sqrt{\rho_k} \left(\frac{1}{k} \sum_{\ell=1}^k N_i(\omega_\ell) - \bar{N}_i \right) &= 0 \\ \lim_{k \rightarrow \infty} \sqrt{\rho_k} \left(\frac{1}{k} \sum_{\ell=1}^k M_i(\omega_\ell) - \bar{M}_i \right) &= 0 \\ \lim_{k \rightarrow \infty} \sqrt{\rho_k} \left(\frac{1}{k} \sum_{\ell=1}^k q_i(\omega_\ell) - \bar{q}_i \right) &= 0 \end{aligned} \right\} \quad i = 1, \dots, m. \quad (4.2)$$

Note that (4.1) implies $\lim_{k \rightarrow \infty} \epsilon_k = 0$.

4.1 Limiting behavior of optimal solutions

We first study the convergence of optimal solutions of problems (3.3). The following lemma can be verified easily.

Lemma 4.1 *Let $\epsilon \geq 0$. Then, for any real numbers a_i and $b_i, i = 1, 2$, we have*

$$\begin{aligned} |\phi_\epsilon(a_1, b_1) - \phi_\epsilon(a_2, b_2)| &\leq 2(|a_1 - a_2| + |b_1 - b_2|), \\ |\phi_\epsilon(a_1, b_1) - \phi_0(a_2, b_2)| &\leq 2(|a_1 - a_2| + |b_1 - b_2|) + \epsilon. \end{aligned}$$

Definition 4.1 [15] Let $\sigma > 0$ and $\kappa \geq 0$ be constants. We say $G : \mathfrak{R}^s \rightarrow \mathfrak{R}^t$ to be *Hölder continuous* on $K \subseteq \mathfrak{R}^s$ with order σ and Hölder constant κ if

$$\|G(u) - G(v)\| \leq \kappa \|u - v\|^\sigma$$

holds for all u and v in K .

This concept is a generalization of the Lipschitz continuity, which is, by definition, Hölder continuity with order $\sigma = 1$. Note that, for two different positive numbers σ and σ' , Hölder continuous functions with order σ and those with order σ' constitute different subclasses. For example, the function $G(u) := \sqrt{\|u\|}$ is Hölder continuous with order $\sigma = \frac{1}{2}$ but not Lipschitz continuous.

Theorem 4.1 *Let f be Hölder continuous in (x, y) on \mathcal{X} with order $\sigma > 0$ and Hölder constant $\kappa(\omega) > 0$ satisfying $\int_\Omega \kappa(\omega) d\zeta(\omega) < +\infty$. Let the parameters ρ_k and ϵ_k be chosen to satisfy (4.1) and (4.2). Suppose that (x^k, y^k) solves problem (3.3) for each k and (x^*, y^*) is an accumulation point of the sequence $\{(x^k, y^k)\}$. Then (x^*, y^*) is an optimal solution of problem (1.1) with probability one.*

Proof. Without loss of generality, we suppose $\lim_{k \rightarrow \infty} (x^k, y^k) = (x^*, y^*)$. Since (x^k, y^k) is an optimal solution of problem (3.3), it follows that

$$\begin{aligned} & \frac{1}{k} \sum_{\ell=1}^k f(x^k, y^k, \omega_\ell) + \rho_k \left\| \Phi_{\epsilon_k} \left(y^k, \frac{1}{k} \sum_{\ell=1}^k F(x^k, y^k, \omega_\ell) \right) \right\|^2 \\ & \leq \frac{1}{k} \sum_{\ell=1}^k f(x, y, \omega_\ell) + \rho_k \left\| \Phi_{\epsilon_k} \left(y, \frac{1}{k} \sum_{\ell=1}^k F(x, y, \omega_\ell) \right) \right\|^2 \end{aligned} \quad (4.3)$$

holds for any $(x, y) \in \mathcal{X}$ and each k .

(a) We first prove that (x^*, y^*) is almost surely a feasible point of problem (1.1). In fact, for an arbitrary $(\bar{x}, \bar{y}) \in \mathcal{F}$, we have from (4.3), the Hölder continuity of f on \mathcal{X} , and (3.2) that

$$\begin{aligned} & \rho_k \left(\left\| \Phi_{\epsilon_k} \left(y^k, \frac{1}{k} \sum_{\ell=1}^k F(x^k, y^k, \omega_\ell) \right) \right\|^2 - \left\| \Phi_{\epsilon_k} \left(\bar{y}, \frac{1}{k} \sum_{\ell=1}^k F(\bar{x}, \bar{y}, \omega_\ell) \right) \right\|^2 \right) \\ & \leq \frac{1}{k} \sum_{\ell=1}^k f(\bar{x}, \bar{y}, \omega_\ell) - \frac{1}{k} \sum_{\ell=1}^k f(x^k, y^k, \omega_\ell) \\ & = \frac{1}{k} \sum_{\ell=1}^k f(\bar{x}, \bar{y}, \omega_\ell) - \frac{1}{k} \sum_{\ell=1}^k f(x^*, y^*, \omega_\ell) + \frac{1}{k} \sum_{\ell=1}^k \left(f(x^*, y^*, \omega_\ell) - f(x^k, y^k, \omega_\ell) \right) \\ & \leq \frac{1}{k} \sum_{\ell=1}^k f(\bar{x}, \bar{y}, \omega_\ell) - \frac{1}{k} \sum_{\ell=1}^k f(x^*, y^*, \omega_\ell) + \|(x^k, y^k) - (x^*, y^*)\|^\sigma \cdot \frac{1}{k} \sum_{\ell=1}^k \kappa(\omega_\ell) \\ & \xrightarrow{k \rightarrow \infty} \mathbb{E}[f(\bar{x}, \bar{y}, \omega)] - \mathbb{E}[f(x^*, y^*, \omega)] \quad \text{w.p.1.} \end{aligned}$$

This indicates that the sequence

$$\rho_k \left\{ \left\| \Phi_{\epsilon_k} \left(y^k, \frac{1}{k} \sum_{\ell=1}^k F(x^k, y^k, \omega_\ell) \right) \right\|^2 - \left\| \Phi_{\epsilon_k} \left(\bar{y}, \frac{1}{k} \sum_{\ell=1}^k F(\bar{x}, \bar{y}, \omega_\ell) \right) \right\|^2 \right\} \quad (4.4)$$

is almost surely bounded above. Since

$$\begin{aligned} & \left\| \Phi_{\epsilon_k} \left(y^k, \frac{1}{k} \sum_{\ell=1}^k F(x^k, y^k, \omega_\ell) \right) - \Phi_{\epsilon_k} \left(y^*, \frac{1}{k} \sum_{\ell=1}^k F(x^*, y^*, \omega_\ell) \right) \right\|^2 \\ & = \sum_{i=1}^m \left[\phi_{\epsilon_k} \left(y_i^k, \frac{1}{k} \sum_{\ell=1}^k F_i(x^k, y^k, \omega_\ell) \right) - \phi_{\epsilon_k} \left(y_i^*, \frac{1}{k} \sum_{\ell=1}^k F_i(x^*, y^*, \omega_\ell) \right) \right]^2 \\ & \leq 4 \sum_{i=1}^m \left[|y_i^k - y_i^*| + \left| \frac{1}{k} \sum_{\ell=1}^k \left(N_i(\omega_\ell)(x^k - x^*) + M_i(\omega_\ell)(y^k - y^*) \right) \right| \right]^2 \\ & \xrightarrow{k \rightarrow \infty} 0 \quad \text{w.p.1} \end{aligned}$$

by Lemma 4.1, we have

$$\lim_{k \rightarrow \infty} \Phi_{\epsilon_k} \left(y^k, \frac{1}{k} \sum_{\ell=1}^k F(x^k, y^k, \omega_\ell) \right) = \Phi_0(y^*, \mathbb{E}[F(x^*, y^*, \omega)]) \quad \text{w.p.1.} \quad (4.5)$$

On the other hand, it follows from $(\bar{x}, \bar{y}) \in \mathcal{F}$ that

$$\lim_{k \rightarrow \infty} \Phi_{\epsilon_k} \left(\bar{y}, \frac{1}{k} \sum_{\ell=1}^k F(\bar{x}, \bar{y}, \omega_\ell) \right) = \Phi_0(\bar{y}, \mathbb{E}[F(\bar{x}, \bar{y}, \omega)]) = 0 \quad \text{w.p.1.}$$

Since the sequence (4.4) is almost surely bounded and $\lim_{k \rightarrow \infty} \rho_k = +\infty$, we have

$$\Phi_0(y^*, \mathbb{E}[F(x^*, y^*, \omega)]) = 0 \quad \text{w.p.1.}$$

Namely, (x^*, y^*) is feasible to (1.1) with probability one.

(b) We next show that (x^*, y^*) is almost surely an optimal solution of problem (1.1). Choose $(\bar{x}, \bar{y}) \in \mathcal{F}$ arbitrarily. It follows that $\phi_0(\bar{y}_i, \mathbb{E}[F_i(\bar{x}, \bar{y}, \omega)]) = 0$ for each i . From Lemma 4.1 and (4.1)–(4.2), we have

$$\begin{aligned} & \rho_k \left\| \Phi_{\epsilon_k} \left(\bar{y}, \frac{1}{k} \sum_{\ell=1}^k F(\bar{x}, \bar{y}, \omega_\ell) \right) \right\|^2 \\ = & \rho_k \sum_{i=1}^m \left[\phi_{\epsilon_k} \left(\bar{y}_i, \frac{1}{k} \sum_{\ell=1}^k F_i(\bar{x}, \bar{y}, \omega_\ell) \right) - \phi_0(\bar{y}_i, \mathbb{E}[F_i(\bar{x}, \bar{y}, \omega)]) \right]^2 \\ \leq & \rho_k \sum_{i=1}^m \left(2 \left| \frac{1}{k} \sum_{\ell=1}^k F_i(\bar{x}, \bar{y}, \omega_\ell) - \mathbb{E}[F_i(\bar{x}, \bar{y}, \omega)] \right| + \epsilon_k \right)^2 \\ = & \sum_{i=1}^m \left[2\sqrt{\rho_k} \left| \left(\frac{1}{k} \sum_{\ell=1}^k N_i(\omega_\ell) - \bar{N}_i \right)^T \bar{x} \right. \right. \\ & \left. \left. + \left(\frac{1}{k} \sum_{\ell=1}^k M_i(\omega_\ell) - \bar{M}_i \right)^T \bar{y} + \left(\frac{1}{k} \sum_{\ell=1}^k q_i(\omega_\ell) - \bar{q}_i \right) \right| + \epsilon_k \sqrt{\rho_k} \right]^2 \\ \xrightarrow{k \rightarrow \infty} & 0 \quad \text{w.p.1.} \end{aligned} \tag{4.6}$$

Moreover, we have from (4.3) that, for every k ,

$$\frac{1}{k} \sum_{\ell=1}^k f(x^k, y^k, \omega_\ell) \leq \frac{1}{k} \sum_{\ell=1}^k f(\bar{x}, \bar{y}, \omega_\ell) + \rho_k \left\| \Phi_{\epsilon_k} \left(\bar{y}, \frac{1}{k} \sum_{\ell=1}^k F(\bar{x}, \bar{y}, \omega_\ell) \right) \right\|^2. \tag{4.7}$$

On the other hand, it follows from the Hölder continuity of f that

$$\begin{aligned} \left| \frac{1}{k} \sum_{\ell=1}^k \left(f(x^*, y^*, \omega_\ell) - f(x^k, y^k, \omega_\ell) \right) \right| & \leq (\|x^k - x^*\| + \|y^k - y^*\|) \cdot \frac{1}{k} \sum_{\ell=1}^k \kappa(\omega_\ell) \\ & \xrightarrow{k \rightarrow \infty} 0 \quad \text{w.p.1,} \end{aligned}$$

which along with (3.2) yields

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{\ell=1}^k f(x^k, y^k, \omega_\ell) = \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{\ell=1}^k f(x^*, y^*, \omega_\ell) = \mathbb{E}[f(x^*, y^*, \omega)] \quad \text{w.p.1.} \tag{4.8}$$

Thus, letting $k \rightarrow +\infty$ in (4.7) and taking (4.6) and (4.8) into account, we obtain

$$\mathbb{E}[f(x^*, y^*, \omega)] \leq \mathbb{E}[f(\bar{x}, \bar{y}, \omega)] \quad \text{w.p.1,}$$

which indicates that (x^*, y^*) is an optimal solution of problem (1.1) with probability one. This completes the proof of the theorem. \blacksquare

4.2 Limiting behavior of stationary points

In general, it is difficult to obtain a global optimal solution of problem (3.3), whereas computation of stationary points is relatively easy. Therefore, it is important to study the limiting behavior of stationary points of problem (3.3). We will use the standard definition of stationarity in nonlinear programming.

Definition 4.2 We say $(x^k, y^k) \in \mathcal{X}$ is *stationary* to (3.3) if there exist Lagrangian multiplier vectors $\alpha^k \in \mathfrak{R}^{s_1}$ and $\beta^k \in \mathfrak{R}^{s_2}$ such that

$$\begin{aligned} \frac{1}{k} \sum_{\ell=1}^k \nabla_{(x,y)} f(x^k, y^k, \omega_\ell) + 2\rho_k \nabla_{(x,y)} \Phi_{\epsilon_k} \left(y^k, \frac{1}{k} \sum_{\ell=1}^k F(x^k, y^k, \omega_\ell) \right) \Phi_{\epsilon_k} \left(y^k, \frac{1}{k} \sum_{\ell=1}^k F(x^k, y^k, \omega_\ell) \right) \\ + \nabla g(x^k, y^k) \alpha^k + \nabla h(x^k, y^k) \beta^k = 0, \end{aligned} \quad (4.9)$$

$$0 \leq \alpha^k \perp -g(x^k, y^k) \geq 0. \quad (4.10)$$

Note that

$$\nabla_x \Phi_{\epsilon_k} \left(y^k, \frac{1}{k} \sum_{\ell=1}^k F(x^k, y^k, \omega_\ell) \right) = \frac{1}{k} \sum_{\ell=1}^k \nabla_x F(x^k, y^k, \omega_\ell) B^k, \quad (4.11)$$

$$\nabla_y \Phi_{\epsilon_k} \left(y^k, \frac{1}{k} \sum_{\ell=1}^k F(x^k, y^k, \omega_\ell) \right) = A^k + \frac{1}{k} \sum_{\ell=1}^k \nabla_y F(x^k, y^k, \omega_\ell) B^k, \quad (4.12)$$

where $A^k := \text{diag}(a_1^k, \dots, a_m^k) \in \mathfrak{R}^{m \times m}$ and $B^k := \text{diag}(b_1^k, \dots, b_m^k) \in \mathfrak{R}^{m \times m}$ with

$$\left. \begin{aligned} a_i^k &:= \partial_a \phi_{\epsilon_k} \left(y_i^k, \frac{1}{k} \sum_{\ell=1}^k F_i(x^k, y^k, \omega_\ell) \right) \\ b_i^k &:= \partial_b \phi_{\epsilon_k} \left(y_i^k, \frac{1}{k} \sum_{\ell=1}^k F_i(x^k, y^k, \omega_\ell) \right) \end{aligned} \right\} \quad i = 1, \dots, m.$$

Here,

$$\partial_a \phi_\epsilon(a, b) = 1 - \frac{a}{\sqrt{a^2 + b^2 + \epsilon^2}}, \quad \partial_b \phi_\epsilon(a, b) = 1 - \frac{b}{\sqrt{a^2 + b^2 + \epsilon^2}}.$$

Theorem 4.2 Suppose both f and $\nabla_{(x,y)} f$ are Hölder continuous in (x, y) on \mathcal{X} with order $\sigma > 0$ and Hölder constant $\kappa(\omega) > 0$ satisfying $\int_\Omega \kappa(\omega) d\zeta(\omega) < +\infty$. Let the parameters ρ_k and ϵ_k be chosen to satisfy (4.1) and (4.2). Let (x^k, y^k) be a stationary point of (3.3) for each k and (x^*, y^*) be an accumulation point of $\{(x^k, y^k)\}$. Suppose that there exists a constant π such that $\theta_k(x^k, y^k) \leq \pi$ for each k and the MPEC-LICQ holds at (x^*, y^*) . Then (x^*, y^*) is almost surely a C -stationary point of (1.1). Furthermore, if the LLSC holds at (x^*, y^*) , it is B -stationary with probability one.

Proof. Assume without loss of generality that $\lim_{k \rightarrow \infty} (x^k, y^k) = (x^*, y^*)$. By the assumptions, we have

$$\frac{1}{k} \sum_{\ell=1}^k f(x^k, y^k, \omega_\ell) + \rho_k \left\| \Phi_{\epsilon_k} \left(y^k, \frac{1}{k} \sum_{\ell=1}^k F(x^k, y^k, \omega_\ell) \right) \right\|^2 = \theta_k(x^k, y^k) \leq \pi, \quad \forall k \quad (4.13)$$

and hence

$$\left\| \Phi_{\epsilon_k} \left(y^k, \frac{1}{k} \sum_{\ell=1}^k F(x^k, y^k, \omega_\ell) \right) \right\|^2 \leq \rho_k^{-1} \left(\pi - \frac{1}{k} \sum_{\ell=1}^k f(x^k, y^k, \omega_\ell) \right), \quad \forall k. \quad (4.14)$$

Note that (4.5) and (4.8) remain valid under the assumptions. Letting $k \rightarrow +\infty$ in (4.14), we have $\Phi_0(y^*, \mathbb{E}[F(x^*, y^*, \omega)]) = 0$ with probability one, which implies that (x^*, y^*) is almost surely a feasible point of (1.1). We next show that (x^*, y^*) is a C-stationary point of problem (1.1) with probability one.

Since (x^k, y^k) is stationary to (3.3), there exist Lagrangian multiplier vectors $\alpha^k \in \mathfrak{R}^{s_1}$ and $\beta^k \in \mathfrak{R}^{s_2}$ satisfying conditions (4.9) and (4.10). Note that, by (4.11)–(4.12), condition (4.9) can be rewritten as

$$\begin{aligned} \frac{1}{k} \sum_{\ell=1}^k \nabla_{(x,y)} f(x^k, y^k, \omega_\ell) - \begin{pmatrix} O \\ I \end{pmatrix} \gamma^k - \frac{1}{k} \sum_{\ell=1}^k \nabla_{(x,y)} F(x^k, y^k, \omega_\ell) \delta^k \\ + \nabla g(x^k, y^k) \alpha^k + \nabla h(x^k, y^k) \beta^k = 0, \end{aligned} \quad (4.15)$$

where

$$\gamma^k := -2\rho_k A^k \Phi_{\epsilon_k} \left(y^k, \frac{1}{k} \sum_{\ell=1}^k F(x^k, y^k, \omega_\ell) \right), \quad (4.16)$$

$$\delta^k := -2\rho_k B^k \Phi_{\epsilon_k} \left(y^k, \frac{1}{k} \sum_{\ell=1}^k F(x^k, y^k, \omega_\ell) \right). \quad (4.17)$$

It follows from (4.13) and (4.8) that the sequence $\left\{ \sqrt{\rho_k} \phi_{\epsilon_k} \left(y_i^k, \frac{1}{k} \sum_{\ell=1}^k F_i(x^k, y^k, \omega_\ell) \right) \right\}$, i.e.,

$$\left\{ \frac{2\sqrt{\rho_k} y_i^k \left(\frac{1}{k} \sum_{\ell=1}^k F_i(x^k, y^k, \omega_\ell) \right) - \epsilon_k^2 \sqrt{\rho_k}}{y_i^k + \frac{1}{k} \sum_{\ell=1}^k F_i(x^k, y^k, \omega_\ell) + \sqrt{(y_i^k)^2 + \left(\frac{1}{k} \sum_{\ell=1}^k F_i(x^k, y^k, \omega_\ell) \right)^2 + \epsilon_k^2}} \right\}, \quad (4.18)$$

is almost surely bounded for each i . Let G and H be defined as in Section 2.

(a) If $i \notin \mathcal{I}_G(x^*, y^*)$, we have $\lim_{k \rightarrow \infty} y_i^k = y_i^* > 0$ and

$$\lim_{k \rightarrow \infty} \frac{1}{k} \sum_{\ell=1}^k F_i(x^k, y^k, \omega_\ell) = \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{\ell=1}^k F_i(x^*, y^*, \omega_\ell) = \mathbb{E}[F_i(x^*, y^*, \omega)] = 0 \quad \text{w.p.1.}$$

It then follows from (4.1) and the boundedness of (4.18) that $\left\{ \sqrt{\rho_k} \left(\frac{1}{k} \sum_{\ell=1}^k F_i(x^k, y^k, \omega_\ell) \right) \right\}$ is almost surely bounded. As a result,

$$\rho_k a_i^k = \frac{\rho_k \left(\frac{1}{k} \sum_{\ell=1}^k F_i(x^k, y^k, \omega_\ell) \right)^2 + \rho_k \epsilon_k^2}{(y_i^k)^2 + \left(\frac{1}{k} \sum_{\ell=1}^k F_i(x^k, y^k, \omega_\ell) \right)^2 + \epsilon_k^2 + y_i^k \sqrt{(y_i^k)^2 + \left(\frac{1}{k} \sum_{\ell=1}^k F_i(x^k, y^k, \omega_\ell) \right)^2 + \epsilon_k^2}}$$

is almost surely bounded. On the other hand, in a similar way to (4.5), we can show that

$$\lim_{k \rightarrow \infty} \phi_{\epsilon_k} \left(y_i^k, \frac{1}{k} \sum_{\ell=1}^k F_i(x^k, y^k, \omega_\ell) \right) = \phi_0(y_i^*, \mathbb{E}[F_i(x^*, y^*, \omega)]) = 0 \quad \text{w.p.1.}$$

In consequence,

$$\lim_{k \rightarrow \infty} \gamma_i^k = - \lim_{k \rightarrow \infty} 2 \rho_k a_i^k \phi_{\epsilon_k} \left(y_i^k, \frac{1}{k} \sum_{\ell=1}^k F_i(x^k, y^k, \omega_\ell) \right) = 0 \quad \text{w.p.1.}$$

Similarly, we can prove that $\lim_{k \rightarrow \infty} \delta_i^k = 0$ with probability one if $i \notin \mathcal{I}_H(x^*, y^*)$.

(b) By the continuity of the functions involved, when k is sufficiently large, there hold

$$\mathcal{I}_g(x^k, y^k) \subseteq \mathcal{I}_g(x^*, y^*), \quad \mathcal{I}_G(x^k, y^k) \subseteq \mathcal{I}_G(x^*, y^*), \quad \mathcal{I}_H(x^k, y^k) \subseteq \mathcal{I}_H(x^*, y^*).$$

Note that, by (4.10), (4.15) can be further rewritten as

$$\begin{aligned} & \frac{1}{k} \sum_{\ell=1}^k \nabla_{(x,y)} f(x^k, y^k, \omega_\ell) - \sum_{i \notin \mathcal{I}_G(x^*, y^*)} \gamma_i^k \begin{pmatrix} 0 \\ e^i \end{pmatrix} - \sum_{i \notin \mathcal{I}_H(x^*, y^*)} \delta_i^k \left(\frac{1}{k} \sum_{\ell=1}^k \nabla_{(x,y)} F_i(x^k, y^k, \omega_\ell) \right) \\ = & \sum_{i \in \mathcal{I}_G(x^*, y^*)} \gamma_i^k \begin{pmatrix} 0 \\ e^i \end{pmatrix} + \sum_{i \in \mathcal{I}_H(x^*, y^*)} \delta_i^k \left(\frac{1}{k} \sum_{\ell=1}^k \nabla_{(x,y)} F_i(x^k, y^k, \omega_\ell) \right) \\ & - \sum_{i \in \mathcal{I}_g(x^*, y^*)} \alpha_i^k \nabla g_i(x^k, y^k) - \nabla h(x^k, y^k) \beta^k, \end{aligned} \quad (4.19)$$

where e^i is the i th unit vector in \mathfrak{R}^m . Note that, from (a), the multiplier sequences that appear on the left-hand side of (4.19) are convergent to zero with probability one. By (a) and the Hölder continuity of $\nabla_{(x,y)} f$ on \mathcal{X} , the left-hand side is convergent to $\mathbb{E}[\nabla_{(x,y)} f(x^*, y^*, \omega)]$ with probability one. Since the MPEC-LICQ holds at (x^*, y^*) , it is not difficult to see that all the multiplier sequences that appear on the right-hand side of (4.19) are convergent with probability one. Letting

$$\alpha^* := \lim_{k \rightarrow \infty} \alpha_i^k, \quad \beta^* := \lim_{k \rightarrow \infty} \beta^k, \quad \gamma^* := \lim_{k \rightarrow \infty} \gamma_i^k, \quad \delta^* := \lim_{k \rightarrow \infty} \delta_i^k$$

and taking a limit in (4.19), we obtain (2.2). Moreover, (2.3)–(2.5) follow from (4.10) and (a) immediately. In addition, since both a_i^k and b_i^k are nonnegative, from (4.16) and (4.17), we have $\gamma_i^* \delta_i^* = \lim_{k \rightarrow \infty} \gamma_i^k \delta_i^k \geq 0$ for each $i \in \mathcal{I}_G(x^*, y^*) \cap \mathcal{I}_H(x^*, y^*)$. Therefore, (x^*, y^*) is a C-stationary point of (1.1) with probability one. If the LLSC holds at (x^*, y^*) , then C-stationarity is equivalent to B-stationarity. This completes the proof of the theorem. \blacksquare

Furthermore, we have the following result.

Theorem 4.3 *Let $(f, \nabla_{(x,y)} f)$ be Hölder continuous in (x, y) on \mathcal{X} with order $\sigma > 0$ and Hölder constant $\kappa(\omega) > 0$ satisfying $\int_{\Omega} \kappa(\omega) d\zeta(\omega) < +\infty$ and the parameters ρ_k and ϵ_k be chosen to satisfy (4.1) and (4.2). Let (x^k, y^k) be a stationary point of (3.3) for each k and (x^*, y^*) be an accumulation point of $\{(x^k, y^k)\}$. Suppose that there exists a constant π such that $\theta_k(x^k, y^k) \leq \pi$ for each k and the MPEC-LICQ holds at (x^*, y^*) . Suppose also that the weak second-order necessary conditions hold at (x^k, y^k) for each k sufficiently large and $\{(x^k, y^k)\}$ is asymptotically weakly nondegenerate. Then (x^*, y^*) is almost surely a B-stationary point of (1.1).*

Let G and H be defined as in Section 2. Roughly speaking, the asymptotically weak nondegeneracy of $\{(x^k, y^k)\}$ means that, for each $i \in \mathcal{I}_G(x^*, y^*) \cap \mathcal{I}_H(x^*, y^*)$, $G_i(x^k, y^k)$ and $H_i(x^k, y^k)$ approach zero in the same order of magnitude. This property is obviously weaker than the LLSC condition. See [5] for more details. Although the results established in this theorem are more interesting and important, its proof is somewhat lengthy and technical. To avoid disturbing the readability, we omit its proof here. One can understand this theorem from Theorem 3.1 in [5] and Theorem 4.2.

5 Choice of Parameters

Suppose that F is given as in Section 4. We now discuss how to choose the parameters ρ_k and ϵ_k so that both (4.1) and (4.2) hold with probability one.

In the case where $(\bar{N}, \bar{M}, \bar{q})$ is known, we can set the parameters as follows: Let $\sigma \in (0, 2)$ and $\lambda > 0$ be given numbers and choose a sequence $\{\bar{\rho}_k\}$ from $(0, +\infty)$ such that $\lim_{k \rightarrow \infty} \bar{\rho}_k = +\infty$. Let $\rho_k := \min\{\bar{\rho}_k, \rho_k^N, \rho_k^M, \rho_k^q\}$ and $\epsilon_k \in (0, \lambda/\rho_k]$, where

$$\begin{aligned}\rho_k^N &:= \min_{1 \leq i \leq m} \left\| \frac{1}{k} \sum_{\ell=1}^k N_i(\omega_\ell) - \bar{N}_i \right\|_1^{-\sigma}, \\ \rho_k^M &:= \min_{1 \leq i \leq m} \left\| \frac{1}{k} \sum_{\ell=1}^k M_i(\omega_\ell) - \bar{M}_i \right\|_1^{-\sigma}, \\ \rho_k^q &:= \left\| \frac{1}{k} \sum_{\ell=1}^k q(\omega_\ell) - \bar{q} \right\|_1^{-\sigma}.\end{aligned}$$

It is easy to see from (3.2) that both (4.1) and (4.2) hold for the above settings.

If some data in $(\bar{N}, \bar{M}, \bar{q})$ are unknown, we suggest to set the parameters as follows.

- Let $\sigma \in (0, 2)$ and $\lambda > 0$ be given scalars. Choose a sequence $\{\bar{\rho}_k\}$ from $(0, +\infty)$ such that

$$\lim_{k \rightarrow \infty} \bar{\rho}_k = +\infty, \quad \lim_{k \rightarrow \infty} \frac{\bar{\rho}_k}{k} = 0. \quad (5.1)$$

- Let $\rho_k := \min\{\bar{\rho}_k, \rho_k^N, \rho_k^M, \rho_k^q\}$, where

$$\begin{aligned}\rho_k^N &:= \min \left\{ \left| \frac{1}{k} \sum_{\ell=1}^k N_{ij}(\omega_\ell) - \bar{N}_{ij} \right|^{-\sigma} : \bar{N}_{ij} \text{ is known} \right\}, \\ \rho_k^M &:= \min \left\{ \left| \frac{1}{k} \sum_{\ell=1}^k M_{ij}(\omega_\ell) - \bar{M}_{ij} \right|^{-\sigma} : \bar{M}_{ij} \text{ is known} \right\}, \\ \rho_k^q &:= \min \left\{ \left| \frac{1}{k} \sum_{\ell=1}^k q_i(\omega_\ell) - \bar{q}_i \right|^{-\sigma} : \bar{q}_i \text{ is known} \right\}.\end{aligned}$$

- Choose $\epsilon_k \in (0, \lambda/\rho_k]$.

Then, we have (4.1)–(4.2) at least in probability. In fact, it is obvious that (4.1) holds with probability one. Moreover, if \bar{N}_{ij} , \bar{M}_{ij} or \bar{q}_i is known, we have

$$\begin{aligned} \lim_{k \rightarrow \infty} \sqrt{\rho_k^N} \left| \frac{1}{k} \sum_{\ell=1}^k N_{ij}(\omega_\ell) - \bar{N}_{ij} \right| &\leq \lim_{k \rightarrow \infty} \left| \frac{1}{k} \sum_{\ell=1}^k N_{ij}(\omega_\ell) - \bar{N}_{ij} \right|^{1-\sigma/2} = 0, \\ \lim_{k \rightarrow \infty} \sqrt{\rho_k^M} \left| \frac{1}{k} \sum_{\ell=1}^k M_{ij}(\omega_\ell) - \bar{M}_{ij} \right| &\leq \lim_{k \rightarrow \infty} \left| \frac{1}{k} \sum_{\ell=1}^k M_{ij}(\omega_\ell) - \bar{M}_{ij} \right|^{1-\sigma/2} = 0, \\ \lim_{k \rightarrow \infty} \sqrt{\rho_k^q} \left| \frac{1}{k} \sum_{\ell=1}^k q_i(\omega_\ell) - \bar{q}_i \right| &\leq \lim_{k \rightarrow \infty} \left| \frac{1}{k} \sum_{\ell=1}^k q_i(\omega_\ell) - \bar{q}_i \right|^{1-\sigma/2} = 0 \end{aligned}$$

with probability one; otherwise, we have

$$\begin{aligned} \lim_{k \rightarrow \infty} \sqrt{\bar{\rho}_k} \left(\frac{1}{k} \sum_{\ell=1}^k N_{ij}(\omega_\ell) - \bar{N}_{ij} \right) &= \lim_{k \rightarrow \infty} \sqrt{\frac{\bar{\rho}_k}{k}} \sqrt{k} \left(\frac{1}{k} \sum_{\ell=1}^k N_{ij}(\omega_\ell) - \bar{N}_{ij} \right) = 0, \\ \lim_{k \rightarrow \infty} \sqrt{\bar{\rho}_k} \left(\frac{1}{k} \sum_{\ell=1}^k M_{ij}(\omega_\ell) - \bar{M}_{ij} \right) &= \lim_{k \rightarrow \infty} \sqrt{\frac{\bar{\rho}_k}{k}} \sqrt{k} \left(\frac{1}{k} \sum_{\ell=1}^k M_{ij}(\omega_\ell) - \bar{M}_{ij} \right) = 0, \\ \lim_{k \rightarrow \infty} \sqrt{\bar{\rho}_k} \left(\frac{1}{k} \sum_{\ell=1}^k q_i(\omega_\ell) - \bar{q}_i \right) &= \lim_{k \rightarrow \infty} \sqrt{\frac{\bar{\rho}_k}{k}} \sqrt{k} \left(\frac{1}{k} \sum_{\ell=1}^k q_i(\omega_\ell) - \bar{q}_i \right) = 0 \end{aligned}$$

in probability, since the convergence in (3.2) is of order $O(k^{-1/2})$ in probability [6], which implies that $\left\{ \sqrt{k} \left(\frac{1}{k} \sum_{\ell=1}^k \psi(\omega_\ell) - \mathbb{E}[\psi(\omega)] \right) \right\}$ is convergent in probability as $k \rightarrow +\infty$. Therefore, from the manner in which ρ_k is determined, we have (4.2) in probability.

Remark 5.1 Another strategy for choosing ρ_k is simply to set $\rho_k := \bar{\rho}_k$ for every k , where $\bar{\rho}_k$ is chosen to satisfy (5.1). However, in order to ensure that more conditions in (4.2) hold with probability one (not just in probability), we make most of the data $(N(\omega_\ell), M(\omega_\ell), q(\omega_\ell))$ in the definition of ρ_k .

6 Extensions to Quasi-Monte Carlo Approach

We have presented a Monte Carlo sampling and penalty approach for solving problem (1.1). Actually, Monte Carlo sampling methods have been proved useful in the evaluation of integration. However, the convergence of Monte Carlo methods is not fast and various techniques have been proposed to speed up the convergence. In this area, the most well-known innovation is the introduction of quasi-Monte Carlo methods, in which the integral is evaluated by using deterministic sequences rather than random sequences. These deterministic sequences have the property that they are well dispersed throughout the domain of integration. Sequences with this property are called *low discrepancy sequences*. See the monograph [14] for more details.

Next, we briefly introduce two advantages of quasi-Monte Carlo methods.

- (i) Since quasi-Monte Carlo methods employ deterministic sequences instead of random sequences, the convergence in (3.2) is valid in a deterministic way for any integrable function $\psi : \Omega \rightarrow \mathfrak{R}$. This is different from Monte Carlo methods, for which convergence is always probabilistic.
- (ii) Quasi-Monte Carlo methods are generally faster than Monte Carlo methods in numerical integration. Actually, the expected convergence in (3.2) is of order $O(k^{-1/2})$ for Monte Carlo methods, whereas the worst case convergence for quasi-Monte Carlo methods is of order $O(\frac{(\log k)^d}{k})$, where k is the number of samples and d is the dimension of the integration.

We may readily develop a quasi-Monte Carlo and penalty approach for solving problem (1.1). In the case where F is affine, we can establish all the results in Section 4 in a similar way, and particularly, those convergence results are deterministic by (i). Moreover, by (ii), the choice of the parameter ρ_k given in Section 5 can also be improved. For example, we may choose the sequence $\{\bar{\rho}_k\}$ from $(0, +\infty)$ such that

$$\lim_{k \rightarrow \infty} \bar{\rho}_k = +\infty, \quad \lim_{k \rightarrow \infty} \frac{\bar{\rho}_k (\log k)^{2d}}{k^2} = 0$$

instead of (5.1). Then, we may expect that the quasi-Monte Carlo sampling and penalty method is faster than the method suggested in Section 3.

7 Applications

Consider a supply side oligopoly market where $(m + 1)$ firms compete to supply a homogeneous product in a non-competitive manner. A dominant firm, called the *leader* hereafter, knows how the other firms (called followers) react to its supply and chooses optimal supply to maximize its profit by expecting the other firms to reach a Nash-Cournot equilibrium after its supply is determined. It is well known that such a market competition can be modeled as a Stackelberg-Nash-Cournot game.

Now suppose that the market demand is unknown at the time when the firms make decisions on their supplies and the demand contains some uncertainties. Assume also that all firms know the distribution of the random factors in the demand. Then each firm may consider the expected profit rather than the profit in a particular demand scenario in its decision making.

In what follows, we demonstrate that this type of Stackelberg leader-follower games can be modeled as (1.1). We start by describing the market demand with the inverse demand function $p(\tau, \omega)$, where τ stands for the total quantity of supply to the market, ω is a random shock with known distribution, and $p(\tau, \omega)$ is the market price.

Let x denote the decision variable of the leader, that is, the quantity supplied by the leader to the market. Let y_i denote the decision variable of the i th follower, that is, the quantity supplied by the i th firm to the market.

The Followers' Decision Problems. Suppose that the leader's supply is x and the aggregate supplies of the followers except the i th firm is $\sum_{j=1, j \neq i}^m y_j$. If the i th firm's supply is y_i , then the market price in this demand scenario is $p(x + \sum_{j=1}^m y_j, \omega)$. The total revenue of the i th firm is $y_i p(x + \sum_{j=1}^m y_j, \omega)$. Suppose that the total cost for the i th firm to produce y_i is $c_i(y_i)$. Then the i th firm's expected profit can be formulated as

$$\mathbb{E} \left[y_i p \left(x + \sum_{j=1}^m y_j, \omega \right) \right] - c_i(y_i).$$

Since the market price depends on y_i (in other words, the i th firm has market power), the i th firm would like to choose an optimal y_i in order to maximize his expected profit. Therefore the i th follower's profit maximization problem can be written as

$$\max_{y_i \geq 0} f_i(y_i) := \mathbb{E} \left[y_i p \left(x + y_i + \sum_{j=1, j \neq i}^m y_j, \omega \right) \right] - c_i(y_i). \quad (7.1)$$

In choosing an optimal decision, the i th firm holds the other firms' supplies as constants. A *Nash-Cournot equilibrium* among the followers is a situation where, given the leader's supply, no firm can improve its expected profit by unilaterally changing his supply. We denote such an equilibrium by $(y_1(x), \dots, y_m(x))$, where each $y_i(x)$ is a global optimal solution of (7.1) with $y_j = y_j(x)$ for all $j \neq i$.

The Leader's Decision Problem. We suppose that the leader expects the followers to choose their outputs as described in (7.1) and maximizes his expected profit based on his knowledge on the market demand distribution and the followers' reaction to his supply. Therefore we can formulate the leader's decision problem as follows:

$$\max_{0 \leq x \leq L} f_0(x) := \mathbb{E} \left[x p \left(x + \sum_{i=1}^m y_i(x), \omega \right) \right] - c_0(x),$$

where $L > 0$ is a constant and $c_0(x)$ is the cost for the leader to produce x .

Stochastic Stackelberg-Nash-Cournot Equilibrium. We investigate a situation where the leader maximizes the expected profit while the followers reach a Nash-Cournot equilibrium. A *Stackelberg-Nash-Cournot equilibrium* is an $(m+1)$ -dimensional vector $(x^*, y_1(x^*), \dots, y_m(x^*))$ such that

$$f_0(x^*) = \max_{0 \leq x \leq L} \mathbb{E} \left[x p \left(x + \sum_{i=1}^m y_i(x), \omega \right) \right] - c_0(x)$$

with

$$y_i(x) \in \text{Arg max}_{y_i \geq 0} \left(\mathbb{E} \left[y_i p \left(x + y_i + \sum_{j=1, j \neq i}^m y_j(x), \omega \right) \right] - c_i(y_i) \right), \quad i = 1, \dots, m. \quad (7.2)$$

If the function $\mathbb{E}[y_i p(x + y_i + \sum_{j=1, j \neq i}^m y_j(x), \omega)] - c_i(y_i)$ is concave in y_i , the Nash-Cournot equilibrium problem (7.2) is equivalent to the following nonlinear complementarity problem:

$$0 \leq y \perp \mathbb{E}[F(x, y, \omega)] \geq 0,$$

where

$$F(x, y, \omega) := -p(x + y^T \mathbf{e}, \omega) \mathbf{e} - p'_\tau(x + y^T \mathbf{e}, \omega) y + \mathbf{c}'(y).$$

Here, $\mathbf{e} := (1, \dots, 1)^T \in \mathfrak{R}^m$ and $\mathbf{c}'(y) := (c'_1(y_1), \dots, c'_m(y_m))^T$. Thus, we can rewrite the stochastic Stackelberg-Nash-Cournot equilibrium problem as an SMPEC:

$$\begin{aligned} \max \quad & \mathbb{E} \left[x p(x + \mathbf{e}^T y, \omega) \right] - c_0(x) \\ \text{s.t.} \quad & 0 \leq x \leq L, \\ & 0 \leq y \perp \mathbb{E}[F(x, y, \omega)] \geq 0. \end{aligned} \tag{7.3}$$

Obviously (7.3) is subsumed by (1.1).

Remark 7.1 Suppose that $p(\tau, \omega) := \alpha(\omega) - \beta(\omega)\tau$ with $\mathbb{E}[\alpha(\omega)] > 0$ and $\mathbb{E}[\beta(\omega)] \geq 0$ and $c_i(y_i)$ is affine. It is easy to show that the function $\mathbb{E}[y_i p(x + y_i + \sum_{j=1, j \neq i}^m y_j(x), \omega)] - c_i(y_i)$ is concave in y_i .

As an application of the proposed methods, we consider a simple case in which there are three followers and the involved functions are given by

$$\begin{aligned} p(\tau, \omega) &:= 20 - (0.002\omega + 0.003)\tau, \\ c_0(x) &:= 9.5x + 60, \\ c_1(y_1) &:= 8.6y_1 + 48, \\ c_2(y_2) &:= 8.9y_2 + 45, \\ c_3(y_3) &:= 9.2y_3 + 75, \end{aligned}$$

respectively. We suppose that the random shock ω is uniformly distributed on $\Omega := [-1, 1]$ and the maximum amount L of the leader is equal to 1800. Then the model (7.3) becomes

$$\begin{aligned} \max \quad & \mathbb{E}[x(20 - (0.002\omega + 0.003)(x + y_1 + y_2 + y_3) + \omega)] - c_0(x) \\ \text{s.t.} \quad & 0 \leq x \leq 1800, \\ & 0 \leq y \perp \mathbb{E}[F(x, y, \omega)] \geq 0 \end{aligned}$$

with $F(x, y, \omega) := N(\omega)x + M(\omega)y + q$, where

$$N(\omega) := \begin{pmatrix} 0.002\omega + 0.003 \\ 0.002\omega + 0.003 \\ 0.002\omega + 0.003 \end{pmatrix}, \quad M(\omega) := \begin{pmatrix} 0.004\omega + 0.006 & 0.002\omega + 0.003 & 0.002\omega + 0.003 \\ 0.002\omega + 0.003 & 0.004\omega + 0.006 & 0.002\omega + 0.003 \\ 0.002\omega + 0.003 & 0.002\omega + 0.003 & 0.004\omega + 0.006 \end{pmatrix},$$

and $q := -(11.4, 11.1, 10.8)^T$. The solution of this problem is $(x^*, y^*) = (1450, 662.5, 562.5, 462.5)$.

We applied the proposed methods to solve the above problem. In our experiments, in order to demonstrate the methods, we treated the expectations \bar{N} and \bar{M} as unknown data although they are easy to be calculated. For the Monte Carlo sampling method, we set $\bar{\rho}_k = k^{3/4}$, $\epsilon_k = \rho_k^{-1}$, and we used the random number generator `rand` in Matlab 6.5 to generate random samples

$\{\xi_1, \dots, \xi_k\}$ from $[0, 1]$ and then let $\omega_i = 2\xi_i - 1 \in \Omega$ for each $i = 1, \dots, k$. For the quasi-Monte Carlo sampling method, we set $\bar{\rho}_k = k$, $\epsilon_k = \rho_k^{-1}$, and used the classical constructions method in [14] to generate samples. Then, we employed the solver `fmincon` in Matlab 6.5 to solve the subproblems (3.3). The initial points were chosen to be $(0, \dots, 0)$ and the computed solutions were used as the starting points in the next iterations. The computational results are shown in Table 1. The results shown in the table reveal that the proposed methods were able to solve the problem successfully and the quasi-Monte Carlo method was faster than the Monte Carlo method.

Table 1: Computational Results

	(x^*, y^*)	
	MC	QMC
$k = 10^2$	(1800.0, 60.5, 54.4, 48.7)	(1800.0, 383.2, 288.4, 199.3)
$k = 10^3$	(1800.0, 469.3, 369.7, 271.5)	(1546.6, 622.2, 522.3, 422.2)
$k = 10^4$	(1546.6, 622.2, 522.3, 422.3)	(1459.2, 658.7, 558.7, 458.7)
$k = 10^5$	(1466.3, 655.7, 555.7, 455.7)	(1450.9, 662.1, 562.1, 462.1)
$k = 10^6$	(1452.9, 661.3, 561.3, 461.3)	(1451.0, 662.2, 562.2, 462.2)
$k = 10^7$	(1453.1, 661.6, 561.6, 461.6)	(1450.9, 662.3, 562.3, 462.3)

8 Conclusion

We have presented Monte Carlo and quasi-Monte Carlo sampling methods with a penalty technique for solving problem (1.1) and, under appropriate assumptions, we have established a comprehensive convergence theory for the proposed methods. Especially, different from the approach proposed in [1], the approximation problems given in this paper are standard differentiable optimization problems and hence they are easy to deal with.

References

- [1] S.I. Birbil, G. Gürkan, and O. Listes, *Simulation-based solution of stochastic mathematical programs with complementarity constraints: Sample-path analysis*, Econometric Institute Report EI 2004-03, Erasmus University Rotterdam, 2004. <https://ep.eur.nl/bitstream/1765/1164/1/ERS+2004+016+LIS.pdf>.
- [2] J.R. Birge and F. Louveaux, *Introduction to Stochastic Programming*, Springer, New York, 1997.

- [3] Y. Chen and M. Florian, *The nonlinear bilevel programming problem: Formulations, regularity and optimality conditions*, Optimization, 32 (1995), 193–209.
- [4] M. Fukushima and G.H. Lin, *Smoothing methods for mathematical programs with equilibrium constraints*, Proceedings of the ICKS'04, IEEE Computer Society, 2004, 206–213.
- [5] M. Fukushima and J.S. Pang, *Convergence of a smoothing continuation method for mathematical problems with complementarity constraints*, Illposed Variational Problems and Regularization Techniques, M. Théra and R. Tichatschke (eds.), Lecture Notes in Economics and Mathematical Systems, Springer, 477 (1999), 105–116.
- [6] P. Hall and J.S. Marron, *Local minima in cross-validation functions*, Journal of the Royal Statistical Society, Series B, 53 (1991), 245–252.
- [7] H. Jiang and D. Ralph, *Smooth methods for mathematical programs with nonlinear complementarity constraints*, SIAM Journal on Optimization, 10 (2000), 779–808.
- [8] G.H. Lin, X. Chen and M. Fukushima, *Smoothing implicit programming approaches for stochastic mathematical programs with linear complementarity constraints*, Technical Report 2003-006, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto, Japan, 2003. <http://www-optima.amp.i.kyoto-u.ac.jp/~fuku/papers/smpec01-rev.pdf>.
- [9] G.H. Lin, X. Chen and M. Fukushima, *Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization*, Technical Report 2005-003, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto, Japan, 2005. http://www.amp.i.kyoto-u.ac.jp/tecrep/ps_file/2005/2005-003.pdf.
- [10] G.H. Lin and M. Fukushima, *A class of stochastic mathematical programs with complementarity constraints: Reformulations and algorithms*, Journal of Industrial and Management Optimization, 1 (2005), 99–122.
- [11] G.H. Lin and M. Fukushima, *Monte Carlo sampling and penalty method for stochastic mathematical programs with complementarity constraints and recourse*, Technical Report 2005-005, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto, Japan, 2005. http://www.amp.i.kyoto-u.ac.jp/tecrep/ps_file/2005/2005-005.pdf.
- [12] G.H. Lin and M. Fukushima, *Regularization method for stochastic mathematical programs with complementarity constraints*, European Series of Applied and Industrial Mathematics (ESAIM): Control, Optimisation and Calculus of Variations, 11 (2005), 252–265.

- [13] Z.Q. Luo, J.S. Pang, and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, United Kingdom, 1996.
- [14] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, SIAM, Philadelphia, 1992.
- [15] J.M. ORTEGA AND W.C. RHEINBOLDT, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, NY, 1970.
- [16] H. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity*, *Mathematics of Operations Research*, 25 (2000), 1–22.
- [17] A. Shapiro, *Monte Carlo sampling approach to stochastic programming*, *European Series of Applied and Industrial Mathematics (ESAIM): Proceedings*, 13 (2003), 65–73.
- [18] A. Shapiro, *Stochastic mathematical programs with equilibrium constraints*, *Journal Optimization Theory and Applications*, to appear. http://www.optimization-online.org/DB_HTML/2004/08/930.html.
- [19] A. Shapiro and H. Xu, *Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation*, Preprint, School of Industrial and System Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA, 2005. http://www.optimization-online.org/DB_HTML/2005/01/1046.html.
- [20] H. Xu, *An implicit programming approach for a class of stochastic mathematical programs with linear complementarity constraints*, *SIAM Journal of Optimization*, to appear.