Math Meth Oper Res (2008) 67:423–441
DOI 10.1007/s00186-007-0201-x

ORIGINAL ARTICLE

Monte Carlo and quasi-Monte Carlo sampling methods for a class of stochastic mathematical programs with equilibrium constraints

Gui-Hua Lin · Huifu Xu · Masao Fukushima

Published online: 8 January 2008

Abstract In this paper, we consider a class of stochastic mathematical programs with equilibrium constraints introduced by Birbil et al. (Math Oper Res 31:739–760, 2006). 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.

Keywords Stochastic mathematical program with equilibrium constraints · Monte Carlo/quasi-Monte Carlo methods · Penalization

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.

G.-H. Lin (✉)
Department of Applied Mathematics, Dalian University of Technology, Dalian 116024, China
e-mail: lin_g_h@yahoo.com.cn

H. Xu
School of Mathematics, University of Southampton, Highfield Southamptom, UK
e-mail: H.Xu@soton.ac.uk

M. Fukushima
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

Mathematics Subject Classification (2000) 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:

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

where $\mathbb{E}$ denotes expectation with respect to the random variable $\omega \in \Omega$, the functions $f : \Re^{n+m} \times \Omega \to \Re$, $g : \Re^{n+m} \to \Re^{s_1}$, $h : \Re^{n+m} \to \Re^{s_2}$, and $F : \Re^{n+m} \times \Omega \to \Re^m$ are all twice continuously differentiable, and the symbol $\perp$ stands for orthogonality of the two vectors on both sides. 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 (Fukushima and Lin 2004, Jiang and Ralph 2000, Luo et al. 1996). Throughout the paper, we suppose that Ω has an infinite number of samples.

Recently there has been quite active research on various formulations of SMPECs such as the *lower-level wait-and-see* model (Lin et al. 2003, Shapiro 2006, Shapiro and Xu 2005, Xu 2006) and the *here-and-now* model (Birbil et al. 2006, Lin et al. 2003, 2007, Lin and Fukushima 2005a,b). Despite the high potential of practical applicability; however, the formulation (1.1) has rarely been studied except the recent work of Birbil et al. (2006) 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 consider a Monte Carlo sampling method combined with a penalty technique for solving problem (1.1). That is, we discretize the true problem by approximating the expected value of the underlying functions with its sample average and then solve the sample averaged approximation problem with a smoothed penalty method. Monte Carlo sampling method is a very popular method in stochastic programming and it is essentially the same as the sample path method. Our focus here is on the combination of the Monte Carlo sampling method with a smoothing penalty method for solving the SMPEC problem where the smoothing parameter and penalty parameter depend on the sample size. This is numerically meaningful in that we do not want to solve a discretized problem with some penalty parameter rather than driving the parameter to infinity to get an exact solution. This distinguishes our approach from the work (Birbil et al. 2006) which discusses the limiting behavior of the optimal solutions of a discretized (approximate) SMPEC without referring to a particular numerical method for solving the discretized SMPEC.

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 $c : \Re^s \to \Re^{s'}$ and a vector $t \in \Re^s$, $\nabla c(t) \in \Re^{s \times s'}$ 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 ith 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

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

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

$$\begin{aligned}&\big\{\nabla g_i(x^*, y^*), \nabla h_j(x^*, y^*), \nabla G_\iota(x^*, y^*), \nabla H_\jmath(x^*, y^*) :\\ &\quad i \in \mathcal{I}_g(x^*, y^*), j \in \{1, \ldots, s_2\}, \iota \in \mathcal{I}_G(x^*, y^*), \jmath \in \mathcal{I}_H(x^*, y^*)\big\}\end{aligned} \tag{2.1}$$

is linearly independent.

Definition 2.2 (Scheel and Scholtes 2000) Suppose that there exist Lagrangian multiplier vectors $\alpha^* \in \Re^{s_1}$, $\beta^* \in \Re^{s_2}$, and $\gamma^*, \delta^* \in \Re^m$ such that

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

$$0 \le \alpha^* \perp -g(x^*, y^*) \ge 0, \tag{2.3}$$

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

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

- We call (x^*, y^*) a *Clarke or C-stationary* point of (1.1) if $\gamma_i^* \delta_i^* \ge 0$ 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 $\gamma_i^* \ge 0$ and $\delta_i^* \ge 0$ hold 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 $\phi_\epsilon : \Re^2 \to \Re$ 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 $\Phi_\epsilon : \Re^{2m} \to \Re^m$ 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

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

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 (Chen and Florian 1995), 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 an integrable function $\psi : \Omega \to \Re$, the Monte Carlo sampling estimate for $\mathbb{E}[\psi(\omega)]$ is obtained by taking independently and identically distributed random samples $\{\omega_1, \ldots, \omega_k\}$ 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.,

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

where $\zeta(\omega)$ is the distribution function of ω. See Birge and Louveaux (1997), Niederreiter (1992) and Shapiro (2003) for more details.

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

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

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:

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

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 $N : \Omega \to \Re^{m\times n}$, $M : \Omega \to \Re^{m\times m}$, and $q : \Omega \to \Re^m$ 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:

$$\lim_{k\to\infty} \rho_k = +\infty, \qquad \limsup_{k\to\infty} \rho_k\epsilon_k < +\infty, \tag{4.1}$$

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

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

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*

$$|\phi_\epsilon(a_1, b_1) - \phi_\epsilon(a_2, b_2)| \leq 2(|a_1 - a_2| + |b_1 - b_2|),$$
$$|\phi_\epsilon(a_1, b_1) - \phi_0(a_2, b_2)| \leq 2(|a_1 - a_2| + |b_1 - b_2|) + \epsilon.$$

Definition 4.1 (Ortega and Rheinboldt 1970) Let $\sigma > 0$ and $\kappa \geq 0$ be constants. We say $G : \Re^s \to \Re^t$ to be *Hölder continuous* on $K \subseteq \Re^s$ 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 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 $\lim_{k\to\infty} (x^k, y^k) = (x^*, y^*)$. Since (x^k, y^k) is an optimal solution of problem (3.3), it follows that

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

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

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

This indicates that the sequence

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

is almost surely bounded above. Since

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

by Lemma 4.1, we have

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

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

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

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

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

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

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

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

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

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

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

which along with (3.2) yields

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

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

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

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

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 $\alpha^k \in \Re^{s_1}$ and $\beta^k \in \Re^{s_2}$ such that

$$\frac{1}{k}\sum_{\ell=1}^{k}\nabla_{(x,y)} f(x^k, y^k, \omega_\ell) + 2\rho_k\, \nabla_{(x,y)} \Phi_{\epsilon_k}\left(y^k, \frac{1}{k}\sum_{\ell=1}^{k} F(x^k, y^k, \omega_\ell)\right)$$

$$\times \Phi_{\epsilon_k}\left(y^k, \frac{1}{k}\sum_{\ell=1}^{k} F(x^k, y^k, \omega_\ell)\right) + \nabla g(x^k, y^k)\alpha^k + \nabla h(x^k, y^k)\beta^k = 0, \quad (4.9)$$

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

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 $A^k := \mathrm{diag}(a_1^k, \ldots, a_m^k) \in \Re^{m\times m}$ and $B^k := \mathrm{diag}(b_1^k, \ldots, b_m^k) \in \Re^{m\times m}$ with

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

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 $\theta_k(x^k, y^k) \le \pi$ 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 $\lim_{k\to\infty}(x^k, y^k) = (x^*, y^*)$. By the assumptions, we have

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

and hence

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

Note that (4.5) and (4.8) remain valid under the assumptions. Letting $k \to +\infty$ 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 $\alpha^k \in \Re^{s_1}$ and $\beta^k \in \Re^{s_2}$ satisfying conditions (4.9) and (4.10). Note that, by (4.11)–(4.12), condition (4.9) can be rewritten as

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

where

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

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

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

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

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

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

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

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

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

In consequence,

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

Similarly, we can prove that $\lim_{k\to\infty} \delta_i^k = 0$ 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

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

where e^i is the ith unit vector in $\Re^m$. 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

$$\alpha^* := \lim_{k\to\infty} \alpha^k, \quad \beta^* := \lim_{k\to\infty} \beta^k, \quad \gamma^* := \lim_{k\to\infty} \gamma^k, \quad \delta^* := \lim_{k\to\infty} \delta^k$$

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 $\gamma_i^* \delta_i^* = \lim_{k\to\infty} \gamma_i^k \delta_i^k \geq 0$ 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. □

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 Sect. 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 Fukushima and Pang (1999) 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 Fukushima and Pang (1999) and Theorem 4.2.

5 Choice of parameters

Suppose that F is given as in Sect. 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 $\lim_{k\to\infty} \bar{\rho}_k = +\infty$. 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

$$\rho_k^N := \min_{1\le i\le m} \left\| \frac{1}{k} \sum_{\ell=1}^{k} N_i(\omega_\ell) - \bar{N}_i \right\|_1^{-\sigma},$$

$$\rho_k^M := \min_{1\le i\le m} \left\| \frac{1}{k} \sum_{\ell=1}^{k} M_i(\omega_\ell) - \bar{M}_i \right\|_1^{-\sigma},$$

$$\rho_k^q := \left\| \frac{1}{k} \sum_{\ell=1}^{k} q(\omega_\ell) - \bar{q} \right\|_1^{-\sigma}.$$

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

$$\lim_{k\to\infty} \bar{\rho}_k = +\infty, \qquad \lim_{k\to\infty} \frac{\bar{\rho}_k}{k} = 0. \tag{5.1}$$

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

$$\rho_k^N := \min\left\{ \left| \frac{1}{k} \sum_{\ell=1}^{k} N_{ij}(\omega_\ell) - \bar{N}_{ij} \right|^{-\sigma} : \ \bar{N}_{ij} \text{ is known} \right\},$$

$$\rho_k^M := \min\left\{ \left| \frac{1}{k} \sum_{\ell=1}^{k} M_{ij}(\omega_\ell) - \bar{M}_{ij} \right|^{-\sigma} : \ \bar{M}_{ij} \text{ is known} \right\},$$

$$\rho_k^q := \min\left\{ \left| \frac{1}{k} \sum_{\ell=1}^{k} q_i(\omega_\ell) - \bar{q}_i \right|^{-\sigma} : \ \bar{q}_i \text{ is known} \right\}.$$

- 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

$$\lim_{k\to\infty} \sqrt{\rho_k^N} \left| \frac{1}{k} \sum_{\ell=1}^{k} N_{ij}(\omega_\ell) - \bar{N}_{ij} \right| \le \lim_{k\to\infty} \left| \frac{1}{k} \sum_{\ell=1}^{k} N_{ij}(\omega_\ell) - \bar{N}_{ij} \right|^{1-\sigma/2} = 0,$$

$$\lim_{k\to\infty} \sqrt{\rho_k^M} \left| \frac{1}{k} \sum_{\ell=1}^{k} M_{ij}(\omega_\ell) - \bar{M}_{ij} \right| \le \lim_{k\to\infty} \left| \frac{1}{k} \sum_{\ell=1}^{k} M_{ij}(\omega_\ell) - \bar{M}_{ij} \right|^{1-\sigma/2} = 0,$$

$$\lim_{k\to\infty} \sqrt{\rho_k^q} \left| \frac{1}{k} \sum_{\ell=1}^{k} q_i(\omega_\ell) - \bar{q}_i \right| \le \lim_{k\to\infty} \left| \frac{1}{k} \sum_{\ell=1}^{k} q_i(\omega_\ell) - \bar{q}_i \right|^{1-\sigma/2} = 0$$

with probability one; otherwise, we have

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

in probability, since the convergence in (3.2) is of order $O(k^{-1/2})$ in probability (Hall and Marron 1991), 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 $k \to +\infty$. 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 (Niederreiter 1992) 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 $\psi : \Omega \to \Re$. 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 $O\left(\frac{(\log k)^d}{k}\right)$, 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 Sect. 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 Sect. 5 can also be improved. For example, we may choose the sequence $\{\bar{\rho}_k\}$ from $(0, +\infty)$ such that

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

instead of (5.1). Then, we may expect that the quasi-Monte Carlo sampling and penalty method is faster than the method suggested in Sect. 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 ith follower, that is, the quantity supplied by the ith 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 ith firm is $\sum_{j=1, j\neq i}^{m} y_j$. If the ith 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 ith firm is $y_i\, p(x + \sum_{j=1}^{m} y_j, \omega)$. Suppose that the total cost for the ith firm to produce y_i is $c_i(y_i)$. Then the ith 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 ith firm has market power), the ith firm would like to choose an optimal y_i in order to maximize his expected profit. Therefore the ith follower's profit maximization problem can be written as

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

In choosing an optimal decision, the ith 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 $(y_1(x), \ldots, y_m(x))$, 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 $(x^*, y_1(x^*), \ldots, y_m(x^*))$ 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

$$y_i(x) \in \operatorname{Arg}\max_{y_i \geq 0} \left(\mathbb{E}\left[y_i\, p\left(x + y_i + \sum_{j=1, j\neq i}^{m} y_j(x), \omega \right)\right] - c_i(y_i) \right),$$
$$i = 1, \ldots, m. \tag{7.2}$$

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, $\mathbf{e} := (1, \ldots, 1)^T \in \Re^m$ and $\mathbf{c}'(y) := (c_1'(y_1), \ldots, c_m'(y_m))^T$. Thus, we can rewrite the stochastic Stackelberg–Nash–Cournot equilibrium problem as an SMPEC:

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

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 $\mathbb{E}[\beta(\omega)] \ge 0$ 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

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

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

$$N(\omega) := \begin{pmatrix} 0.002\omega + 0.003 \\ 0.002\omega + 0.003 \\ 0.002\omega + 0.003 \end{pmatrix},$$

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

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 calculate. 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

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)

6.5 to generate random samples $\{\xi_1, \ldots, \xi_k\}$ from $[0, 1]$ and then let $\omega_i = 2\xi_i - 1 \in \Omega$ for each $i = 1, \ldots, k$. 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 Niederreiter (1992) 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 $(0, \ldots, 0)$ 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.

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 Birbil et al. (2006), the approximation problems given in this paper are standard differentiable optimization problems and hence they are easy to deal with.

References

Birbil SI, Gürkan G, Listes O (2006) Solving stochastic mathematical programs with complementarity constraints using simulation. Math Oper Res 31:739–760

Birge JR, Louveaux F (1997) Introduction to stochastic programming. Springer, New York

Chen Y, Florian M (1995) The nonlinear bilevel programming problem: formulations, regularity and optimality conditions. Optimization 32:193–209

Fukushima M, Lin GH (2004) Smoothing methods for mathematical programs with equilibrium constraints. Proceedings of the ICKS'04. IEEE Computer Society, New York, pp 206–213

Fukushima M, Pang JS (1999) Convergence of a smoothing continuation method for mathematical problems with complementarity constraints, illposed variational problems and regularization techniques. In: Théra M, Tichatschke R (eds) Lecture Notes in Economics and Mathematical Systems, vol 477. Springer, Heidelberg, pp 105–116

Hall P, Marron JS (1991) Local minima in cross-validation functions. J Roy Stat Soc Ser B 53:245–252

Jiang H, Ralph D (2000) Smooth SQP methods for mathematical programs with nonlinear complementarity constraints. SIAM J Optim 10:779–808

Lin GH, Fukushima M (2005a) A class of stochastic mathematical programs with complementarity constraints: Reformulations and algorithms. J Ind Manage Optim 1:99–122

Lin GH, Fukushima M (2005b) Regularization method for stochastic mathematical programs with complementarity constraints. Eur Ser Appl Ind Math (ESAIM): Control Optim Calculus Variations 11: 252–265

Lin GH, Chen X, Fukushima M (2003) Smoothing implicit programming approaches for stochastic mathematical programs with linear complementarity constraints. Technical Report 2003–2006, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto, Japan

Lin GH, Chen X, Fukushima M (2007) Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization. Math Program (to appear)

Luo ZQ, Pang JS, Ralph D (1996) Mathematical programs with equilibrium constraints. Cambridge University Press, Cambridge

Niederreiter H (1992) Random number generation and Quasi-Monte Carlo methods. SIAM, Philadelphia

Ortega JM, Rheinboldt WC (1970) Iterative solution of nonlinear equations in several variables. Academic Press, New York

Scheel H, Scholtes S (2000) Mathematical programs with complementarity constraints: stationarity, optimality, and sensivity. Math Oper Res 25:1–22

Shapiro A (2003) Monte Carlo sampling approach to stochastic programming. Eur Ser Appl Ind Math (ESAIM): Proc 13:65–73

Shapiro A (2006) Stochastic mathematical programs with equilibrium constraints. J Optim Theory Appl 128:223–243

Shapiro A, Xu H (2005) Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation. Preprint, School of Industrial and System Engineering, Georgia Institute of Technology, Atlanta

Xu H (2006) An implicit programming approach for a class of stochastic mathematical programs with linear complementarity constraints. SIAM J Optim 16:670–696