Working Paper Series

28/2007

Stochastic approximation approaches to the stochastic variational inequality problem

Jiang, H. and Xu, H.

CAMBRIDGE
Judge Business School

These papers are produced by Judge Business School, University of Cambridge. They are circulated for discussion purposes only. Their contents should be considered preliminary and are not to be quoted without the authors' permission.

Author contact details are as follows:

Houyuan Jiang
Judge Business School
University of Cambridge
h.jiang@jbs.cam.ac.uk

Huifu Xu
School of Mathematics
University of Southampton
h.xu@soton.ac.uk

Please address enquiries about the series to:

Research Support Manager
Judge Business School
Trumpington Street
Cambridge CB2 1AG, UK
Tel: 01223 760546 Fax: 01223 339701
E-mail: research-support@jbs.cam.ac.uk

Stochastic Approximation Approaches to the Stochastic Variational Inequality Problem

Houyuan Jiang*and Huifu Xu†

February 2007, revised September 2007

Abstract. Stochastic approximation methods have been extensively studied in the literature for solving systems of stochastic equations and stochastic optimization problems where function values and first order derivatives are not observable but can be approximated through simulation. In this paper, we investigate stochastic approximation methods for solving stochastic variational inequality problems (SVIP) where the underlying functions are the expected value of stochastic functions. Two types of methods are proposed: stochastic approximation methods based on projections and stochastic approximation methods based on reformulations of SVIP. Global convergence results of the proposed methods are obtained under appropriate conditions.

Keywords: Stochastic variational inequalities, stochastic complementarity problems, stochastic approximation, projection method, simulation.

1 Introduction

Consider the stochastic variational inequality problem (SVIP): Finding $x \in \mathbb{R}^n$ satisfying

$$(y - x)^T \mathbb{E}\left[f(x, \xi(\theta))\right] \geq 0, \quad \forall y \in \mathcal{Y}, \tag{1.1}$$

where $\xi(\theta)$ is a random variate defined on a probability space (Υ, Λ, P), $f(\cdot, \xi) : \mathbb{R}^n \to \mathbb{R}^n$ is continuous for every realization of $\xi(\theta)$, $\mathbb{E}\left[f(x, \xi(\theta))\right]$ is the expected value of $f(x, \xi(\theta))$ over $\xi(\theta)$, and $\mathcal{Y} \subseteq \mathbb{R}^n$ is a closed convex set.

SVIP has been investigated in [15, 16, 33, 35] and it is a natural extension of deterministic VIP (VIP for short). Over the past several decades, VIP has been effectively applied to modeling a range of equilibrium problems in engineering, economics, game theory, and networks; see books

*Correspondence author. Judge Business School, University of Cambridge, Trumpington Street, Cambridge CB2 1AG, UK (Email: h.jiang@jbs.cam.ac.uk Fax: +44 1223 339701).

†School of Mathematics, University of Southampton, Highfield, Southampton SO17 1BJ, UK (Email: H.Xu@soton.ac.uk Fax +44 23 80595147).

[8, 22]. While many practical problems only involve deterministic data, there are some important instances where problem data contain some uncertain factors and consequently SVIP models are needed to reflect uncertainties. For example, in an oligopoly competition of a future market, market demand is uncertain and firms have to choose their strategies to maximize their expected profits. In structural engineering, design of a structure or an object may involve random factors such as temperature and extraneous forces. Applications of SVIP can also be found in inventory or pricing competition among several firms that provide substitutable goods or services [4, 23]. Some stochastic dynamic games [1, 9] can be formulated as examples of SVIP. In Section 6, we present several detailed examples of these applications.

SVIP is closely related to some other interesting stochastic problems studied in the literature. When $\mathcal{Y} = \mathbb{R}^n_+$, SVIP reduces to the following stochastic nonlinear complementarity problem (SNCP): Finding $x \in \mathbb{R}^n$ satisfying

$$0 \leq x \perp \mathbb{E}\left[f(x, \xi(\theta))\right] \geq 0, \tag{1.2}$$

where $x \perp y$ means that $x^T y = 0$ for $x, y \in \mathbb{R}^n$. When $\mathcal{Y} = \mathbb{R}^n$, SVIP further reduces to a system of stochastic equations (SSE): Finding $x \in \mathbb{R}^n$ satisfying

$$\mathbb{E}\left[f(x, \xi)\right] = 0. \tag{1.3}$$

Note that SVIP is also related to the following smooth stochastic optimization problem:

$$\begin{array}{ll} \min & G(x) \equiv \mathbb{E}[g(x, \xi(\theta))], \\ \text{s.t.} & x \in \mathcal{Y}, \end{array} \tag{1.4}$$

where $g(\cdot, \xi) : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable for every realization of $\xi(\theta)$. The first order necessary conditions of the stochastic optimization problem gives rise to a symmetric SVIP in the sense that the Jacobian of G is symmetric.

Many numerical methods have been proposed for VIP but few can be applied directly to solving SVIP because of the complexity of $\mathbb{E}[f(x, \xi(\theta))]$. To explain this, let $F(x) = \mathbb{E}\left[f(x, \xi)\right]$. Then SVIP (1.1) can be rewritten as

$$(y - x)^T F(x) \geq 0, \quad \forall y \in \mathcal{Y}. \tag{1.5}$$

If $F(x)$ can be evaluated either analytically or numerically, then (1.5) can be regarded as a VIP and consequently it can be solved by existing numerical methods, which are documented in [8]. However, it might not be easy to evaluate $\mathbb{E}\left[f(x, \xi)\right]$ in the following situations: (a) ξ is a random vector with a known probability distribution, but calculations of the expected value involve multi-dimensional integration, which is computationally expensive if not impossible; (b) the function $f(x, \xi)$ is known, but the distribution of ξ is unknown and the information on ξ can only be obtained using past data or sampling; (c) $\mathbb{E}\left[f(x, \xi)\right]$ is not observable and it must be approximately evaluated through simulation. Under these circumstances, existing numerical methods for VIP are not applicable to SVIP and new methods are needed.

In this paper, we study the stochastic approximation (SA) method for solving SVIP and SNCP. Since its introduction by Robbins and Monro [31], SA has been extensively studied and applied to solving various practical stochastic problems arising in engineering and economics despite its slow convergence; see books [2, 7, 20, 27, 31, 37], sample references [18, 30, 36, 39], and an excellent survey paper by Fu [11] for motivations, justifications and applications. SA is based on a recursive procedure. For SSE (1.3), SA generates the next iterate x^{k+1} from the current iterate x^k by

$$x^{k+1} = x^k - a_k f(x^k, \xi^k),$$

where $a_k \geq 0$ is a pre-fixed step-length and ξ^k is drawn from ξ stochastically and independently. Under suitable conditions on f, samples of ξ and step-length a_k, SA is proved to converge almost surely to the solution of SSE (1.3); see [27].

Recently SA has been used to solve SVIP and SNCP which are reformulated from competitive revenue management problems (Mahajan and van Ryzin [23]) and stochastic game theoretic problems (Flam [10]). SA is also a key computational tool for solving other revenue management models in [3, 38]. Given SA's historical popularity and its emerging applications in engineering and economics, we feel it is necessary to systematically present a theoretical treatment of SA for SVIP and SNCP.

The main contributions of this paper can be summarized as follows: (a) we propose two SA methods based on projection for solving SVIP and prove global convergence results with probability one under fairly standard conditions (Theorems 3.1 and 3.2); (b) we reformulate SVIP as optimization problems using gap and D-gap functions and apply the existing SA methods for the stochastic optimization problem to the reformulated problems; convergence results (Theorem 4.1) are obtained for the case when the underlying function in SVIP is affine with respect to deterministic variables; furthermore we show how derivative-free iterative schemes for VIP can be extended to solve SVIP (Theorem 4.2); (c) we propose a derivative-free SA method based on the Fischer-Burmeister function for solving general SNCP and obtain the global convergence result without the uniform strong monotonicity condition on the underlying function (Theorem 5.1); we also present a counter example to show that the underlying function of nonsmooth equations reformulated from strongly monotone SNCP with Fischer-Burmeister function may not necessarily retain strong monotonicity, which is a key condition required to ensure convergence of the SA method for solving the nonsmooth equation based reformulation of SNCP.

In the literature, several other numerical approaches have been proposed for solving SVIP and SNCP. Among others, the sample path optimization (SPO) method and the sample average approach (SAA) have been well recognized. SPO is a simulation based approach considered by Plambeck, Fu, Robinson and Suri [29] and analyzed rigorously by Robinson [33]. The basic idea of the SPO method is to construct a sequence of computable functions $\{F_k\}$ which converges almost surely to an uncomputable original function F as k increases. Gürkan, Özge and Robinson [15, 16] consider an SVIP model where the expectation or limit function F is dynamically estimated by F_k by simulation. At each iteration k, an instance of VIP is obtained based on averaging effects through observing a large number of instances of the random parameters. Instances of VIP are solved by the PATH solver in which the automatic differentiation solver is used to estimate gradients of F_k. They discuss conditions under which the approximating problems can be shown to have solutions with probability 1 and provide bounds for the closeness of those solutions to solutions of the limit problem.

Note that SPO is closely related to SAA in which $F_k(x)$ is constructed by the sample average of $f(x, \xi)$ as follows

$$F_k(x) = \frac{1}{N} \sum_{i=1}^{N} f(x, \xi_i),$$

where ξ_i, $i = 1, \cdots, N$, is independently identically distributed samples of random variate ξ. Over the past few years, SAA has been increasingly studied for the stochastic optimization problem; see Shapiro in [35] for a comprehensive review of SAA. Note that the exponential convergence of SAA for SVIP and SNCP can be obtained under some mild conditions.

SA is an alternative to SPO, SAA and other existing stochastic methods. On the one hand, SPO and SAA are powerful computational tools typically when the underlying function $f(x, \xi)$ has a smooth and closed form. On the other hand, SA is more suitable for solving problems where the underlying function $f(x, \xi)$ is nonsmooth and/or has no closed form, that is, there is no off-the-shelf solver for the deterministic subproblem.

Note that the SNCP model (1.2) is different from the stochastic complementarity models recently considered in [5, 17, 21]. In the latter, a deterministic decision vector is sought to satisfy NCPs parameterized by all possible realizations of a random variate. This results in a deterministic overdetermined system of NCPs which usually do not have a solution. Chen and Fukushima [5] use NCP functions to reformulate NCPs into systems of nonsmooth equations and consider least-squared minimization of the residual of the reformulated equations. Consequently it can be proved that solutions for such a reformulated problem exist under suitable conditions.

The rest of the paper is organized as follows. In the next section, we present some results related to the projection operator after introducing some basic definitions. In Section 3, two classical projection based methods for the deterministic VIP are extended for solving SVIP. Under appropriate conditions, global convergence of those iterative schemes are established. In Section 4, we propose more SA methods that are based on reformulations of SVIP into the stochastic optimization problem or SSE. Global convergence of some of those iterative schemes are established too. In Section 5, we develop numerical methods specifically for solving SNCP. In Section 6, we collect several practical problems to illustrate how SVIP and SNCP can be used as appropriate mathematical models. We make some concluding remarks in Section 7.

2 Definitions and preliminaries

In this section we introduce some necessary definitions related to VIP and present some preliminary results about the projection operator on a convex set in the context of VIP.

Definition 2.1 ([8]) *Let $\mathcal{Y}$ be a convex subset of $\mathbb{R}^n$ and $D \in \mathbb{R}^{n\times n}$ a symmetric positive definite matrix. The projection operator $\Pi_{\mathcal{Y},D} : \mathbb{R}^n \to \mathcal{Y}$ is called the* skewed projection mapping *onto $\mathcal{Y}$ if for every fixed $x \in \mathbb{R}^n$, $\Pi_{\mathcal{Y},D}(x)$ is the solution of the following convex optimization problem:*

$$\begin{array}{rl} \min_y & \frac{1}{2}\|y-x\|_D^2 \equiv \frac{1}{2}(y-x)^T D(y-x), \\ \text{s.t.} & y \in \mathcal{Y}, \end{array}$$

where $\|s\|_D = \sqrt{s^T Ds}$ is the D-norm of $s \in \mathbb{R}^n$.

It is known [8] that for any $s \in \mathbb{R}^n$ and any symmetric positive definite matrix $D \in \mathbb{R}^{n\times n}$,

$$\lambda_{\min}(D)\|s\|^2 \le \|s\|_D^2 \le \lambda_{\max}(D)\|s\|^2, \tag{2.6}$$

where $\lambda_{\min}(D) > 0$ and $\lambda_{\max}(D) > 0$ are the smallest and the largest eigenvalues of D respectively. Here $\|\cdot\|$ denotes the standard 2-norm.

Definition 2.2 ([2, 8, 14]) *Let the function F be a mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$ and $\mathcal{Y}$ a subset of $\mathbb{R}^n$. F is said to be* strongly monotone *on $\mathcal{Y}$ with modulus $\sigma > 0$ if $(F(x) - F(y))^T(x-y) \ge \sigma\|x-y\|^2$ for all $x, y \in \mathcal{Y}$; F is said to be* strictly monotone *on $\mathcal{Y}$ if $(F(x)-F(y))^T(x-y) > 0$, for all $x \ne y \in \mathcal{Y}$; F is said to be* monotone *on $\mathcal{Y}$ if $(F(x)-F(y))^T(x-y) \ge 0$, for all $x, y \in \mathcal{Y}$; F is said to be* inversely strongly monotone (ISM) *on $\mathcal{Y}$ under the D-norm with modulus $\mu > 0$ if $(F(x)-F(y))^T(x-y) \ge \mu\|F(x)-F(y)\|_D^2$, for all $x, y \in \mathcal{Y}$, where D is a symmetric positive definite matrix in $\mathbb{R}^{n\times n}$. When D is the identity matrix, ISM is also called co-coercive in [8]. F is said to be* Lipschitz continuous *on $\mathcal{Y}$ with modulus $L > 0$ if $\|F(x)-F(y)\| \le L\|x-y\|$, for all $x, y \in \mathcal{Y}$.*

Remark 2.1 *(a) If F is ISM on $\mathcal{Y}$, then F is both Lipschitz and monotone on $\mathcal{Y}$ (not necessarily strongly monotone; see [14] for a counter example). If F is both strongly monotone and Lipschitz continuous on $\mathcal{Y}$, then F is ISM on $\mathcal{Y}$; see [8, Page 164]. (b) The properties of F described in the definition may be obtained by the corresponding properties of $f(x,\xi)$ with respect to x for almost all ξ. For instance, if $f(x,\xi)$ is uniformly strongly monotone with respect to x, that is, there exists $\sigma > 0$ such that for almost every realization ξ of $\xi(\theta)$, and*

$$(f(y,\xi) - f(x,\xi))^T(y-x) \ge \sigma\|y-x\|^2, \forall \xi, x, y \in \mathcal{Y},$$

then $\mathbb{E}[f(x,\xi)]$ is strongly monotone.

Proposition 2.1 below summarizes some main properties of the projection mapping, which will be used in the proofs of our main convergence results in 3 and 4. A proof for Proposition 2.1 can be found in the appendix.

Proposition 2.1 *Let D be a symmetric positive definite matrix in $\mathbb{R}^{n\times n}$, $\mathcal{Y}$ a closed convex subset of $\mathbb{R}^n$ and $\Pi_{\mathcal{Y},D} : \mathbb{R}^n \to \mathcal{Y}$ the skewed projection mapping as defined in Definition 2.1 . Then*

(a) $\Pi_{\mathcal{Y},D}$ *is nonexpansive under the D-norm, i.e.,*

$$\|\Pi_{\mathcal{Y},D}(x) - \Pi_{\mathcal{Y},D}(y)\|_D \le \|x - y\|_D, \ \forall x, y \in \mathcal{Y}.$$

(b) The projection mapping $\Pi_{\mathcal{Y},D}$ *is ISM under the D-norm with modulus* 1.

(c) x^* *is a solution of the (1.1) if and only if the following holds*

$$\Pi_{\mathcal{Y},D}[x^* - aD^{-1}F(x^*)] = x^*,$$

where a is a positive constant.

(d) For $0 \le a \le 4\mu\lambda_{\max}(D)$, the mapping $x - \Pi_{\mathcal{Y},D}(x - aD^{-1}F(x))$ is ISM on $\mathcal{Y}$ with modulus $1 - \frac{a}{4\mu\lambda_{\min}(D)}$ if F is ISM on $\mathcal{Y}$ with modulus μ.

Remark 2.2 *When D is the identity matrix, (b) and (d) of Proposition 2.1 are proved in [14].*

3 Stochastic Approximation Methods Based on Projections

We consider the following Robbins-Monro type stochastic approximation (iterative) scheme [31] for solving the SVIP (1.5):

$$x^{k+1} = \Pi_{\mathcal{Y},D}[x^k - a_k(F(x^k) + \omega^k + R^k)], \tag{3.7}$$

where $F(x^k)$ is the true value of F at x^k and $F(x^k) + \omega^k + R^k$ is an "approximation" of F at x^k, ω^k is a stochastic error and R^k is a deterministic system error. For ease of exposition, we assume throughout the rest of this section that $R^k \equiv 0$ which means there is no deterministic error in calculation.

To explain how iterative scheme (3.7) works, let us consider a special case when $F(x^k)+\omega^k = f(x^k, \xi^k)$ where ξ^k is a particular sample of random variate $\xi(\theta)$. In other words, at iteration k we simply use a sample ξ^k of ξ to calculate $f(x^k, \xi)$ and regard it as an approximation of $\mathbb{E}[f(x^k, \xi)] \equiv F(x^k)$. Obviously in this case, we do not need to know the probability distribution of ξ for approximating $F(x^k)$.

In what follows, we analyze convergence of the sequence generated by (3.7). Let $\mathcal{F}_k$ denote an increasing sequence of σ-algebras such that x^k is $\mathcal{F}_k$ measurable. We need to make the following assumptions.

Assumption 3.1 *(a) The stepsize a_k satisfies $a_k > 0$, $a_k \to 0$, $\sum_{k=0}^{\infty} a_k = \infty$, and $\sum_{k=0}^{\infty}(a_k)^2 < \infty$; (b) $\mathbb{E}[\omega^k|\mathcal{F}_k] = 0$; (c) $\sum_{k=1}^{\infty}(a_k)^2\mathbb{E}[\|w^k\|^2|\mathcal{F}_k] < \infty$ holds almost surely. (d) F is globally Lipschitz with modulus L over $\mathcal{Y}$; (e) F is strongly monotone with modulus σ over $\mathcal{Y}$.*

A few comments about Assumption 3.1 are in order. Part (a) is a standard rule for stepsize choices in SA. See [27] for instance. Parts (b) and (c) state the stochastic error ω^k is unbiased and the scale of variance is under control. These assumptions are also standard in the literature of SA methods. Parts (d) and (e) are specifically made for SVIP.

Recall that a sequence of random variables $\{X^k\}$ converges almost surely to random variable X if $\mathrm{P}(\lim_{n\to\infty} X^k = X) = 1$. Our first result on SA uses the lemma below which is a generalization of the martingale convergence theorem.

Lemma 3.1 ([32]) *Let $\{\mathcal{F}_k\}$ be an increasing sequence of σ-algebras and V_k, α_k, β_k and γ_k be nonnegative random variables adapted to $\mathcal{F}_k$. If it holds almost surely that $\sum_{k=1}^{\infty}\alpha_k < \infty$, $\sum_{k=1}^{\infty}\beta_k < \infty$ and*

$$\mathbb{E}(V_{k+1}|\mathcal{F}_k) \le (1+\alpha_k)V_k - \gamma_k + \beta_k,$$

then $\{V_k\}$ is convergent almost surely and $\sum_{k=1}^{\infty}\gamma_k < \infty$ almost surely.

Theorem 3.1 *Let $\{x^k\}$ be generated by iterative scheme (3.7). Suppose that Assumption 3.1 (b), (c), (d) and (e) are satisfied for this scheme and the following conditions on stepsize hold*

$$\inf_{k\ge 0} a_k > 0, \qquad \sup_{k\ge 0} a_k \le \frac{2\sigma}{L^2\lambda_{\max}(D^{-1})}. \tag{3.8}$$

Then $\{x^k\}$ converges to the unique solution x^* of SVIP (1.1) almost surely.

Proof. The existence and uniqueness of a solution for SVIP (1.1) is guaranteed by the strong monotonicity property of F under Assumption 3.1 (e). We next prove convergence. By iterative scheme (3.7) and Proposition 2.1(c), we obtain

$$\begin{aligned}
& \mathbb{E}[\|x^{k+1} - x^*\|_D^2 | \mathcal{F}_k] \\
= \ & \mathbb{E}[\|\Pi_{\mathcal{Y},D}[x^k - a_k D^{-1}(F(x^k) + w^k)] - \Pi_{\mathcal{Y},D}(x^* - a_k D^{-1} F(x^*))\|_D^2 | \mathcal{F}_k] \\
\leq \ & \mathbb{E}[\|x^k - x^* - a_k D^{-1}(F(x^k) - F(x^*) + w^k)\|_D^2 | \mathcal{F}_k] \\
& \text{(By Proposition 2.1(a))} \\
= \ & \|x^k - x^*\|_D^2 + (a_k)^2 (F(x^k) - F(x^*))^T D^{-1} (F(x^k) - F(x^*)) + (a_k)^2 \mathbb{E}[(w^k)^T D^{-1} w^k | \mathcal{F}_k] \\
& -2a_k (x^k - x^*)^T (F(x^k) - F(x^*)) - 2a_k (x^k - x^*)^T \mathbb{E}[w^k | \mathcal{F}_k] \\
& +(a_k)^2 (F(x^k) - F(x^*))^T D^{-1} \mathbb{E}[w^k | \mathcal{F}_k] \\
= \ & \|x^k - x^*\|_D^2 + (a_k)^2 (F(x^k) - F(x^*))^T D^{-1} (F(x^k) - F(x^*)) + (a_k)^2 \mathbb{E}[(w^k)^T D^{-1} w^k | \mathcal{F}_k] \\
& -2a_k (x^k - x^*)^T (F(x^k) - F(x^*)) \\
& \text{(By Assumption 3.1 (b))} \\
\leq \ & \|x^k - x^*\|_D^2 + (a_k)^2 L^2 \tfrac{\lambda_{\max}(D^{-1})}{\lambda_{\min}(D)} \|x^k - x^*\|_D^2 + (a_k)^2 \mathbb{E}[\lambda_{\max}(D^{-1}) \|w^k\|^2 | \mathcal{F}_k] \\
& -2a_k \tfrac{\sigma}{\lambda_{\max}(D)} \|x^k - x^*\|_D^2 \\
= \ & \|x^k - x^*\|_D^2 (1 - 2\delta_k) + \beta_k \\
& \text{(By (2.6), (3.8), Assumption 3.1 (d) and (e))} \\
\leq \ & \|x^k - x^*\|_D^2 (1 + 0) - \delta_k \|x^k - x^*\|_D^2 + \lambda_{\max}(D^{-1})(a_k)^2 \mathbb{E}[\|w^k\|^2 | \mathcal{F}_k] \\
= \ & \|x^k - x^*\|_D^2 (1 + \alpha_k) - \gamma_k + \beta_k,
\end{aligned}$$

where $\delta_k = 2a_k \frac{\sigma}{\lambda_{\max}(D)} - (a_k)^2 L^2 \frac{\lambda_{\max}(D^{-1})}{\lambda_{\min}(D)}$, $\alpha_k \equiv 0 \geq 0$, $\beta_k = \lambda_{\max}(D^{-1})(a_k)^2 \mathbb{E}[\|w^k\|^2 | \mathcal{F}_k] \geq 0$ and $\gamma_k = \delta_k \|x^k - x^*\|_D^2 \geq 0$. Under Assumptions 3.1(c), we have $\sum_{k=1}^{\infty} \alpha_k < \infty$ and $\sum_{k=1}^{\infty} \beta_k < \infty$ almost surely. It follows from Lemma 3.1 that $\|x^k - x^*\|_D$ converges almost surely. and $\sum_{k=1}^{\infty} \gamma_k < \infty$ holds almost surely. By condition (3.8), $\{\delta_k\}$ is bounded away from zero which, implies that $\{x^k\}$ converges to x^* almost surely because condition (3.8) shows that $\{\delta_k\}$ is bounded away from zero. ■

Theorem 3.1 is an extension of Theorem 12.1.8 of [8], which is a projection method for solving VIP. Conditions (d) and (e) in Assumption 3.1 are quite strong when they are put together. In what follows, we replace these two conditions with the following two weaker conditions.

- (d') There exists a solution x^* of SVIP such that for any $x \in \mathcal{Y}$, $\|F(x)\| \leq \lambda(1 + \|x - x^*\|)$ for a constant $\lambda > 0$.
- (e') At the solution x^*, $\inf_{x \in \mathcal{Y}: \rho \geq \|x - x^*\| \geq \varepsilon} F(x)^T (x - x^*) > 0$, for any $\rho > \varepsilon > 0$.

The example below shows that condition (e') is strictly weaker than condition (e).

Example 3.1 *Consider SVIP (1.5) with single variable where*

$$F(x) = \begin{cases} \frac{1}{2}(x+1), & x < 1, \\ \sqrt{x}, & x \geq 1, \end{cases} \tag{3.9}$$

and $\mathcal{Y} = \mathbb{R}_+$. Then F is strictly monotone but not strongly monotone. Therefore Assumption 3.1 (e) is not satisfied in this case. However it is not difficult to verify that condition (e$'$) holds. Too see this, notice that the problem has a unique solution $x^* = 0$. Condition (e$'$) requires that

$$\inf_{x\in\mathbb{R}_+:\rho\geq\|x-x^*\|\geq\varepsilon} F(x)^T x > 0.$$

for any $\rho > \varepsilon > 0$. By the definition of F, we have

$$\inf_{x\in\mathbb{R}_+:\rho\geq\|x-x^*\|\geq\varepsilon} F(x)^T x \geq \begin{cases} \frac{1}{2}\varepsilon(\varepsilon+1), & \varepsilon < 1, \\ 1, & \varepsilon \geq 1. \end{cases}$$

Note that this function is globally Lipschitz continuous with modulus 0.5 *(the maximum of the absolute value of its derivative).*

Despite it is weaker than Assumption 3.1(e), condition (e') is a bit difficult to verify. We discuss sufficient conditions for (e') in Proposition 3.1 for which a proof can be found in appedix.

Proposition 3.1 *Suppose that F is monotone on $\mathcal{Y}$. Then*

(a) (e$'$) holds either F is strictly monotone at x^* or $-F(x^*)$ is in the interior of the polar cone of the tangent cone of $\mathcal{Y}$ at x^*;

(b) condition (e$'$) implies that SVIP (1.1) has a unique solution.

Theorem 3.2 *Let sequence $\{x^k\}$ be generated by iterative scheme (3.7). Suppose that Assumption 3.1 holds with (d) and (e) being replaced by (d$'$) and (e$'$) for this scheme. Suppose also that F is monotone at x^*. Then $\{x^k\}$ almost surely converges to the unique solution x^*.*

Proof. First, Proposition 3.1, monotonicity of F and condition (e') imply the uniqueness of the solution. We next prove convergence. By iterative scheme (3.7) and the fact that x^* is a solution of SVIP, we have

$$
\begin{aligned}
& \mathbb{E}[\|x^{k+1} - x^*\|_D^2 | \mathcal{F}_k] \\
= \;& \mathbb{E}[\|\Pi_{\mathcal{Y},D}[x^k - a_k D^{-1}(F(x^k) + w^k)] - \Pi_{\mathcal{Y},D}(x^*)\|_D^2 | \mathcal{F}_k] \\
\leq \;& \mathbb{E}[\|x^k - x^* - a_k D^{-1}(F(x^k) + w^k)\|_D^2 | \mathcal{F}_k] \\
& \text{ (By Proposition 2.1(a))} \\
= \;& \|x^k - x^*\|_D^2 + (a_k)^2 F(x^k)^T D^{-1} F(x^k) + (a_k)^2 \mathbb{E}[(w^k)^T D^{-1} w^k | \mathcal{F}_k] \\
& -2a_k (x^k - x^*)^T F(x^k) - 2a_k (x^k - x^*)^T \mathbb{E}[w^k | \mathcal{F}_k] + 2(a_k)^2 F(x^k)^T D^{-1} \mathbb{E}[w^k | \mathcal{F}_k] \\
= \;& \|x^k - x^*\|_D^2 + (a_k)^2 F(x^k)^T D^{-1} F(x^k) + (a_k)^2 \mathbb{E}[(w^k)^T D^{-1} w^k | \mathcal{F}_k] \\
& -2a_k (x^k - x^*)^T F(x^k) \\
\leq \;& \|x^k - x^*\|_D^2 + (a_k)^2 \lambda_{\max}(D^{-1}) \|F(x^k)\|^2 + (a_k)^2 \lambda_{\max}(D^{-1}) \mathbb{E}[\|w^k\|^2 | \mathcal{F}_k] \\
& -2a_k (x^k - x^*)^T F(x^k) \\
\leq \;& \|x^k - x^*\|_D^2 + 3(a_k)^2 \lambda_{\max}(D^{-1}) \lambda^2 (1 + \|x^k - x^*\|^2) \\
& +(a_k)^2 \lambda_{\max}(D^{-1}) \mathbb{E}[\|w^k\|^2 | \mathcal{F}_k] - 2a_k (x^k - x^*)^T F(x^k) \quad \text{(By Assumption 3.1 (d'))} \\
\leq \;& \|x^k - x^*\|_D^2 + 3(a_k)^2 \lambda_{\max}(D^{-1}) \lambda^2 \left(1 + \frac{1}{\lambda_{\min}(D)} \|x^k - x^*\|_D^2\right) \\
& +(a_k)^2 \lambda_{\max}(D^{-1}) \mathbb{E}[\|w^k\|^2 | \mathcal{F}_k] - 2a_k (x^k - x^*)^T F(x^k) \\
\equiv \;& \|x^k - x^*\|_D^2 (1 + \alpha_k) - \gamma_k + \beta_k,
\end{aligned}
$$

where

$$\alpha_k = \frac{3\lambda_{\max}(D^{-1})\lambda^2}{\lambda_{\min}(D)} (a_k)^2 \geq 0,$$

$$\beta_k = \left(3\lambda_{\max}(D^{-1})\lambda^2 + \lambda_{\max}(D^{-1}) \mathbb{E}[\|w^k\|^2 | \mathcal{F}_k]\right) (a_k)^2 \geq 0$$

and

$$\gamma_k = 2a_k (x^k - x^*)^T F(x^k) = 2a_k [(x^k - x^*)^T (F(x^k) - F(x^*)) + (x^k - x^*)^T F(x^*)] \geq 0.$$

Under Assumption 3.1 (a), (b) and (c), $\sum_{k=1}^{\infty} \alpha_k < \infty$ and almost surely $\sum_{k=1}^{\infty} \beta_k < \infty$. It follows from Lemma 3.1 that $\|x^k - x^*\|$ converges almost surely and

$$\sum_{k=1}^{\infty} a_k (x^k - x^*)^T F(x^k) < \infty \tag{3.10}$$

almost surely. Suppose $\{x^k\}$ does not converge to x^* almost surely. Then there exist a constant $\varepsilon > 0$ and an index ℓ such that $\|x^k - x^*\| \geq \varepsilon$ holds almost surely for all $k \geq \ell$. By Condition (d'), we have that $F(x^k)^T (x^k - x^*) > \delta > 0$ for any $k \geq \ell$ with a positive constant δ. This shows that

$$\sum_{k=\ell}^{\infty} a_k F(x^k)^T (x^k - x^*) \geq \sum_{k=\ell}^{\infty} a_k \delta = \delta \sum_{k=\ell}^{\infty} a_k = \infty,$$

which contradicts to (3.10). Therefore $\{x^k\}$ converges to x^*. ■

Remark 3.1 *Theorems 3.1 and 3.2 address convergence of the same iterative scheme (3.7) under different and non-overlapping assumptions. Theorem 3.1 assumes that (3.8) holds which* *implies the stepsize is bounded away from zero as $k \to \infty$ while Theorem 3.2 replaces this condition with Assumption 3.1(a) which requires stepsize go to zero as $k \to \infty$. This is the exact reason why we can weaken Assumption 3.1(c) and (d) in Theorem 3.1 to Assumption 3.1(c') and (d') in Theorem 3.2.*

4 Stochastic Approximation Methods Based on Reformulations

Apart from projection methods, many numerical methods have been developed for solving deterministic VIP [8]. By introducing suitable merit functions, one can reformulate VIP into equivalent smooth constrained or unconstrained optimization problems for which many efficient methods are readily available. Our purpose in this section is to show some of the above methods can be extended for solving SVIP. We prefer unconstrained optimization reformulations to constrained ones because the latter involves two projections: one due to reformulation of SVIP as a constrained optimization problem and another due to the application of SA for the reformulated optimization problems.

Consider the traditional SA method for solving the stochastic optimization problem (1.4) with $\mathcal{Y} = \mathbb{R}^n$:

$$x^{k+1} = x^k - a_k(\nabla G(x^k) + \omega^k), \tag{4.11}$$

where a_k is stepsize and ω^k is the stochastic approximation error of ∇G at x^k. Observe that finding stationary points of (1.4) is equivalent to finding solutions of (1.3) with $F(x) = \nabla G(x)$. This means that the above iterative scheme (4.11) is a special case of iterative scheme (3.7). Therefore, it is not a surprise to obtain almost sure convergence results for iterative scheme (4.11). As a matter of fact, such a convergence result can be found from many SA text books; see for instance [27, Chapter 5]. The usual conditions that ensure such a convergence result are strong convexity of G (or equivalently strong monotonicity of ∇G) and the Lipschitz continuous property of ∇G, but can be replaced by their weaker counterparts Assumption 3.1 (d') and (e') respectively. We shall investigate those conditions for some popular reformulations of VIP.

Regularized gap and D-gap functions proposed in [12, 41] are those merit functions that have desirable properties for designing various efficient numerical methods for the deterministic VIP. For any positive scalars α and β ($\beta < \alpha$), the regularized gap function is defined by

$$\Phi_\alpha(x) = \max_{y \in \mathcal{Y}} F(x)^T(x - y) - \frac{\alpha}{2}\|y - x\|^2,$$

and the D-gap function by

$$\Psi_{\alpha\beta} = \Phi_\alpha(x) - \Phi_\beta(x).$$

It is known [12] that

$$\Phi_\alpha(x) = F(x)^T(x - \Pi_{\mathcal{Y}}[x - \alpha^{-1}F(x)]) - \frac{\alpha}{2}\|\Pi_{\mathcal{Y}}[x - \alpha^{-1}F(x)) - x\|^2,$$

i.e., $\Pi_{\mathcal{Y}}[x - \alpha^{-1}F(x)]$ is the unique optimal solution of the maximization problem that defines $\Phi_\alpha(x)$. Note that the original regularized gap function is defined in a more general setting that allows the norm $\|\cdot\|$ used in $\Phi_\alpha(x)$ to be replaced by $\|\cdot\|_D$ with D a symmetric positive definite matrix.

Based on either the regularized gap function or the D-gap function, VIP can be cast as either a constrained optimization problem

$$\begin{array}{rl} \min & \Phi_\alpha(x) \\ \text{s.t.} & x \in \mathcal{Y} \end{array} \tag{4.12}$$

or an unconstrained optimization problem

$$\begin{array}{rl} \min & \Psi_{\alpha\beta}(x) \\ \text{s.t.} & x \in \mathbb{R}^n \end{array} \tag{4.13}$$

in the sense that any global solution of the reformulated optimization problem is a solution of VIP, and vice versa. When F is continuously differentiable, both the regularized gap function and the D-gap function are proved to be continuously differentiable [12, 41], but not necessarily twice continuously differentiable in general. Moreover,

$$\nabla\Phi_\alpha(x) = F(x) - (\nabla F(x) - \alpha I)(\Pi_{\mathcal{Y}}[x - \alpha^{-1}F(x)] - x).$$

When F is further assumed to be strongly monotone over $\mathcal{Y}$, a stationary point of either (4.12) or (4.13) is the unique solution of the VIP [12, 41]. Those analytical properties pave a way for solving the VIP using numerical methods of smooth optimization problems. In the context of the SVIP, we take that $F(x) = \mathbb{E}[f(x,\xi)]$ as defined in Section 1. Then our aim is to find solutions of SVIP by solving its corresponding stochastic reformulations of (4.12) and (4.13).

Stochastic approximation methods for solving the stochastic optimization problem have been extensively investigated in the literature. Here we apply the SA method in [27] to (4.12) and (4.13). According to Theorem 5.3 in [27], convergence of the SA method in [27] relies on Assumption 3.1 (d') and (e'). When F is nonlinear, it is difficult to prove these properties for either $\Phi_\alpha(x)$ or $\Psi_{\alpha\beta}(x)$. In what follows we consider the case where F is affine.

Proposition 4.1 *Suppose that $F : \mathbb{R}^n \to \mathbb{R}^n$ is an affine mapping such that $F(x) = Ax + b$ and $A \in \mathbb{R}^{n\times n}$ is positive definite. Then (a) if $\beta > \alpha > 0$ are chosen such that $A + A^T - \alpha I - \beta^{-1}A^TA$ is positive definite, then $\Psi_{\alpha\beta}(x)$ is strongly convex; (b) if $\alpha > 0$ is chosen such that $A + A^T - \alpha I$ is positive definite then $\Phi_\alpha(x)$ is strongly convex; (c) both $\nabla\Phi_\alpha(x)$ and $\nabla\Psi_{\alpha\beta}(x)$ are globally Lipschitz continuous.*

Proof. (a) This is proved in Proposition 1 of [26]. (b) follows from part (a) by taking β to infinity in which case $\Psi_{\alpha\beta}(x)$ reduces to $\Phi_\alpha(x)$.

(c) This can be easily proved by checking $\nabla\Phi_\alpha(x)$ and $\nabla\Psi_{\alpha\beta}(x)$ and the fact that the projection operator is nonexpansive; see Proposition 2.1. ■

Theorem 4.1 *Suppose that $F : \mathbb{R}^n \to \mathbb{R}^n$ is an affine mapping such that $F(x) = Ax + b$ and $A \in \mathbb{R}^{n\times n}$ is positive definite.*

(a) Assume $\beta > \alpha > 0$ are chosen such that $A + A^T - \alpha I - \beta^{-1}A^T A$ is positive definite. Let $x^{k+1} = x^k - a_k(\nabla\Psi_{\alpha\beta}(x^k) + w^k)$. If Assumptions 3.1 (a), (b) and (c) hold for this scheme[1], then $\{x^k\}$ converges to the unique solution of SVIP almost surely.

(b) Assume $\alpha > 0$ is chosen such that $A + A^T - \alpha I$ is positive definite. Let $x^{k+1} = \Pi_{\mathcal{Y}}[x^k - a_k(\nabla\Phi_\alpha(x^k) + w^k)]$. If Assumptions 3.1 (a), (b) and (c) hold, for this scheme, then sequence $\{x^k\}$ generated converges to the unique solution of SVIP almost surely.

Proof. Proposition 4.1 implies Assumption 3.1 (d') and (e'), which are equivalent to Conditions (i) and (ii) of Theorem 5.3 of [27] respectively. Therefore the results in (a) and (b) follow from Theorem 5.3 of [27]. ■

Remark 4.1 *The computational efforts of the stochastic approximation methods used in (a) and (b) of Theorem 4.1 are comparable since the former needs to evaluate the projection of $x^k - a_k(\nabla\Phi_\alpha(x^k) + w^k)$ over $\mathcal{Y}$ while the latter applies projections twice when evaluating $\nabla\Psi_{\alpha\beta}(x^k)$.*

One of the challenges in solving SVIP is to evaluate or approximate $\Phi_\alpha(x)$ and $\Psi_{\alpha\beta}(x)$ both involving evaluation of $F(x)$. It is even more challenging to evaluate or approximate $\nabla\Phi_\alpha(x)$ and $\nabla\Psi_{\alpha\beta}(x)$ both involving evaluations of Jacobians of F. Therefore, it is computationally advantageous to adopt derivative free iterative schemes that have been proposed for solving VIP.

Based on the merit function $\Psi_{\alpha\beta}$, Yamashita *et. al.* [41] propose a derivative free iterative scheme for solving VIP with the search direction below:

$$d = r(x) + \rho s(x),$$

where

$$r(x) = y_\alpha(x) - y_\beta(x), \qquad s(x) = \alpha(x - y_\alpha(x)) - \beta(x - y_\beta(x)),$$

and $y_\alpha(x) = \Pi_{\mathcal{Y}}[x - \alpha^{-1}F(x)]$ is the unique solution of the maximization problem that defines $\Phi_\alpha(x)$; see the beginning of this section. The derivative-free search direction d is proved to be a descent direction of the merit function $\Psi_{\alpha\beta}$ and the proposed derivative-free iterative scheme is proved to converge to the unique solution of VIP when F is strongly monotone over $\mathbb{R}^n$.

To extend the above derivative-free iterative scheme to SVIP, we need the following result.

Lemma 4.1 ([41]) *Let $d = r(x) + \rho s(x)$. Suppose F is strongly monotone with modulus σ over $\mathbb{R}^n$.*

[1]This means that a_k, ω^k and $\mathcal{F}_k$ in the assumption refer to this scheme. The same comment applies to part (b) of this theorem, Theorem 4.2, Theorem 5.1 and Corollary 5.1.

(a) $\nabla \psi_{\alpha\beta}(x)^T d \leq -\frac{\sigma}{2}(\|r(x)\| + \rho\|s(x)\|)^2$.

(b) If x is not a solution of VIP, then

$$\nabla \Psi_{\alpha\beta}(x)^T d \leq -\frac{\sigma}{2}\|d\|^2 \tag{4.14}$$

for sufficiently small positive ρ.

(c) If $r(x) = s(x) = 0$, then x is a solution of VIP.

(d) The level set of $\psi_{\alpha\beta}(x)$ is bounded.

Proof. (a), (c) and (d) are proved in [41]. (b) follows (a) by the definition of d. ■

We are now ready to state the SA method that extends the above derivative-free iterative scheme of [41] based on the unconstrained reformulation (4.13) of SVIP: Given the current iterate x^k,

$$x^{k+1} = x^k - a_k(d^k + \omega^k), \tag{4.15}$$

where

$$d^k = r(x^k) + \rho s(x^k),$$

and ω^k represents the stochastic error when approximating d^k from sampling. In computational implementation, $d^k + \omega^k$ is replaced by a sample $d^k(\xi) = r(x^k, \xi) + \rho s(x^k, \xi)$, where

$$r(x, \xi) = y_\alpha(x, \xi) - y_\beta(x, \xi), \qquad s(x, \xi) = \alpha(x - y_\alpha(x, \xi)) - \beta(x - y_\beta(x, \xi)),$$

and $y_\alpha(x, \xi) = \Pi_{\mathcal{Y}}[x - \alpha^{-1} f(x, \xi)]$.

Theorem 4.2 *Suppose that Assumptions 3.1(a), (b), (c), (d) and (e) hold for iterative scheme (4.15), and $\nabla \Psi_{\alpha\beta}(x)$ is globally Lipschitz continuous. Then sequence $\{x^k\}$ generated by this scheme almost surely converges to the unique solution x^* of SVIP when $\rho > 0$ is sufficiently small and $\beta > \alpha > 0$.*

Proof. By virtue of the mean value theorem, there exists y^k located on the line segment between x^k and x^{k+1} such that

$$\Psi_{\alpha\beta}(x^{k+1}) = \Psi_{\alpha\beta}(x^k) + \nabla \Psi_{\alpha\beta}(y^k)^T (x^{k+1} - x^k).$$

Since $\nabla \Psi_{\alpha\beta}$ is globally Lipschitz continuous, there exists $L > 0$ such that

$$\|\nabla \Psi_{\alpha\beta}(y^k) - \nabla \Psi_{\alpha\beta}(x^k)\| \leq L\|x^{k+1} - x^k\|.$$

By iterative scheme (4.15) and the fact that x^* is a solution of SVIP, we have

$$\begin{aligned}
\mathbb{E}[\Psi_{\alpha\beta}(x^{k+1})|\mathcal{F}_k] &= \mathbb{E}[\Psi_{\alpha\beta}(x^k) + \nabla\Psi_{\alpha\beta}(y^k)^T(x^{k+1} - x^k)|\mathcal{F}_k] \\
&= \mathbb{E}[\Psi_{\alpha\beta}(x^k) + a_k\nabla\Psi_{\alpha\beta}(y^k)^T(d^k + w^k)|\mathcal{F}_k] \\
&= \Psi_{\alpha\beta}(x^k) + a_k\mathbb{E}[\nabla\Psi_{\alpha\beta}(y^k)(d^k + w^k)|\mathcal{F}_k] \\
&= \Psi_{\alpha\beta}(x^k) + a_k\mathbb{E}[\nabla\Psi_{\alpha\beta}(x^k)^T(d^k + w^k)|\mathcal{F}_k] \\
&\quad +(\nabla\Psi_{\alpha\beta}(y^k) - \nabla\Psi_{\alpha\beta}(x^k))^T(d^k + w^k)|\mathcal{F}_k] \\
&= \Psi_{\alpha\beta}(x^k) + a_k\nabla\Psi_{\alpha\beta}(x^k)^T d^k + 0 \\
&\quad +a_k\mathbb{E}[(\nabla\Psi_{\alpha\beta}(y^k) - \nabla\Psi_{\alpha\beta}(x^k))^T(d^k + w^k)|\mathcal{F}_k] \\
&\leq \Psi_{\alpha\beta}(x^k) + a_k\nabla\Psi_{\alpha\beta}(x^k)^T d^k + a_k\mathbb{E}[L\|y^k - x^k\|\|d^k + w^k\||\mathcal{F}_k] \\
&\leq \Psi_{\alpha\beta}(x^k) + a_k\nabla\Psi_{\alpha\beta}(x^k)^T d^k + L(a_k)^2\mathbb{E}[\|d^k + w^k\|^2] \\
&\leq \Psi_{\alpha\beta}(x^k) + a_k\nabla\Psi_{\alpha\beta}(x^k)^T d^k + 2L(a_k)^2(\|d^k\|^2 + \mathbb{E}[\|w^k\|^2|\mathcal{F}_k] \\
&\leq \Psi_{\alpha\beta}(x^k) - a_k\tfrac{\sigma}{2}\|d^k\|^2 + 2L(a_k)^2(\|d^k\|^2 + \mathbb{E}[\|w^k\|^2|\mathcal{F}_k] \\
&\quad \text{(By Lemma 4.1}(b)) \\
&\leq \Psi_{\alpha\beta}(x^k) + a_k(2La_k - \sigma)\|d^k\|^2 + 2L(a_k)^2\mathbb{E}[\|w^k\|^2|\mathcal{F}_k] \\
&= \Psi_{\alpha\beta}(x^k) - \gamma_k + \beta_k,
\end{aligned}$$

where

$$\gamma_k = -a_k(2La_k - \sigma)\|d^k\|^2,$$

and

$$\beta_k = 2L(a_k)^2\mathbb{E}[\|w^k\|^2|\mathcal{F}_k].$$

By Assumption 3.1 (a) and (c),

$$\sum_{k=1}^{\infty}\beta_k < \infty.$$

By applying Lemma 3.1 to the recursive equation

$$\mathbb{E}[\Psi_{\alpha\beta}(x^{k+1})|\mathcal{F}_k] \leq \Psi_{\alpha\beta}(x^k) - \gamma_k + \beta_k,$$

we show almost surely that $\Psi_{\alpha\beta}(x^k)$ is convergent, $\{\Psi_{\alpha\beta}(x^k)\}$ is bounded, and $\sum_{k=1}^{\infty}\gamma_k < \infty$. The latter implies that $\sum_{k=1}^{\infty}\|d^k\|^2 < \infty$ almost surely, and $\lim_{k\to\infty} d^k = 0$ almost surely. By Lemma 4.1 (d), almost sure boundedness of $\{\Psi_{\alpha\beta}(x^k)\}$ implies that $\{x^k\}$ is bounded almost surely. Furthermore, since $\nabla_{\alpha,\beta}\Psi(x^k)$ is bounded almost surely, Lemma 4.1 (a) implies that

$$0 = \lim_{k\to\infty}\nabla\Psi_{\alpha\beta}(x^k)^T d^k \leq \lim_{k\to\infty} -\frac{\sigma}{2}(\|r(x^k)\| + \rho\|s(x^k)\|)^2 \leq 0, \text{almost surely}.$$

Therefore, for any accumulation point x^* of $\{x^k\}$, $r(x^*) = s(x^*) = 0$, i.e., x^* is a solution of SVIP according to Lemma 4.1 (c). By the strong monotonicity property of F, x^* is the unique solution of SVIP and $\{x^k\}$ converges to x^*. ■

5 Stochastic Approximation Methods for Stochastic Nonlinear Complementarity Problems

In the preceding sections, we proposed stochastic approximation methods for solving SVIP (1.1). Theoretically, these methods can be applied to solving SNCP (1.2) as the latter is a special case of the former. However, we are motivated to consider specific iterative scheme for SNCP for three main reasons. (a) The SA methods proposed so far are designed for general SVIP without exploiting specific structures of SNCP; in particular the methods in Section 4 based on gap functions are typically designed for SVIP rather than SNCP as this is well known in the deterministic case. (b) The conditions imposed for convergence of SA methods for SVIP may be weakened when SVIP is reduced to SNCP. For instance, Theorem 4.1 only applies to the case when F is a linear affine function, and Theorem 4.2 requires F to be strongly monotone. Alternative methods for SNCP, which require weaker conditions for the convergence analysis, may be possible. (c) Numerical methods based on NCP functions such as the Fischer-Burmeister function are very popular and powerful for solving deterministic NCPs. It is therefore natural for us to consider specialized SA methods based on these NCP functions for solving SNCP.

Specifically, we reformulate SNCP as a stochastic nonsmooth system of equations and then solve the equations via least-squared minimization. Recall that Fischer-Burmeister function $\phi : \mathbb{R}^2 \to \mathbb{R}$ is defined as

$$\phi(a,b) = \sqrt{a^2+b^2} - a - b.$$

Using this function, SNCP (1.2) is equivalent to the following system of stochastic equations

$$H(x) \equiv \begin{pmatrix} \phi(x_1, F_1(x)) \\ \vdots \\ \phi(x_n, F_n(x)) \end{pmatrix} = 0 \tag{5.16}$$

in the sense that the solution set of (5.16) coincides with that of (1.2). One of the main benefits in using the Fischer-Burmeister function is that it is semismooth everywhere and continuously differentiable of any order at any point except the origin. Also the function is globally Lipschitz continuous. It is natural to propose SA methods for solving SSE (5.16) for solutions of the SNCP. By Proposition 4.4 of [2], convergence of SA for SSE usually requires H in (5.16) to be strongly monotone and globally Lipschitz continuous. However the following example shows that H in (5.16) is not necessarily monotone even when F is strongly monotone.

Example 5.1 *Let* $F(x) = Ax + b$ *where* $A = \begin{pmatrix} 2 & 1 \\ -10 & 20 \end{pmatrix}$, *and* $b = (0, -20)^T$. *It is easy to verify that* $A + A^T$ *is positive definite. Therefore* A *is positive definite and* F *is strongly monotone. In what follows, we show that* H *is not monotone at the point* $(0,1)$, *which is equivalent to showing that* $\nabla H(0,1)$ *is not positive semi-definite. Since* $F_1(0,1) = 1$,

$F_2(0,1) = 0$, *H is continuously differentiable at* $(0,1)$ *and* $\nabla H(0,1) = \begin{pmatrix} -1 & 0 \\ 10 & -20 \end{pmatrix}$. *Therefore* $\nabla H(0,1) + \nabla H(0,1)^T = \begin{pmatrix} -2 & 10 \\ 10 & -40 \end{pmatrix}$. *The above matrix has two real valued eigenvalues:* 0.4709 *and* -42.4709. *This shows* $\nabla H(0,1) + \nabla H(0,1)^T$, *or equivalently,* $\nabla H(0,1)$ *is indefinite. By continuity of* ∇H *at* $(0,1)$, $\nabla H(\cdot)$ *is indefinite in a small neighborhood of point* $(0,1)$. *This shows that* H *is not monotone in the neighborhood.*

The above example discourages us to consider a Robbins-Monro type iterative scheme for solving SSE (5.16). In what follows, we consider SA methods based on a minimization reformulation of SNCP. Let

$$\min_x \Psi(x) \equiv \frac{1}{2}\|H(x)\|^2. \tag{5.17}$$

Under suitable conditions [8], the global solution set of (5.17) is the same as the solution set of SNCP. Analogous to Theorem 4.1, one may attempt to propose SA methods for solving SNCP based on (5.17). If one can prove strong convexity of Ψ and global Lipschitz continuity of $\nabla\Psi$, then it follows from Theorem 5.3 of [27] that the SA method based on the merit function Ψ converges almost surely to a global solution of (5.17) and hence a solution of the SNCP.

Instead of looking for conditions that ensure convergence of SA methods applied to reformulations (5.16) and (5.17), we propose an SA method in the spirit of iterative scheme (4.15), which is a derivative-free approach. Consider the following SA scheme for solving (5.17):

$$x^{k+1} = x^k - a_k(d^k + \omega^k), \tag{5.18}$$

where

$$d_i^k = -\phi(x_i^k, F_i(x^k))\nabla_b\phi(x_i^k, F_i(x^k)), i = 1, \cdots, n, \tag{5.19}$$

and ω^k is the stochastic error when approximation of d^k is obtained from a sample. For instance, if ξ^k is a sample of $\xi(\theta)$, then we may choose $d^k + w^k = -\phi(x_i^k, f_i(x^k, \xi^k))\nabla_b\phi(x_i^k, f_i(x^k, \xi^k))$.

The search direction is used for developing derivative-free iterative scheme for solving NCP in [13, 19]. The following result states that under proper conditions, d^k is a descent direction of Ψ at x^k and $d^k = 0$ implies that x^k is a solution of SNCP.

Lemma 5.1

(a) x is a solution of SNCP (1.2) if and only if $d = 0$, where

$$d_i = -\phi(x_i, F_i(x))\nabla_b\phi(x_i, F_i(x)), i = 1, \cdots, n.$$

(b) If F is continuously differentiable and there exists $\sigma_k > 0$ such that $d^T\nabla F(x^k)d \geq \sigma_k\|d\|^2$, then d^k is a descent direction of Ψ at x^k and

$$\nabla\Psi(x^k)^T d^k \leq -\sigma_k\|d^k\|^2. \tag{5.20}$$

Proof. (a) is proved in [13, Lemma 4.1]. The result in (b) is analogous to [13, Lemma 4.1]. The only difference is that modulus σ^k depends on x^k whereas the modulus used in [13] is a constant. We omit details for the proof. ■

Next, we analyze convergence of sequence $\{x^k\}$ generated by iterative scheme (5.18). We require conditions similar to those in iterative scheme (3.1). A proof for Lemma 5.2 below is provided in the appendix.

Lemma 5.2 *Let $\psi(a,b) = \phi(a,b)^2$ where ϕ is the Fischer-Burmeister function. Then (a) ψ is continuously continuously differentiable; (b) ψ is twice continuously differentiable over $\mathbb{R}^2\backslash\{(0,0\}$; (c) $\nabla\psi$ is locally Lipschitz over $\mathbb{R}^2$; (d) the Clarke generalized Jacobian of $\nabla\psi$ is bounded over $\mathbb{R}^2$; (e) $\nabla\psi$ is globally Lipschitz continuous over $\mathbb{R}^2$.*

Proposition 5.1 *Suppose that F is globally Lipschitz continuous and twice continuously differentiable, and there exist positive constants C_1, C_2 such that*

$$\max_i(\|x\| + C_1)\|\nabla^2 F_i(x)\| \leq C_2, \qquad \forall x \in \mathbb{R}^n. \tag{5.21}$$

Then $\nabla\Psi$ is globally Lipschitz continuous over $\mathbb{R}^n$.

A proof for Proposition 5.1 can be found in the Appendix. Note that Condition (5.21) is satisfied when F is a linear affine function.

Theorem 5.1 *Suppose that Assumption 3.1 (a), (b), (c) and (d) hold, for iterative scheme (5.18). Assume that F is twice continuously differentiable, that Condition (5.21) is satisfied, and there exist $t \in (1,2)$ and $C > 0$ and a continuous function $\sigma(x) > 0$ such that*

$$(F(y) - F(x))^T(y - x) \geq \min(\sigma(x)\|y - x\|^2, C\|y - x\|^t),\ \forall y \in \mathbb{R}^n, \tag{5.22}$$

and the stepsize satisfies

$$0 < a_k \leq \frac{\sigma(x^k)}{2L}. \tag{5.23}$$

Then sequence $\{x^k\}$ generated by this scheme almost surely converges to the unique solution x^* of SVIP.

Before providing a proof, we note that condition (5.23) on stepsize choice here is more restrictive than Assumption 3.1 (a) because it must be bounded above by $\frac{\sigma(x^k)}{2L}$, which implies that we need some knowledge at x^k when choosing a_k. Technically this is feasible because we do not have to select the sequence of stepsizes at the beginning of the iterative scheme. This condition is automatically satisfied when F is strongly monotone. But our intention here is to cover other monotone functions which are not strongly monotone. For instance the function in

Example 3.1 is monotone but not strongly monotone. This function, however, satisfies Condition (5.22). Note that the ξ-monotonicity [8, Definition 2.3.1] implies (5.22).

Proof of Theorem 5.1. We first check the condition (5.20) of Lemma 5.1 (b). Consider (5.22). Let $d \in \mathbb{R}^n$ with $\|d\| = 1$ be fixed. Let $y = x + \tau d$. Then for $\tau > 0$ sufficiently small, $\frac{1}{4}\tau^2\|d\|^2$ is dominated by $\frac{1}{4}\|\tau d\|^t$. Since F is continuously differentiable, it follows from (5.22) that

$$d^T \nabla F(x^k) d = \lim_{\tau \downarrow 0} (F(x^k + \tau d) - F(x^k))^T d/\tau \geq \sigma(x^k).$$

This shows that the condition in Lemma 5.1 (b) holds.

Next, we will use Lemma 3.1 to prove our main result. By virtue of the mean value theorem, there exists y^k located at a point on the line segment between x^k and x^{k+1} such that

$$\Psi(x^{k+1}) = \Psi(x^k) + \nabla\Psi(y^k)^T(x^{k+1} - x^k).$$

By Proposition 5.1, $\nabla\Psi$ is globally Lipschitz continuous. Therefore there exists $L > 0$ such that

$$\|\nabla\Psi(y^k) - \nabla\Psi(x^k)\| \leq L\|x^{k+1} - x^k\|.$$

By iterative scheme (5.18) and the fact that x^* is a solution of SNCP, we have

$$\begin{aligned}
\mathbb{E}[\Psi(x^{k+1})|\mathcal{F}_k] &= \mathbb{E}[\Psi(x^k) + \nabla\Psi(y^k)^T(x^{k+1} - x^k)|\mathcal{F}_k] \\
&= \mathbb{E}[\Psi(x^k) + a_k\nabla\Psi(y^k)^T(d^k + w^k)|\mathcal{F}_k] \\
&= \Psi(x^k) + a_k\mathbb{E}[\nabla\Psi(y^k)(d^k + w^k)|\mathcal{F}_k] \\
&= \Psi(x^k) + a_k\mathbb{E}[\nabla\Psi(x^k)^T(d^k + w^k)|\mathcal{F}_k] + \\
&\quad\ a_k\mathbb{E}[(\nabla\Psi(y^k) - \nabla\Psi(x^k))^T(d^k + w^k)|\mathcal{F}_k] \\
&= \Psi(x^k) + a_k\nabla\Psi(x^k)^T d^k + 0 + a_k\mathbb{E}[(\nabla\Psi(y^k) - \nabla\Psi(x^k))^T(d^k + w^k)|\mathcal{F}_k] \\
&\leq \Psi(x^k) + a_k\nabla\Psi(x^k)^T d^k + a_k\mathbb{E}[L\|y^k - x^k\|\|d^k + w^k\||\mathcal{F}_k] \\
&\quad\ \text{(By Proposition 5.1)} \\
&\leq \Psi(x^k) + a_k\nabla\Psi(x^k)^T d^k + L(a_k)^2\mathbb{E}[\|d^k + w^k\|^2] \\
&\quad\ \text{(By (5.18))} \\
&\leq \Psi(x^k) + a_k\nabla\Psi(x^k)^T d^k + 2L(a_k)^2(\|d^k\|^2 + \mathbb{E}[\|w^k\|^2|\mathcal{F}_k] \\
&\quad\ \text{(By Assumption 3.1 (b))} \\
&\leq \Psi(x^k) - a_k\sigma(x^k)\|d^k\|^2 + 2L(a_k)^2(\|d^k\|^2 + \mathbb{E}[\|w^k\|^2|\mathcal{F}_k] \\
&\quad\ \text{(By (5.20))} \\
&= \Psi(x^k) + a_k(2La_k - \sigma(x^k))\|d^k\|^2 + 2L(a_k)^2\mathbb{E}[\|w^k\|^2|\mathcal{F}_k] \\
&= \Psi(x^k) - \gamma_k + \beta_k,
\end{aligned}$$

where $\gamma_k = -a_k(2La_k - \sigma(x^k))\|d^k\|^2$, and $\beta_k = 2L(a_k)^2\mathbb{E}[\|w^k\|^2|\mathcal{F}_k]$. By (5.23) and Assumption 3.1 (a), $\gamma_k \geq 0$, and by Assumptions 3.1 (a) and (c), $\sum_{k=1}^{\infty}\beta_k < \infty$. Applying Lemma 3.1 to the recursive equation

$$\mathbb{E}[\Psi(x^{k+1})|\mathcal{F}_k] \leq \Psi(x^k) - \gamma_k + \beta_k,$$

we show that $\Psi(x^k)$ is convergent almost surely and $\sum_{k=1}^{\infty}\gamma_k < \infty$.

We next show that $\Psi(x^k)$ is convergent to 0 for every sample path corresponding to convergence of $\Psi(x^k)$. Let the sample path be fixed.

First, we prove that x^k is bounded. Assume for the sake of a contradiction that $\|x^k\| \to \infty$ for the sample path at which $\Psi(x^k)$ is convergent. Let

$$J \equiv \{i \in \{1, 2, \cdots, n\} : \{x_i^k\} \text{ is unbounded}\}.$$

Then $J \neq \emptyset$. Define $y^k \in \mathbb{R}^n$ as follows

$$y_i^k \equiv \begin{cases} 0, & \text{if } i \in J, \\ x_i^k, & \text{if } i \notin J. \end{cases}$$

Then $\{y^k\}$ is bounded, that is, there exists a constant $\bar{C} > 0$ such that y^k is located within $\bar{C}\mathcal{B}$, where $\mathcal{B}$ denotes the closed unit ball in $\mathbb{R}^n$. Let

$$\sigma = \inf_{x \in \bar{C}\mathcal{B}} \sigma(x).$$

For sufficiently large k, $\sigma(y^k)\|x^k - y^k\|^2 > C\|x^k - y^k\|^t$. By (5.22), we have

$$\begin{aligned} C\left(\sum_{i\in J}(x_i^k)^2\right)^{t/2} &= C\|x^k - y^k\|^t \\ &\leq \sum_{i=1}^{n}(x_i^k - y_i^k)(F_i(x^k) - F_i(y^k)) \\ &\leq \sqrt{\sum_{i\in J}(x_i^k)^2}\sum_{i=1}^{n}|F_i(x^k) - F_i(y^k)|. \end{aligned}$$

Following a similar argument as in the proof of [13, Theorem 3.2], we can prove that there exists an index $i_0 \in J$ such that $|x_{i_0}^k| \to \infty$, and $|F_{i_0}(x^k)| \to \infty$. By [13, Lemma 3.1], $\psi(x_{i_0}^k, F_{i_0}(x^k)) \to \infty$, as $k \to \infty$, which implies $\Psi(x^k) \to \infty$, which contradicts the fact that $\{\Psi(x^k)\}$ is convergent. Therefore, $\{x^k\}$ is bounded.

Because $\{x^k\}$ is bounded, there exists $\hat{\sigma} > 0$ such that $\sigma(x^k) \geq \hat{\sigma} > 0$. Consequently we can derive from $\sum_{k=1}^{\infty}\gamma_k < \infty$ that $d^k \to 0$ as $k \to \infty$. By Lemma 5.1 (a) and Assumption 3.1 (a), we show that $\{x^k\}$ converges to the unique solution of SNCP (1.2). Almost sure convergence follows from the fact that the argument above holds for every sample path corresponding to convergence of $\Psi(x^k)$. ■

Note that Conditions (5.21) and (5.22) play an important role in the above theorem. The next example shows that these two conditions do hold for some functions that are not strongly monotone.

Example 5.2 *Consider the function in Example 3.1. We shall prove that (5.22) holds when* $C = \frac{1}{4}$ *and* $t = \frac{3}{2}$ *and*

$$\sigma(x) = \begin{cases} \frac{1}{4\sqrt{x}} & \text{if } x \geq 1, \\ \frac{1}{4} & \text{otherwise.} \end{cases}$$

We only need to prove that for any $x, y \in \mathbb{R}$,

$$(F(y) - F(x))(y - x) \geq E(x, y), \tag{5.24}$$

where

$$E(x,y) = \begin{cases} \frac{1}{4}(y-x)^{\frac{3}{2}} & \text{if } 1 < x, 9x < y, \\ \frac{1}{4\sqrt{x}}(y-x)^2 & \text{if } 1 < x, 1 \leq y \leq 9x, \\ \frac{1}{4\sqrt{x}}(y-x)^2 & \text{if } 1 < x, 0 \leq y < 1, \\ \frac{1}{4\sqrt{x}}(y-x)^2 & \text{if } 1 < x, y < 0, \\ \frac{1}{4}(y-x)^2 & \text{if } 0 \leq x \leq 1, y < 1, \\ \frac{1}{4}(y-x)^2 & \text{if } 0 \leq x \leq 1, 1 \leq y \leq 4, \\ \frac{1}{4}(y-x)^{\frac{3}{2}} & \text{if } 0 \leq x \leq 1, 4 < y, \\ \frac{1}{4}(y-x)^2 & \text{if } x < 0, y \leq 1, \\ \frac{1}{4}(y-x)^{\frac{3}{2}} & \text{if } x < 0, 1 < y. \end{cases}$$

The inequality (5.24) can be proved using the definition of F and some simple and direct calculations. Here we only provide some simple clues for each case but omit tedious detail.

$$\begin{array}{ll} \text{If } 1 < x, 9x < y: & \sqrt{y-x} \leq \sqrt{y}, 3\sqrt{x} \leq \sqrt{y}. \\ \text{If } 1 < x, 1 \leq y \leq 9x: & (\sqrt{y}-\sqrt{x})(\sqrt{y}+\sqrt{x}) = y - x, \sqrt{y} \leq 3\sqrt{x}. \\ \text{If } 1 < x, 0 \leq y < 1: & -y \leq -2y\sqrt{x}, \sqrt{x} \leq x. \\ \text{If } 1 < x, y < 0: & -(1+x) \leq -2\sqrt{x}, -x \leq -xy. \\ \text{If } 0 \leq x \leq 1, y < 1: & \text{trivial.} \\ \text{If } 0 \leq x \leq 1, 1 \leq y \leq 4: & y \leq 4\sqrt{y} - 3, x \leq 1. \\ \text{If } 0 \leq x \leq 1, 4 < y: & 2 \leq \sqrt{y}, x \leq \sqrt{y}, \sqrt{y-x} \leq \sqrt{y}. \\ \text{If } x < 0, y \leq 1: & \text{trivial.} \\ \text{If } x < 0, 1 < y: & 1 \leq \sqrt{y}, \sqrt{y-x} \leq \sqrt{y} - x. \end{array}$$

Finally Condition (5.21) can be verified easily by calculating the first and the second order derivative of F.

When F is an affine strongly monotone function, Conditions (5.21) and (5.22) in Theorem 5.1 are satisfied, and Condition (5.23) is redundant. Hence we have the following result.

Corollary 5.1 *Suppose that Assumption 3.1 (a), (b) and (c) hold for for iterative scheme (5.18). Suppose also that F is a strongly monotone affine function, that is, $F(x) = Ax + b$ where A is positive definite. Then sequence $\{x^k\}$ generated by iterative scheme (5.18) almost surely converges to the unique solution x^* of SVIP.*

Before we conclude this section, it is worthwhile to point out what has been done in the literature on SA for SVIP and SNCP. Flam [10] proposes a projection based SA for solving SNCP that is formulated from a number of optimization problems describing equilibrium systems. A major difference between Flam's SA and iterative scheme (3.7) is that the projection in the former method is carried out in the feasible solution space of variable x_i rather than $\mathcal{Y} = \mathbb{R}^n_+$. Under similar conditions to Assumption 3.1, Flam proves convergence of his SA method.

6 Sample Applications

In this section, we present a few SVIP and SNCP examples arising from the areas of economics, engineering and operations management.

Example 6.1 Stochastic User Equilibrium *[40]. Network equilibrium models are commonly used for predictions of traffic patterns in transportation networks which are subject to congestion. Network equilibrium is characterized by Wardrop's two principles. The first principle states that the journey times in all routes actually used are equal to or less than those which would be experienced by a single vehicle on any unused route. The traffic flows satisfying this principle are usually referred to as user equilibrium flows. The second principle states that at equilibrium the average journey time is minimum.*

A variant of user equilibrium is the stochastic user equilibrium (SUE) in which each traveller attempts to minimize their perceived dis-utility/costs, where these costs are composed of a deterministic measured cost and a random term. For each origin-destination (OD) pair j in the traffic network and a particular path r of OD j, the user's dis-utility function is defined by $u_r = \theta_0 d_r + \theta_1 \mathbb{E}[C_r] + \theta_2 \mathbb{E}[\max(0, C_r - \tau_j)]$, where d_r represents the composite of attributes such as distance which are independent of time/flow, C_r denotes the travel time on path r which is implicitly determined by the flows on all arcs on path r, τ_j denotes the longest acceptable travel time for j, θ_0 is the weight placed on these attributes, θ_1 is the weight placed on time, and θ_2 is the penalty coefficient when the actual travel time on j exceeds τ_j. Let x_r denote the traffic flow on path r and $\mathcal{R}_j$ denote the collection of all feasible paths for j. Assume that the total demand for OD pair j is q_j. Then the feasible set of the traffic flow across the whole network can be expressed as $\mathcal{X} = \left\{x : \sum_{r\in\mathcal{R}_j} x_r = q_j, \forall j, x_r \geq 0, \forall r\right\}$, which is a convex set.

A vector $x^* \in \mathcal{X}$ is a stochastic user equilibrium if and only if $(x - x^*)^T u(x^*) \geq 0, \forall x \in \mathcal{X}$, which is an SVIP.

Example 6.2 Electricity Supply Networks *[8].Oligopolistic pricing models have wide applicability in spatially separated electricity markets. The aim of these models is to determine the amount of electricity produced by each competing firm, the flow of power, and the transmission prices through the links of the electricity network. We describe a simplified, single-period, spatially separated, oligopolistic electricity pricing model with random demand. A slightly different example is presented in [28].*

Consider an electricity network with node set $\mathcal{N}$ and arc set $\mathcal{A}$. As many as n firms compete *to supply electricity to the network. Each firm i owns generation facilities in a subset of nodes $\mathcal{N}_i \subset \mathcal{N}$. Let $\mathcal{G}_{ij}$ denote the set of generation plants by firm i at node $j \in \mathcal{N}_i$ and $q_{ij\ell}$ the amount of electricity produced by firm i at plant $\ell \in \mathcal{G}_{ij}$. Market demand at node j is described by an inverse demand function $p_j(\sum_{i=1}^n d_{ij}, \xi_j)$ which is a function of the total amount of electricity $\sum_{i=1}^n d_{ij}$ supplied to node j and a random shock ξ_j, where d_{ij} is the amount of electricity delivered to node j by firm i.*

Let r_{ia} be the amount of electricity transmitted through arc $a \in \mathcal{A}$ by firm i. Then $d_{ij}, j \in \mathcal{N}_i$, $q_{ij\ell}, j \in \mathcal{N}_i, \ell \in \mathcal{G}_{ij}$ and $r_{ia}, a \in \mathcal{A}$ are all decision variables for firm i, which is collectively denoted by x_i. Firm i needs to make a decision before demands at nodes are realized and its decision problem is to maximize their profit which is equal to the total revenue minus the total production cost and transmission cost:

$$u_i(x_i, x_{-i}) = \mathbb{E}\left[\sum_{j\in\mathcal{N}_i} p_j(\sum_{i=1}^n d_{ij}, \xi)d_{ij} - \sum_{j\in\mathcal{N}_i}\sum_{\ell\in\mathcal{G}_{ij}} C_{ij\ell}(q_{ij\ell}) - \sum_{a\in\mathcal{A}} \rho_a r_{ia}\right],$$

where x_{-i} is the joint decision variables for all other firms except firm i, $C_{ij\ell}$ is the production cost function at plant ℓ of node j for firm i, and ρ_a is the unit transmission cost on arc a.

In order for the electricity flow to be feasible, it must satisfy the flow balance at each node, production capacity at each plant, and transmission capacity on each arc. Let $\mathcal{X}_i(x_{-i})$, which is precisely defined on page 30 of [8], denote the feasible set of electricity flow for firm i given the joint decision variables x_{-i} for all other firms. Overall, firm i's profit maximization problem is $\max_{x_i} u_i(x_i, x_{-i})$ subject to $x_i \in \mathcal{X}_i(x_{-i})$. Following Proposition 1.4.3 of [8], this game theoretic model for the electricity supply network problem can be converted into an SVIP.

Example 6.3 Newsvendor Competition *[24]. The newsvendor (or newsboy) model is an important mathematical model in supply chain management, operations management and applied economics used to determine optimal inventory levels. It is typically characterized by fixed prices and uncertain demand.*

Suppose there are n players in the market who produce the same product. Let p_i and c_i be the unit price and unit production cost of newsvendor i. Assume the number of customers who prefer to buy the product from newsvendor i is D_i, which is a random variable. Customers always purchase the product from their unique preferred newsvendor provided that the product is available. However a proportion, say o_{ij}, of the customers of newsvendor j will purchase the product from newsvendor i if they find that newsvendor j does not have any product left unsold. Let q_i be the production level for newsvendor i. Given the production levels q_{-i} for all other newsvendors, newsvendor i chooses their optimal production level q_i by maximizing their expected profit: $u_i(q_i, q_{-i}) = p_i\mathbb{E}[\min(q_i, D_i + \sum_{j\neq i} o_{ij}\max(D_j - q_j, 0))] - c_i q_i$.

All newsvendors play an oligopolistic game by choosing their production levels appropriately. An optimal solution for the newsvendor game is a Nash equilibrium which states that no newsvendor will increase their expected profit by unilaterally altering their production level. It is well *known from Proposition 1.4.2 of [8] that any oligopolistic Nash game can be converted into an example of VIP under the conditions that the utility function is concave and continuously differentiable with respect to the player's own strategy. Hence this oligopolistic Nash game can be converted into an example of SVIP. Several other newsvendor examples of SNCP and SVIP can be found from [4, 23] and references of [24].*

Example 6.4 Wireless Networks *[25]. Wireless networks have dramatically changed the world and our daily life. Many wireless network problems can be formulated as game theoretic models. Consider a multipacket reception wireless network with n nodes and an uplink communication channel where the nodes communicate with a common base station.*

Let ξ_i denote the channel state of node i, which is a continuous random variable and $p_i(\xi_i)$ the decision variable which is a Lebesgue measurable function that maps channel state ξ to a transmission policy. The objective of node i is to find an optimal transmission policy function $p_i(\xi_i)$ that maximizes its individual utility (the expected value of a complicated function of $p_i(\xi_i)$). A particular transmission policy called threshold policy *characterizes p_i by a single real-valued parameter x_i, that is, $p_i(\xi_i) = 0$ if $\xi_i \leq x_i$, and 1 otherwise, where $x_i \in [0, M]$. In this case, the utility can be reformulated as a function of x_i, denoted by $T_i(x_i, x_{-i})$, where x_{-i} denotes the parameters that determine the policies of the other nodes. Consequently node i's decision problem is to maximize $T_i(x_i, x_{-i})$ for given x_{-i} subject to $x_i \in [0, M]$. This is a stochastic minimization problem with a single variable x_i.*

A wireless network solution is a Nash equilibrium $(x_1^*, \cdots, x_n^*)$ where no node is better off by unilaterally altering its strategies x_i. Similar to Example 6.3, under certain conditions, the multipacket reception wireless network problem can be reformulated as an SVIP.

7 Conclusions

In this paper, we have proposed several SA methods for solving stochastic variational inequality and stochastic nonlinear complementarity problems. They are iterative schemes generalized from their deterministic counterparts. Iterative scheme (3.7) is a projection-type method and it does not involve any calculation of derivatives. Therefore it is more suitable for those problems with a simple feasible set but a relatively complex structure of underlying functions. Iterative scheme (4.11) is based on the gap functions of SVIP and it allows us to explore high-order derivative information so that faster convergent algorithms can be designed. See the next paragraph for simultaneous perturbation stochastic approximation. Iterative scheme (5.18) is specifically proposed for SNCP and is based on the well-known Fischer-Burmeister function. It is interesting to note that implementation of this scheme does not require differentiability of the underlying functions although we required second order derivatives in the convergence analysis. Numerical efficiency of those proposed iterative schemes remains to be investigated.

A couple of topics are worth further investigation. First, averaging SA methods [30] are proven to speed up convergence for solving SSE and the stochastic optimization problem. It would be interesting to see whether or not averaging SA methods can also speed up convergence when they are used for solving SVIP and SNCP. Second, some SA methods such as simultaneous perturbation stochastic approximation [18] that incorporate higher order derivative information of the underlying functions for SSE and the stochastic optimization problem have been extensively examined in the literature. The same method might be explored for SVIP and SNCP.

Acknowledgements: The authors are thankful to three anonymous referees, Associate Editor Ji-Feng Zhang, Jong-Shi Pang, Danny Ralph, Stefan Scholtes, Defeng Sun and Roger Wets for their constructive comments and to Roger Wets for bringing reference [7] to our attention.

References

[1] T. Basar and G. Olsder, *Dynamic Noncooperative Game Theory*, SIAM, 1999.

[2] D.P. Bertsekas and J.N. Tsitsiklis, *Neuro-Dynamic Programming*, Athena Scientific, 1996.

[3] D. Bertsimas and S. de Boer, Simulation-based booking limits for airline revenue management, *Operations Research* 53 (2005) 90–106.

[4] F.Y. Chen, H. Yan and L. Yao, A newsvendor pricing game, *IEEE Transactions on Systems, Man & Cybernetics: Part A*, 34 (2004) 450–456.

[5] X. Chen and M. Fukushima, Expected residual minimization method for stochastic linear complementarity problems, *Mathematics of Operations Research*, to appear.

[6] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New-York, 1983.

[7] Y. Ermoliev, Stohastic quasigradient methods, In Y. Ermoliev and R.J-B. Wets (Eds.), *Numerical Techniques for Stochastic Optimization Problems*, Springer-Verlag: Berlin, Heidelberg, pp. 141–186 (1988).

[8] F. Facchinei and J.S. Pang, *Finite-dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, 2003.

[9] J. Filar and K. Vrieze, *Competitive Markov Decision Processes*, Springer-Verlag, 1997.

[10] S.D. Flam, Learning equilibrium play: A myopic approach, *Computational Optimization and Applications* 14 (1999) 87–102.

[11] M.C. Fu, Optimization for simulation: Theory vs. Practice, *INFORMS Journal on Computing* 14 (2002) 192–215.

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

[13] C. Geiger and C. Kanzow, On the resolution of monotone complementarity problems, *Computational Optimization and Applications* 5 (1996) 155–173.

[14] E.G. Golshtein and N.V. Tretyakov, *Modified Lagrangian and Monotone Maps in Optimization*, John Wiley and Sons, New York, 1996.

[15] G. Gürkan and A.Y. Özge and S.M. Robinson, Sample path solution of stochastic variational inequalities, with applications to option pricing, In J.M. Charnes and D.M. Morrice and D.T. Brunner and J.J. Swai (eds.): *Proceedings of the 1996 Winter Simulation Conference*, 1996, pp. 337–344.

[16] G. Gürkan, A.Y. Özge, and S.M. Robinson, Sample-path solution of stochastic variational inequalities, *Mathematical Programming* 84 (1999) 313-333.

[17] A. Haurie and F. Moresino, S-adapted oligopoly equilibria and approximations in stochastic variational inequalities, *Annals of Operations Research* 114 (2002) 183–201.

[18] Y. He and M. Fu and S.I. Marcus, Convergence of Simultaneous Perturbation Stochastic Approximation for Nondifferentiable Optimization, *IEEE Transactions on Automatic Control* 48 (2003) 1459–1463.

[19] H. Jiang, Unconstrained minimization approaches to nonlinear complementarity problems, *Journal of Global Optimization* 9 (1996) 169–181.

[20] H.J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Springer-Verlag, New York, 2003.

[21] G.H. Lin and M. Fukushima, New reformulations and smoothed penalty method for stochastic nonlinear complementarity problems, Technical Report, Kyoto University, 2003.

[22] Z.-Q. Luo, J.-S. Pang, and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, 1996.

[23] S. Mahajan and G. van Ryzin, Inventory competition under dynamic consumer choice, *Operations Research* 49 (2001) 646–657.

[24] S. Netessine and N. Rudi, Centralized and competition inventory models with demand substitution, *Operations Research* 51 (2003) 329-335.

[25] M.H. Ngo abd V.K. Krishnamurthy, Game theoretic cross-layer transmission policies in multipacket reception wireless networks, *IEEE Transactions on Signal Processing* 55 (2007) 1911–1926.

[26] J.M. Peng and M. Fukushima, A hybrid Newton method for solving the variational inequality problem via the D-gap function, *Mathematical Programming* 86 (1999) 367–386.

[27] G.C. Pflug, *Optimization of stochastic models, The Interface Between Simulation and Optimization*, Kluwer Academic Publishers, 1996.

[28] P.O. Pineau and P. Murto, An oligopolistic investment model of the finnish electricity market, *Annals of Operations Research* 121 (2003) 123–148.

[29] E.L. Plambeck, B.R. Fu, S.M. Robinson and R. Suri, Sample-path optimization of convex stochastic performances functions, *Mathematical Programming*, Vol. 75, pp. 137–176, 1996.

[30] B.T. Polyak and A.B. Juditsky, Acceleration of stochastic approximation by averaging, *SIAM Journal on Control and Optimization*, 30 (1992) 838–855.

[31] H. Robbins and S. Monro, On a stochastic approximation method, *Annals of Mathematical Statistics* 22 (1951) 400-407.

[32] H. Robbins and D. Siegmund, A convergence theorem for nonnegative almost supermartingales and some applications, In J.S. Rustagi, *Optimizing Methods in Statistics*, Academic Press, 1971, pp. 233–257.

[33] S.M. Robinson, Analysis of sample-path optimization, *Mathematics of Operations Research* 21 (1996) 513-528.

[34] A. Shapiro, Stochastic Programming by Monte Carlo Simulation Methods, Published electronically in: Stochastic Programming E-Print Series, 2000.

[35] A. Shapiro, Monte Carlo sampling methods, In A. Ruszczyński and A. Shapiro (eds.), *Handbooks in Operations Research and Management Science: Stochastic Programming*, Elsevier, 2003, pp. 353–425.

[36] J.C. Spall, Adaptive stochastic approximation by the simultaneous perturbation method, *IEEE Transactions on Automatic Control* 45 (2000) 1839–1853.

[37] J.C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, Wiley-Interscience, 2003.

[38] G. van Ryzin and J.I. McGill, Yield management without forecasting or optimization: An adaptive algorithm for determining seat protection levels, *Management Science* 46 (2000) 568–573.

[39] I.J. Wang and E.K.P. Chong, A deterministic analysis of stochastic approximation with randomized directions, *IEEE Transactions on Automatic Control* 43 (1998) 1745–1749.

[40] D. Watling, User equilibrium traffic network assignment with stochastic travel times and late arrival penalty, *European Journal of Operational Research 175* (2006) 1539–1556.

[41] N. Yamashita, K. Taji and M. Fukushima, Unconstrained optimization reformulations of variational inequality problems, *Journal of Optimization Theory and Applications* 92 (1997) 439–456.

8 Appendix

Proof of Proposition 2.1. Proofs for (a) and (c) can be found in [8].

(b) Let $H(x,y) = \frac{1}{2}(y-x)^T D(y-x)$. Then $\nabla_y H(x,y) = D(y-x)$. In view of the optimization problem in Definition 2.1, for any fixed $x \in \mathbb{R}^n$, $y^* = \Pi_{\mathcal{Y},D}(x)$ is the unique optimal solution. Then the first order necessary condition of the optimization problem indicates

$$\nabla_y H(x, y^*)^T (y - y^*) \geq 0, \qquad \forall y \in \mathcal{Y},$$

which with the symmetric property of D implies that

$$(y - \Pi_{\mathcal{Y},D}(x))^T D(y - \Pi_{\mathcal{Y},D}(x)) \geq 0, \qquad \forall y \in \mathcal{Y}.$$

Let x take two different values $x^1, x^2 \in \mathbb{R}^n$, let y take values of $\Pi_{\mathcal{Y},D}(x^1)$, $\Pi_{\mathcal{Y},D}(x^2)$. Then we obtain from the inequalities above

$$(\Pi_{\mathcal{Y},D}(x^2) - x^1)^T D(\Pi_{\mathcal{Y},D}(x^2) - \Pi_{\mathcal{Y},D}(x^1)) \geq 0,$$

and

$$(\Pi_{\mathcal{Y},D}(x^1) - x^2)^T D(\Pi_{\mathcal{Y},D}(x^1) - \Pi_{\mathcal{Y},D}(x^2)) \geq 0.$$

Adding the last two inequalities, we obtain

$$(\Pi_{\mathcal{Y},D}(x^1) - \Pi_{\mathcal{Y},D}(x^2))^T D(x^1 - x^2) \geq \|\Pi_{\mathcal{Y},D}(x^1) - \Pi_{\mathcal{Y},D}(x^2)\|_D^2,$$

i.e., $\Pi_{\mathcal{Y},D}$ is ISM under the D-norm with modulus 1.

(d) Let $E(x) = x - \Pi_{\mathcal{Y},D}(x - aD^{-1}F(x))$. Since $\Pi_{\mathcal{Y},D}$ is ISM under the D-norm with modulus 1, for any $x^1, x^2 \in \mathbb{R}^n$, we have

$$\begin{aligned}(\Pi_{\mathcal{Y},D}(x^1 - aD^{-1}F(x^1)) - \Pi_{\mathcal{Y},D}(x^2 - aD^{-1}F(x^2)))^T D(x^1 - aD^{-1}F(x^1) - x^2 + aD^{-1}F(x^2))\\ \geq \|\Pi_{\mathcal{Y},D}(x^1 - aD^{-1}F(x^1)) - \Pi_{\mathcal{Y},D}(x^2 - aD^{-1}F(x^2))\|_D^2.\end{aligned}$$

By the definition of $E(x)$, this is equivalent to the following

$$(x^1 - x^2 - E(x^1) + E(x^2))^T D(E(x^1) - E(x^2) - aD^{-1}F(x^1) + aD^{-1}F(x^2)) \geq 0.$$

Further rearrangements of the above inequality yield

$$\begin{aligned}(x^1 - x^2)^T D(E(x^1) - E(x^2)) &\geq (x^1 - x^2)^T D(aD^{-1}F(x^1) - aD^{-1}F(x^2)) + \|E(x^1) - E(x^2)\|_D^2\\ &\qquad -(E(x^1) - E(x^2))^T D(aD^{-1}F(x^1) - aD^{-1}F(x^2))\\ &\geq a\mu\|F(x^1) - F(x^2)\|^2 + \|E(x^1) - E(x^2)\|_D^2\\ &\qquad -a(E(x^1) - E(x^2))^T (F(x^1) - F(x^2))\\ &= \|E(x^1) - E(x^2)\|_D^2 - \tfrac{a}{4\mu}\|E(x^1) - E(x^2)\|^2\\ &\qquad +a\mu\|F(x^1) - F(x^2) - \tfrac{1}{2\mu}(E(x^1) + E(x^2))\|^2\\ &\geq \|E(x^1) - E(x^2)\|_D^2 - \tfrac{a}{4\mu}\|E(x^1) - E(x^2)\|^2\\ &\geq \|E(x^1) - E(x^2)\|_D^2 - \tfrac{a}{4\mu\lambda_{\min}(D)}\|E(x^1) - E(x^2)\|_D^2\\ &= (1 - \tfrac{a}{4\mu\lambda_{\min}(D)})\|E(x^1) - E(x^2)\|_D^2,\end{aligned}$$

where the second inequality follows the fact that F is ISM on $\mathcal{Y}$ with modulus μ, and the last inequality from (2.6). The result follows. ∎

Proof of Proposition 3.1. Part (a). Let

$$a(\varepsilon) := \min_{x\in\mathcal{Y}:\rho\geq\|x-x^*\|\geq\varepsilon} (F(x)-F(x^*))^T(x-x^*)$$

and

$$b(\varepsilon) := \min_{x\in\mathcal{Y}:\rho\geq\|x-x^*\|\geq\varepsilon} F(x^*)^T(x-x^*).$$

Since $\mathcal{Y}$ is a closed convex set, both $a(\varepsilon)$ and $b(\varepsilon)$ are well defined. It is easy to observe that

$$\inf_{x\in\mathcal{Y}:\rho\geq\|x-x^*\|\geq\varepsilon} F(x)^T(x-x^*) \geq a(\varepsilon)+b(\varepsilon).$$

The fact that x^* is a solution implies that $b(\varepsilon)\geq 0$, and the monotonicity property of F at x^* implies $a(\varepsilon)\geq 0$. It is easy to verify that $a(\varepsilon)>0$ when F is strictly monotone at x^*. In what follows we show that $b(\varepsilon)>0$ when $-F(x^*)$ is in the interior of the polar cone of the tangent cone of $\mathcal{Y}$ at x^*. To see this, let $T_{\mathcal{Y}}(x^*)$ denote the tangent cone of $\mathcal{Y}$ at x^* and $d\in T_{\mathcal{Y}}(x^*)$. By definition, $F(x^*)^Td>0$. Assume for a contradiction that for any $t>0$, there exists $x(t)\in\mathcal{Y}$, $\|x(t)-x^*\|\geq\varepsilon$ such that $F(x^*)^T(x(t)-x^*)\leq t$. Divide both sides of the inequality by $\|x(t)-x^*\|$ and let $d(t)=(x(t)-x^*)/\|x(t)-x^*\|$. Then by driving t to zero, we may get a subsequence of $\{d(t)\}$ such that it converges to d^* and $F(x^*)^Td^*\leq 0$. This contradicts the assumption as $d^*\in T_{\mathcal{Y}}(x^*)$. This shows $b(\varepsilon)>0$.

Part (b). Let x^{**} be another solution and (e') holds at the point. The monotonicity property of F implies that

$$(F(x^{**})-F(x^*))^T(x^{**}-x^*)=0.$$

Using this relation and (e'), we have

$$F(x^{**})^T(x^{**}-x^*)=F(x^*)^T(x^{**}-x^*)>0$$

and by symmetry

$$F(x^*)^T(x^*-x^{**})=F(x^{**})^T(x^*-x^{**})>0$$

a contradiction! The proof is complete. ∎

Proof of Lemma 5.2. (a) is proved in [13], (b) is obvious, and (c) is proved in Example 7.4.9 of [8]. We only need to prove (d) and (e).

(d) It follows from (b) that ψ is twice continuously differentiable over $\mathbb{R}^2\backslash\{(0,0\}$ and

$$\nabla^2\psi(a,b)=2\nabla\phi(a,b)\nabla\phi(a,b)^T+2\phi(a,b)\nabla^2\phi(a,b), \qquad \forall (a,b)\neq(0,0).$$

Therefore the Clarke generalized Jacobian of $\nabla\psi$ coincides with $\nabla^2\psi$ on $\mathbb{R}^2\backslash\{(0,0\}$. It is easy to calculate $\nabla\phi(a,b)$ and to prove that $\nabla\phi$ is bounded over $\mathbb{R}^2\backslash\{(0,0\}$. In order to prove that $\nabla^2\psi$ is bounded over $\mathbb{R}^2\backslash\{(0,0\}$, we only need to show that $\phi(a,b)\nabla^2\phi(a,b)$ is bounded over

$\mathbb{R}^2\backslash\{(0,0\}$. The latter holds by showing that $\phi(a,b)\frac{\partial^2\phi(a,b)}{\partial a^2}$, $\phi(a,b)\frac{\partial^2\phi(a,b)}{\partial b^2}$ and $\phi(a,b)\frac{\partial^2\phi(a,b)}{\partial a\partial b}$ are bounded over $\mathbb{R}^2\backslash\{(0,0\}$.

By a simple calculation

$$\frac{\partial^2\phi(a,b)}{\partial a^2} = \frac{1}{\sqrt{a^2+b^2}} - \frac{a^2}{\sqrt{(a^2+b^2)^3}}.$$

Therefore

$$\phi(a,b)\frac{\partial^2\phi(a,b)}{\partial a^2} = 1 - \frac{a+b}{\sqrt{a^2+b^2}} - \phi(a,b)\frac{a^2}{\sqrt{(a^2+b^2)^3}}.$$

The third term on the right hand side of the above equation can be written as

$$\frac{a^2}{a^2+b^2}\left(1-\frac{a+b}{\sqrt{a^2+b^2}}\right),$$

which is obviously bounded since

$$\left|\frac{a+b}{\sqrt{a^2+b^2}}\right| \le 2.$$

Thus $\phi(a,b)\frac{\partial^2\phi(a,b)}{\partial a^2}$ is bounded over $\mathbb{R}^2\backslash\{(0,0\}$. Similarly, we can prove that $\phi(a,b)\frac{\partial^2\phi(a,b)}{\partial b^2}$ is bounded over $\mathbb{R}^2\backslash\{(0,0\}$ as b and a are symmetric in ϕ.

Finally, we consider $\phi(a,b)\frac{\partial^2\phi(a,b)}{\partial a\partial b}$. Since

$$\phi(a,b)\frac{\partial^2\phi(a,b)}{\partial a\partial b} = -\frac{ab}{a^2+b^2}\left(1-\frac{a+b}{\sqrt{a^2+b^2}}\right),$$

it is easy to see that the right hand side is bounded. This shows $\phi(a,b)\nabla^2\phi(a,b)$ is bounded over $\mathbb{R}^2\backslash\{(0,0\}$.

So far we have shown that $\nabla^2\psi$ is bounded over $\mathbb{R}^2\setminus\{0,0\}$ except origin $(0,0)$. By [6, Definition 2.6.1], the Clarke generalized Jacobian of $\nabla\psi$ at $(0,0)$ is the convex hull of the limiting Jacobians of $\nabla^2\psi(x)$ as $x\to(0,0)$, and it is compact by Proposition 2.6.2 of [6]. Hence the boundedness of $\partial\nabla\psi(0,0)$ follows from that of $\nabla^2\psi(x)$.

(e) For any $x,y\in\mathbb{R}^2$, it follows from Proposition 2.6.5 of [6] that

$$\nabla\psi(x)-\nabla\psi(y)\in \mathrm{Co}\partial\nabla\psi([x,y])(y-x), \tag{8.25}$$

where $\mathrm{Co}\partial\nabla\psi([x,y])$ is the convex hull of all the Clarke generalized Jacobians of $\nabla\psi$ at z that belongs to the line segment between x and y. By (iv), there exists a positive constant L such that $\|\partial\nabla\psi(z)\|\le L$. This, together with (8.25), implies that for any $x,y\in\mathbb{R}^2$

$$\|\nabla\psi(x)-\nabla\psi(y)\|\le L\|x-y\|$$

i.e., $\nabla\psi$ is globally Lipschitz over $\mathbb{R}^2$. ■

Proof of Proposition 5.1. By definition

$$\Psi(x)=\frac{1}{2}\sum_{i=1}^{n}\psi(x_i,F_i(x)),$$

and

$$\nabla\Psi(x) = \sum_{i=1}^{n}[\nabla_a\psi(x_i, F_i(x))e_i + \nabla_b\psi(x_i, F_i(x))\nabla F_i(x)],$$

where e_i is a unit n-dimensional vector with the i-th component being 1 and others being 0.

By Lemma 5.2, $\nabla\psi$ is globally Lipschitz continuous over $\mathbb{R}^2$, and by the assumption, F is a globally Lipschitz continuous. Therefore $\nabla_a\psi(x_i, F_i(x))$ is globally Lipschitz continuous as it is a composition of two globally Lipschitz functions over $\mathbb{R}^n$.

To prove that $\nabla\Psi$ is globally Lipschitz continuous over $\mathbb{R}^n$, we are now left to prove that for any $i = 1, \cdots, n$, $\nabla_b\psi(x_i, F_i(x))\nabla F_i(x)$ to be globally Lipschitz continuous over $\mathbb{R}^n$. By proposition 2.6.6 of [6], $\nabla_b\psi(x_i, F_i(x))\nabla F_i(x)$ is locally Lipschitz and its generalized Jacobian at x is contained in the following set

$$\Omega(x) \equiv \nabla F_i(x)\partial\nabla_b\psi(x_i, F_i(x))^T + \nabla_b\psi(x_i, F_i(x))\nabla^2 F_i(x).$$

It suffices to prove $\{\Omega(x) : x \in \mathbb{R}^n\}$ is bounded in order to prove that $\nabla_b\psi(x_i, F_i(x))\nabla F_i(x)$ to be globally Lipschitz continuous over $\mathbb{R}^n$. The boundedness of $\{\Omega(x) : x \in \mathbb{R}^n\}$ follows from the facts below:

- $\partial\nabla_b\psi(x_i, F_i(x))$ is bounded by Lemma 5.2,
- $\nabla F_i(x)$ is bounded because F is globally Lipschitz,
- $$\begin{aligned} \|\nabla_b\psi(x_i, F_i(x))\| &= 2\|\phi(x_i, F_i(x))\nabla_b\phi(x_i, F_i(x))\| \\ &\le 4\max(|x_i|, |F_i(x)|) \\ &\le 4\max(|x_i|, |F_i(x)|) \\ &= O(\|x\| + C_1) \end{aligned}$$

  where C_1 is a positive constant,
- Condition (5.21).

The proof is complete. ■