Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 16, No. 2, pp. 261–287, 2020

A CLASS OF STOCHASTIC SECOND-ORDER-CONE COMPLEMENTARITY PROBLEMS*

Guo Sun, Liying Yu, Gui-Hua Lin† and Xiaodai Dong

Abstract: The second-order-cone programs constitute an important class of convex optimization problems and as their Karush-Kuhn-Tucker conditions, the second-order-cone complementarity problems have attracted much attention of researchers. In this paper, we consider a class of stochastic second-order cone complementarity problems (SSOCCP). Since uncertainty often occurs in practice, we can find many applications of the considered SSOCCP. With the help of some complementarity functions, we transform the SSOCCP as a stochastic programming problem and by employing some approximation techniques, we present an approximation method. Under some moderate conditions, we establish a comprehensive convergence analysis. Furthermore, we derive some results related to exponential convergence rate and error bounds. Finally, we report some numerical experiments on a chance-constrained optimal power flow problem.

Key words: *SSOCCP, complementarity function, error bound, Monte Carlo approximation, optimal power flow*

Mathematics Subject Classification: *90C33, 90C15*

1 Introduction

In this paper, we consider the stochastic second-order-cone (SOC) complementarity problem (SSOCCP) of finding vectors $x, y \in \Re^n$ and $z \in \Re^l$ such that

$$x \in \mathcal{K}, \quad y \in \mathcal{K}, \quad x^T y = 0, \quad \mathbb{E}_\xi[F(x, y, z, \xi)] = 0, \tag{1.1}$$

where $F : \Re^n \times \Re^n \times \Re^l \times \Omega \to \Re^n \times \Re^l$ with Ω referring to the support of the random variable ξ, $\mathbb{E}_\xi$ denotes the expectation operator with respect to ξ, and $\mathcal{K} := \mathcal{K}^{n_1} \times \cdots \times \mathcal{K}^{n_m}$ with $n_1 + \cdots + n_m = n$ and the n_i-dimensional second-order-cone $\mathcal{K}^{n_i}$ is defined by

$$\mathcal{K}^{n_i} := \{(x_1, x_2) \in \Re \times \Re^{n_i - 1} \mid \|x_2\| \le x_1\}.$$

Problem (1.1) is clearly a generalization of SOC complementarity problems, which have been extensively studied in the optimization world (see the survey paper [6] for details), and it is also closely related to the stochastic variational inequality problem studied in [13] (see [17, 26, 29, 31] for details about recent developments on stochastic variational inequality

*This work was supported in part by NSFC (Nos. 11671250, 11431004, 71831008), Shandong Province Natural Science Fund (No. ZR2014AM012), Humanity and Social Science Foundation of Ministry of Education of China (No. 15YJA630034), and NSERC (No. 261524).

†Corresponding author.

and complementarity problems). In addition, problem (1.1) may be regarded as the Karush-Kuhn-Tucker (KKT) conditions of stochastic SOC programs

$$\begin{array}{ll} \min & \mathbb{E}_\xi[f(u,\xi)] \\ \text{s.t.} & g(u) \le 0,\ h(u) = 0, \\ & x \in \mathcal{K}, \end{array}$$

provided that the orders of expectation and gradient involved above are exchangeable under some dominated conditions. As is well-known to us, SOC programs and their stochastic versions have lots of applications in engineering design, portfolio optimization, etc. [1, 3]. Therefore, it is meaningful to study the theoretical and algorithmic aspects of the SSOCCP (1.1).

Note that the SSOCCP given above is different from the one considered in the recent work [16] and formulated as

$$x \in \mathcal{K}, \quad y \in \mathcal{K}, \quad x^T y = 0, \quad F(x,y,z,\xi) = 0 \qquad a.e.\ \xi \in \Omega, \tag{1.2}$$

where *a.e.* is the abbreviation for "almost every". According to the works [8, 17, 26] on stochastic complementarity problems, which have attracted much attention in the recent optimization literature, problem (1.1) can be called an expected value formulation for problem (1.2). From this point of view, our study may be regarded as a further supplement of the work [16], in which the authors discuss the so-called expected residual minimization formulation for problem (1.2) only.

As in the deterministic SOC complementarity theory, the SOC complementarity functions will play a key role in our study. Recall that a function $\phi : \Re^\nu \times \Re^\nu \to \Re^\nu$ is called an SOC complementarity function associated with a second-order-cone $\mathcal{K}^\nu$ if

$$\phi(s,t) = 0 \quad \Longleftrightarrow \quad s \in \mathcal{K}^\nu,\ t \in \mathcal{K}^\nu,\ s^T t = 0. \tag{1.3}$$

Two well-known SOC complementarity functions associated with $\mathcal{K}^\nu$ will be used in this paper: one is the natural residual function associated with $\mathcal{K}^\nu$ defined as

$$\phi_{\rm NR}(s,t) := s - [s-t]_+ \tag{1.4}$$

and the other is the vector-valued Fischer-Burmeister function associated with $\mathcal{K}^\nu$ defined as

$$\phi_{\rm FB}(s,t) := s + t - (s^2 + t^2)^{1/2}, \tag{1.5}$$

where $[\,\cdot\,]_+$ denotes the projection operator onto the convex cone $\mathcal{K}^\nu$, $s^2 := s \circ s$ with '$\circ$' referring to the Jordan product operator, and $s^{1/2}$ is the unique square root in $\mathcal{K}^\nu$ of s under the Jordan algebra. See Section 2 for details about the Jordan algebras.

By means of the SOC complementarity function (1.3), it is easy to see that the SSOCCP (1.1) is equivalent to the minimization problem

$$\min_{(x,y,z)} \ \Theta(x,y,z) := \|\mathbb{E}_\xi[F(x,y,z,\xi)]\|^2 + \|\Phi(x,y)\|^2, \tag{1.6}$$

where Φ is defined by

$$\Phi(x,y) := \begin{bmatrix} \phi(x^1,y^1) \\ \vdots \\ \phi(x^m,y^m) \end{bmatrix} \tag{1.7}$$

with $x := (x^1, \dots, x^m) \in \Re^{n_1} \times \cdots \times \Re^{n_m}$ and $y := (y^1, \dots, y^m) \in \Re^{n_1} \times \cdots \times \Re^{n_m}$, provided that the optimal value of (1.6) equals to zero. In general, there are two main difficulties in dealing with (1.6): One is the existence of expectation, which generally has no analytical expression. The other is the nonsmoothness of the SOC complementarity function ϕ. In Section 3, with the help of the Monte Carlo approximation techniques and some smoothing techniques given in [12], we suggest an approximation method for solving problem (1.6) and establish a comprehensive convergence analysis for the method. In Section 4, we investigate the exponential convergence of the approximation method and, in Section 5, we derive some results related to error bounds, which belong to an important area in optimization theory. Finally, in Section 5, we consider an application of the theoretical results in a practical engineering problem.

Throughout, we assume that $F(x, y, z, \xi)$ is twice continuously differentiable with respect to (x, y, z) and continuously integrable with respect to ξ over the compact set Ω. For a differentiable function $H : \Re^n \to \Re^m$ and a vector $x \in \Re^n$, $\nabla H(x)$ denotes the transposed Jacobian of H at x. Given a vector $x \in \Re^n$ and a convex set $X \subseteq \Re^n$, $\mathrm{dist}(x, X)$ denotes the distance from x to X and $\mathrm{int}X/\mathrm{cl}X/\mathrm{co}X/\mathrm{bd}X$ denote the interior/closure/convex hull/boundary of X respectively. For an $m \times n$ matrix $A := (a_{ij})$, $\|A\|_{\mathcal{F}}$ denotes its Frobenius norm, that is, $\|A\|_{\mathcal{F}} := (\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2)^{1/2}$. For a vector $x := (x_1, x_2) \in \Re \times \Re^{n-1}$, we denote its reflection of the x_1-axis by $\hat{x} := (x_1, -x_2)$ and by $R_x := \{tx \mid t \in \Re\}$ and $R_x^- := \{tx \mid t < 0\}$ respectively. Moreover, we denote by $\mathrm{gph}\Gamma$ the graph of a set-valued map Γ and by $\mathbb{B}_\varepsilon(x)$ and $\mathbb{B}$ the open ball centered at x with radius $\varepsilon > 0$ and the open unit ball respectively. In addition, I and O stand for the identity matrix and null matrix with suitable dimensions and $\mathcal{C}^\circ$ stands for the polar of a cone $\mathcal{C}$.

2 Preliminaries

In this section, we review some background materials that will be used later on.

2.1 Jordan algebras

For any $s = (s_1, s_2) \in \Re \times \Re^{\nu-1}$ and $t = (t_1, t_2) \in \Re \times \Re^{\nu-1}$, their Jordan product is defined by

$$s \circ t := (s^T t,\ t_1 s_2 + s_1 t_2).$$

Under the Jordan product, the identity element is $e := (1, 0, \dots, 0)^T \in \Re^\nu$ and for simplicity, we denote by $x^2 = x \circ x$. Moreover, if $s \in \mathcal{K}^\nu$, there exists a unique vector $s^{1/2} \in \mathcal{K}^\nu$ such that $(s^{1/2})^2 = s$.

For any vector $s = (s_1, s_2) \in \Re \times \Re^{\nu-1}$, its spectral factorization with respect to the second-order-cone $\mathcal{K}^\nu$ can be decomposed as

$$s = \lambda_1 u^1 + \lambda_2 u^2,$$

where λ_1 and λ_2 are the spectral values given by

$$\lambda_i := s_1 + (-1)^i \parallel s_2 \parallel, \quad i = 1, 2,$$

and u^1 and u^2 are the spectral vectors given by

$$u^i := \begin{cases} \frac{1}{2}\big(1, (-1)^i \frac{s_2}{\|s_2\|}\big) & \text{if } s_2 \neq 0, \\ \frac{1}{2}\big(1, (-1)^i w\big) & \text{if } s_2 = 0, \end{cases} \quad i = 1, 2$$

with w being an arbitrary unit vector in $\Re^{\nu-1}$. The projection function $[s]_+$ used in Section 1 can be calculated by

$$[s]_+ = [\lambda_1]_+ u^1 + [\lambda_2]_+ u^2,$$

where $[\lambda]_+ := \max\{\lambda, 0\}$ for a scalar $\lambda \in \Re$. In [12], it is showed that both $s^{1/2}$ and s^2 can also be rewritten by the spectral values and vectors of s. Thus, the functions defined in (1.4) and (1.5) can be represented as

$$\phi_{\mathrm{NR}}(s,t) = s - [s-t]_+ = s - ([\lambda_1]_+ u^1 + [\lambda_2]_+ u^2),$$

where $\{\lambda_1, \lambda_2\}$ and $\{u^1, u^2\}$ are given by, for $i = 1, 2$,

$$\lambda_i := s_1 - t_1 + (-1)^i \|s_2 - t_2\|, \tag{2.1}$$

$$u^i := \begin{cases} \frac{1}{2}\big(1, (-1)^i \frac{s_2 - t_2}{\|s_2 - t_2\|}\big) & \text{if } s_2 \neq t_2, \\ \frac{1}{2}\big(1, (-1)^i w\big) & \text{if } s_2 = t_2, \end{cases} \tag{2.2}$$

and

$$\phi_{\mathrm{FB}}(s,t) = s + t - (s^2 + t^2)^{1/2} = s + t - (\sqrt{\lambda_1} u^1 + \sqrt{\lambda_2} u^2),$$

where $\{\lambda_1, \lambda_2\}$ and $\{u^1, u^2\}$ are given by, for $i = 1, 2$,

$$\lambda_i := \|s\|^2 + \|t\|^2 + 2(-1)^i \|s_1 s_2 + t_1 t_2\|,$$

$$u^i := \begin{cases} \frac{1}{2}\big(1, (-1)^i \frac{s_1 s_2 + t_1 t_2}{\|s_1 s_2 + t_1 t_2\|}\big) & \text{if } s_1 s_2 + t_1 t_2 \neq 0, \\ \frac{1}{2}\big(1, (-1)^i w\big) & \text{if } s_1 s_2 + t_1 t_2 = 0, \end{cases}$$

with $w \in \Re^{\nu-1}$ being an arbitrary vector satisfying $\|w\| = 1$. Note that both ϕ_{NR} and ϕ_{FB} are locally Lipschitz continuous but not differentiable everywhere [12].

2.2 Variational analysis

This subsection contains some background materials on nonsmooth analysis that will be used particularly for the investigation on local error bounds in Section 4.

For a cone $\mathcal{C}$, we denote by $\mathcal{C}^\circ$ its *polar*. For a set-valued map $\Gamma : \Re^n \rightrightarrows \Re^n$, we denote the *Kuratowski-Painlevé upper* (*outer*) and *lower* (*inner*) *limit* by

$$\limsup_{x \to x_0} \Gamma(x) := \left\{ \xi \in \Re^n \,\middle|\, \exists \text{ sequences } x_k \to x_0 \text{ and } \xi_k \to \xi \text{ with } \xi_k \in \Gamma(x_k) \text{ for each } k \right\},$$

$$\liminf_{x \to x_0} \Gamma(x) := \left\{ \xi \in \Re^n \,\middle|\, \forall \text{ sequence } x_k \to x_0, \exists \xi_k \in \Gamma(x_k) \ (\forall k) \text{ such that } \xi_k \to \xi \right\}.$$

Definition 2.1 (Normal cones [18])**.** Let $X \subseteq \Re^n$ be nonempty and $x_0 \in \mathrm{cl}X$. The convex cone

$$\mathcal{N}_X^\pi(x_0) := \big\{ \xi \in \Re^n \,|\, \exists \sigma > 0 \text{ s.t. } \xi^T(x - x_0) \leq \sigma \|x - x_0\|^2, \ \forall x \in X \big\}$$

is called the *proximal normal cone* to X at x_0. The convex cone

$$\mathcal{N}_X^F(x_0) := \left\{ \xi \in \Re^n \,\middle|\, \limsup_{x \to x_0, x \in X} \frac{\langle \xi, x - x_0 \rangle}{\|x - x_0\|} \leq 0 \right\}$$

is called the *Fréchet* (*regular*) *normal cone* to X at x_0. The nonempty cone

$$\mathcal{N}_X(x_0) := \limsup_{x \to x_0} \mathcal{N}_X^F(x) = \limsup_{x \to x_0} \mathcal{N}_X^\pi(x)$$

is called the *limiting* (*Mordukhovich* or *basic*) *normal cone* to X at x_0. The *Clarke normal cone* is the closure of the convex hull of the limiting normal cone, i.e.,

$$\mathcal{N}_X^c(x_0) := \mathrm{clco}\mathcal{N}_X(x_0).$$

The limiting normal cone leads to the definition of coderivative of a set-valued map.

Definition 2.2 (Coderivative of a set-valued map [18])**.** Let $\Upsilon : \Re^n \rightrightarrows \Re^q$ be an arbitrary set-valued map (assigning to each $\tau \in \Re^n$ a set $\Upsilon(\tau) \subset \Re^q$, which may be empty) and $(\tilde{x}, \tilde{y}) \in \mathrm{clgph}\Upsilon$, where $\mathrm{gph}\Upsilon := \{(\tau, \upsilon)|\ \upsilon \in \Upsilon(\tau)\}$ denotes the graph of the set-valued map Υ. The set-valued map $D^*\Upsilon(\tilde{\tau}, \tilde{\upsilon})$ from $\Re^q$ into $\Re^n$ by

$$D^*\Upsilon(\tilde{\tau}, \tilde{\upsilon})(\zeta) := \{\varrho \in \Re^n|\ (\varrho, -\zeta) \in \mathcal{N}_{\mathrm{gph}\Upsilon}(\tilde{\tau}, \tilde{\upsilon})\}$$

is called the *coderivative* of Υ at the point $(\tilde{\tau}, \tilde{\upsilon})$. By convention, for $(\tilde{\tau}, \tilde{\upsilon}) \notin \mathrm{clgph}\Upsilon$, we define $D^*\Upsilon(\tilde{\tau}, \tilde{\upsilon})(\zeta) = \emptyset$.

To better understand this concept, we give the following example.

Example 2.1. Let $\mathcal{X}$ and $\mathcal{Y}$ be finite-dimensional Hilbert spaces and $\mathcal{A} : \mathcal{X} \to \mathcal{Y}$ be a linear operator. The graph of the set-valued map $\mathcal{A}$ is given by

$$\mathrm{gph}\mathcal{A} := \{(x, \mathcal{A}x)|\ x \in \mathcal{X}\},$$

which is a linear space, and the normal cone to $\mathrm{gph}\mathcal{A}$ at $(\overline{x}, \mathcal{A}\overline{x})$ can be written as

$$\begin{aligned} \mathcal{N}_{\mathrm{gph}\mathcal{A}}(\overline{x}, \mathcal{A}\overline{x}) &:= \{(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}|\ \langle (x^*, y^*), (z, \mathcal{A}z)\rangle = 0, \forall z \in \mathcal{X}\} \\ &= \{(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}|\ \langle x^*, z\rangle + \langle y^*, \mathcal{A}z)\rangle = 0, \forall z \in \mathcal{X}\} \\ &= \{(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}|\ \langle x^* + \mathcal{A}^*y^*, z\rangle = 0, \forall z \in \mathcal{X}\} \\ &= \{(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}|\ x^* + \mathcal{A}^*y^* = 0\}. \end{aligned}$$

Then, we have

$$D^*\mathcal{A}(\overline{x}, \mathcal{A}\overline{x})(y) = \{v|\ (v, -y) \in (x^*, y^*) \in \mathcal{N}_{\mathrm{gph}\mathcal{A}}(\overline{x}, \mathcal{A}\overline{x})\} = \{v|\ v - \mathcal{A}^*y^* = 0\} = \mathcal{A}^*y.$$

We now give some concepts for Lipschitz behavior of a set-valued map. Examples for these concepts can be found in [18].

The following concept for Lipschitz behavior is introduced by Aubin in [2].

Definition 2.3 (Pseudo-Lipschitz continuity)**.** A set-valued map $\Upsilon : \Re^n \rightrightarrows \Re^q$ is said to be *pseudo-Lipschitz continuous* around $(\tilde{\tau}, \tilde{\upsilon}) \in \mathrm{gph}\Upsilon$ if there exist a neighborhood $\mathbb{U}$ of $\tilde{\tau}$, a neighborhood $\mathbb{V}$ of $\tilde{\upsilon}$, and $\rho \geq 0$ such that

$$\Upsilon(\tau) \cap \mathbb{V} \subset \Upsilon(\tau^{'}) + \rho\|\tau^{'} - \tau\|\mathrm{cl}\mathbb{B}, \qquad \forall \tau^{'}, \tau \in \mathbb{U}.$$

On the other hand, the following upper-Lipschitz behavior is studied by Robinson in [23].

Definition 2.4 (Upper-Lipschitz continuity)**.** A set-valued map $\Upsilon : \Re^n \rightrightarrows \Re^q$ is said to be *upper-Lipschitz continuous* at $\tilde{\tau} \in \Re^n$ if there exist a neighborhood $\mathbb{U}$ of $\tilde{\tau}$ and $\rho \geq 0$ such that

$$\Upsilon(\tau) \subset \Upsilon(\tilde{\tau}) + \rho\|\tau - \tilde{\tau}\|\mathrm{cl}\mathbb{B}, \qquad \forall \tau \in \mathbb{U}.$$

Some conditions weaker than both the Aubin's pseudo-Lipschitz continuity and the Robinson's upper-Lipschitz continuity include the clamness condition [25] and the pseudo upper-Lipschitz continuity in [33].

Definition 2.5 (Calmness)**.** A set-valued map $\Upsilon : \Re^n \rightrightarrows \Re^q$ is called to be *calm* at $(\tilde{\tau}, \tilde{\upsilon}) \in \mathrm{gph}\Upsilon$ if there exist a neighborhood $\mathbb{U}$ of $\tilde{\tau}$, a neighborhood $\mathbb{V}$ of $\tilde{\upsilon}$, and $\rho \geq 0$ such that

$$\Upsilon(\tau) \cap \mathbb{V} \subset \Upsilon(\tilde{\tau}) + \rho\|\tilde{\tau} - \tau\|\mathrm{cl}\mathbb{B}, \quad \forall \tau \in \mathbb{U}.$$

3 Monte Carlo Approximation Method

In this section, we devote to developing approximation methods for solving the SSOCCP (1.1). Our focus is on the minimization approach (1.6). As mentioned in Section 1, there are two main difficulties in dealing with (1.6): One is the existence of expectation and the other is the nonsmoothness of the SOC complementarity functions. For the former, popular strategy is to employ the Monte Carlo sampling techniques to approximate the expectation and, for the latter, we may employ some smoothing techniques. But this is not always the case. For the two SOC complementarity functions introduced in Section 1, $\|\phi_{\rm FB}\|^2$ is actually a smooth function although $\phi_{\rm FB}$ is nonsmooth, while $\|\phi_{\rm NR}\|^2$ and $\phi_{\rm NR}$ are both nonsmooth functions.

In general, for an integrable function $\psi : \Omega \to \Re$, the Monte Carlo sampling estimate for $\mathbb{E}_\xi[\psi(\xi)]$ is obtained by taking independently and identically distributed (iid) random samples $\Omega_k := \{\xi^1, \ldots, \xi^{N_k}\}$ from Ω and letting $\mathbb{E}_\xi[\psi(\xi)] \approx \frac{1}{N_k}\sum_{\xi^i \in \Omega_k} \psi(\xi^i)$. We assume that $\{N_k\}$ tends to infinity as k increases. The strong law of large numbers guarantees that this procedure converges with probability one (abbreviated by "w.p.1" below), that is,

$$\lim_{k\to\infty} \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \psi(\xi^i) = \mathbb{E}_\xi[\psi(\xi)] \qquad \text{w.p.1.} \tag{3.1}$$

When ϕ is taken to be $\phi_{\rm NR}$ and $\phi_{\rm FB}$, (1.6) becomes the following two problems respectively:

$$\min_{(x,y,z)} \ \Theta_{\rm NR}(x,y,z) := \|\mathbb{E}_\xi[F(x,y,z,\xi)]\|^2 + \|\Phi_{\rm NR}(x,y)\|^2, \tag{3.2}$$

$$\min_{(x,y,z)} \ \Theta_{\rm FB}(x,y,z) := \|\mathbb{E}_\xi[F(x,y,z,\xi)]\|^2 + \|\Phi_{\rm FB}(x,y)\|^2. \tag{3.3}$$

Note that $\|\Phi_{\rm FB}\|^2$ is actually a smooth function although $\phi_{\rm FB}$ is nonsmooth, while $\|\Phi_{\rm NR}\|^2$ and $\phi_{\rm NR}$ are both nonsmooth functions. Here, we employ the following smoothing technique presented in [12] for $\phi_{\rm NR}$: Given a scalar $\mu > 0$, let

$$\phi^\mu_{\rm NR}(s,t) := s - \mu\left(g\left(\frac{\lambda_1}{\mu}\right)u^1 + g\left(\frac{\lambda_2}{\mu}\right)u^2\right),$$

where $g(a) := \frac{\sqrt{a^2+4}+a}{2}$, $\{\lambda_1, \lambda_2\}$ and $\{u^1, u^2\}$ are the same as in (2.1) and (2.2) respectively. It is shown in [12] that, for each $(s,t) \in \Re^{2\nu}$,

$$\lim_{\mu\to 0+} \phi^\mu_{\rm NR}(s,t) = \phi_{\rm NR}(s,t)$$

and $\phi^\mu_{\rm NR}$ is a smooth function with

$$\nabla\phi^\mu_{\rm NR}(s,t) = \begin{bmatrix} I - M_\mu(s,t) \\ M_\mu(s,t) \end{bmatrix},$$

where

$$M_\mu(s,t) := \begin{cases} a_\mu(s,t)I, & \text{if} \quad s_2 - t_2 = 0 \\ \begin{bmatrix} b_\mu(s,t) & \frac{d_\mu(s,t)(s_2-t_2)^T}{\|s_2-t_2\|} \\ \frac{d_\mu(s,t)(s_2-t_2)}{\|s_2-t_2\|} & \frac{(b_\mu(s,t)-c_\mu(s,t))(s_2-t_2)(s_2-t_2)^T}{\|s_2-t_2\|^2} + c_\mu(s,t)I \end{bmatrix}, & \text{if} \quad s_2 - t_2 \neq 0 \end{cases}$$

for $s=(s_1,s_2)\in\Re\times\Re^{\nu-1}$ and $t=(t_1,t_2)\in\Re\times\Re^{\nu-1}$ with

$$a_\mu(s,t) := g'(\tfrac{s_1-t_1}{\mu}), \tag{3.4}$$
$$b_\mu(s,t) := \tfrac{1}{2}(g'(\tfrac{\lambda_2}{\mu}) + g'(\tfrac{\lambda_1}{\mu})), \tag{3.5}$$
$$c_\mu(s,t) := \frac{g(\frac{\lambda_2}{\mu})-g(\frac{\lambda_1}{\mu})}{\frac{\lambda_2}{\mu}-\frac{\lambda_1}{\mu}}, \tag{3.6}$$
$$d_\mu(s,t) := \tfrac{1}{2}(g'(\tfrac{\lambda_2}{\mu}) - g'(\tfrac{\lambda_1}{\mu})). \tag{3.7}$$

In addition, from the proof of Proposition 5.1 in [12], it is not difficult to see that there exists a positive constant C such that

$$\|\phi^\mu_{\rm NR}(s,t) - \phi_{\rm NR}(s,t)\| \le C\mu \tag{3.8}$$

holds for each $(s,t)\in\Re^{2\nu}$.

Taking a smoothing parameter $\mu_k>0$ and iid samples $\Omega_k := \{\xi^1,\dots,\xi^{N_k}\}$ from Ω, the corresponding approximation problems of (3.2) and (3.3) are

$$\min_{(x,y,z)} \ \Theta^k_{\rm NR}(x,y,z) := \Big\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F(x,y,z,\xi^i)\Big\|^2 + \|\Phi^{\mu_k}_{\rm NR}(x,y)\|^2 \tag{3.9}$$

and

$$\min_{(x,y,z)} \ \Theta^k_{\rm FB}(x,y,z) := \Big\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F(x,y,z,\xi^i)\Big\|^2 + \|\Phi_{\rm FB}(x,y)\|^2 \tag{3.10}$$

respectively. Suppose that $\mu_k\to 0+$ as $k\to\infty$. We next study the limiting behavior of the above approximations.

Theorem 3.1. *Let (x^k,y^k,z^k) be a globally optimal solution of problem* (3.9) *or* (3.10) *for each k and $(\bar x,\bar y,\bar z)$ be an accumulation point of the sequence $\{(x^k,y^k,z^k)\}$. Then $(\bar x,\bar y,\bar z)$ is a globally optimal solution of problem* (3.2) *or* (3.3) *with probability one.*

Proof. We first prove the case of $\phi_{\rm NR}$. Without loss of generality, we assume that $\lim_{k\to\infty}(x^k,y^k,z^k) = (\bar x,\bar y,\bar z)$. Let B be a compact and convex set containing the whole sequence $\{(x^k,y^k,z^k)\}$. By the continuity of F and $\nabla_{(x,y,z)}F$ on the compact set $B\times\Omega$, there exists a constant $\bar C>0$ such that

$$\|F(x,y,z,\xi)\| \le \bar C, \ \ \|\nabla_{(x,y,z)}F(x,y,z,\xi)\|_{\mathcal F} \le \bar C, \quad \forall (x,y,z,\xi)\in B\times\Omega. \tag{3.11}$$

It is sufficient to show that, for each (x,y,z),

$$\big\|\mathbb{E}_\xi\left[F(\bar x,\bar y,\bar z,\xi)\right]\big\|^2 + \|\Phi_{\rm NR}(\bar x,\bar y)\|^2 \le \big\|\mathbb{E}_\xi\left[F(x,y,z,\xi)\right]\big\|^2 + \|\Phi_{\rm NR}(x,y)\|^2 \tag{3.12}$$

holds with probability one.

In fact, for each k, since (x^k,y^k,z^k) solves (3.9), we have

$$\begin{aligned}\Big\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F(x^k,y^k,z^k,\xi^i)\Big\|^2 + \|\Phi^{\mu_k}_{\rm NR}(x^k,y^k)\|^2 \quad \le \quad & \Big\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F(x,y,z,\xi^i)\Big\|^2 \\ & +\|\Phi^{\mu_k}_{\rm NR}(x,y)\|^2 \end{aligned}\tag{3.13}$$

for any (x,y,z). Note that

$$
\begin{aligned}
&\Big\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}F(x^k,y^k,z^k,\xi^i)-\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}F(\bar{x},\bar{y},\bar{z},\xi^i)\Big\|\\
\le\ & \frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\big\|F(x^k,y^k,z^k,\xi^i)-F(\bar{x},\bar{y},\bar{z},\xi^i)\big\|\\
\le\ & \frac{1}{N_k}\sum_{\xi^i\in\Omega_k}\int_0^1\|\nabla_{(x,y,z)}F(tx^k+(1-t)\bar{x},ty^k+(1-t)\bar{y},tz^k+(1-t)\bar{z},\xi^i)\|_{\mathcal{F}}\\
&\qquad\qquad\times\|(x^k,y^k,z^k)-(\bar{x},\bar{y},\bar{z})\|\,dt\\
\le\ & \bar{C}\|(x^k,y^k,z^k)-(\bar{x},\bar{y},\bar{z})\|\\
\to\ & 0 \qquad \text{as } k\to\infty,
\end{aligned}
$$

where the second inequality follows from the mean-value theorem and the third inequality follows from (3.11). We then have from (3.1) that

$$
\begin{aligned}
\lim_{k\to\infty}\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}F(x^k,y^k,z^k,\xi^i) &= \lim_{k\to\infty}\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}F(\bar{x},\bar{y},\bar{z},\xi^i)\\
&= \mathbb{E}_\xi\left[\,F(\bar{x},\bar{y},\bar{z},\xi)\,\right] \qquad \text{w.p.1}, \qquad (3.14)
\end{aligned}
$$

that is,

$$
\lim_{k\to\infty}\Big\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k}F(x^k,y^k,z^k,\xi^i)\Big\|^2=\big\|\mathbb{E}_\xi\left[\,F(\bar{x},\bar{y},\bar{z},\xi)\,\right]\big\|^2.
$$

On the other hand, since

$$
\begin{aligned}
&\big\|\phi_{\rm NR}^{\mu_k}(x^k,y^k)-\phi_{\rm NR}^{\mu_k}(\bar{x},\bar{y})\big\|\\
\le\ & \big\|\phi_{\rm NR}^{\mu_k}(x^k,y^k)-\phi_{\rm NR}(x^k,y^k)\big\|+\big\|\phi_{\rm NR}(x^k,y^k)-\phi_{\rm NR}(\bar{x},\bar{y})\big\|\\
&+\big\|\phi_{\rm NR}(\bar{x},\bar{y})-\phi_{\rm NR}^{\mu_k}(\bar{x},\bar{y})\big\|\\
\le\ & 2C\mu_k+\big\|\phi_{\rm NR}(x^k,y^k)-\phi_{\rm NR}(\bar{x},\bar{y})\big\|\\
\to\ & 0 \qquad \text{as } k\to\infty,
\end{aligned}
$$

where the second inequality follows from (3.8), we have

$$
\lim_{k\to\infty}\|\Phi_{\rm NR}^{\mu_k}(x^k,y^k)\|^2=\|\Phi_{\rm NR}(\bar{x},\bar{y})\|^2 \quad \text{and} \quad \lim_{k\to\infty}\|\Phi_{\rm NR}^{\mu_k}(x,y)\|^2=\|\Phi_{\rm NR}(x,y)\|^2. \qquad (3.15)
$$

Letting $k\to\infty$ in (3.13) and applying (3.1) again, we can get (3.12) immediately.

For the case of $\phi_{\rm FB}$, noting that the smoothing techniques are not required, we can show the conclusion in a similar but simpler way. This completes the proof. □

Since (3.2) and (3.3) are nonconvex optimization problems in general, we need to investigate the limiting behavior of its stationary points. To this end, the following definitions are useful.

Definition 3.2 ([10]). Let $H:\Re^p\to\Re^q$ be locally Lipschitz continuous. The *Clarke generalized gradient* of H at w is defined as

$$
\partial H(w):={\rm co}\Big\{\lim_{w'\to w,w'\in D_H}\nabla H(w')\Big\},
$$

where D_H denotes the set of points at which H is differentiable.

Definition 3.3 ([14]). Let $H : \Re^p \to \Re^q$ be locally Lipschitz continuous and $H^\mu : \Re^p \to \Re^q$ be a function such that H^μ is continuously differentiable everywhere for any $\mu > 0$ and $\lim_{\mu\to 0+} H^\mu(w) = H(w)$ for any $w \in \Re^p$. We say that H^μ satisfies the *Jacobian consistency* with H if

$$\lim_{\mu\to 0+} \operatorname{dist}\big(\nabla H^\mu(w), \partial H(w)\big) = 0$$

holds for any $w \in \Re^p$.

For simplicity, we denote by

$$\Psi_{\rm NR}(x,y) := \|\Phi_{\rm NR}(x,y)\|^2, \quad \Psi^\mu_{\rm NR}(x,y) := \|\Phi^\mu_{\rm NR}(x,y)\|^2,$$

and, for $x = (x^1,\dots,x^m) \in \Re^{n_1}\times\cdots\times\Re^{n_m}$ and $y = (y^1,\dots,y^m) \in \Re^{n_1}\times\cdots\times\Re^{n_m}$, we let

$$\Psi^i_{\rm NR}(x^i,y^i) := \|\phi_{\rm NR}(x^i,y^i)\|^2, \quad \Psi^{\mu,i}_{\rm NR}(x^i,y^i) := \|\phi^\mu_{\rm NR}(x^i,y^i)\|^2$$

for each i. Then we have

$$\partial\Psi_{\rm NR}(x,y) = \partial\Psi^1_{\rm NR}(x^1,y^1)\times\cdots\times\partial\Psi^m_{\rm NR}(x^m,y^m), \tag{3.16}$$

where $\partial\Psi^i_{\rm NR}(x^i,y^i) = 2\partial(\phi_{\rm NR}(x^i,y^i))\phi_{\rm NR}(x^i,y^i),\ i = 1,\dots,m$, and

$$\nabla\Psi^\mu_{\rm NR}(x,y) = \begin{bmatrix} \nabla\Psi^{\mu,1}_{\rm NR}(x^1,y^1) \\ \vdots \\ \nabla\Psi^{\mu,m}_{\rm NR}(x^m,y^m) \end{bmatrix}. \tag{3.17}$$

Theorem 3.4. *Suppose that (x^k,y^k,z^k) is a stationary point of problem (3.9) or (3.10) for each k and $(\bar x,\bar y,\bar z)$ is an accumulation point of the sequence $\{(x^k,y^k,z^k)\}$. Then $(\bar x,\bar y,\bar z)$ is a stationary point of problem (3.2) or (3.3) with probability one.*

Proof. First of all, we consider the case of (3.9). Without loss of generality, we may assume $\lim_{k\to\infty}(x^k,y^k,z^k) = (\bar x,\bar y,\bar z)$. Let B and $\bar C > 0$ be the same as those in the proof of Theorem 3.1. For each k, since (x^k,y^k,z^k) is stationary to problem (3.9), we have

$$\frac{2}{N_k^2}\sum_{\xi^i\in\Omega_k}\nabla_{(x,y,z)}F(x^k,y^k,z^k,\xi^i)\sum_{\xi^i\in\Omega_k}F(x^k,y^k,z^k,\xi^i) + \begin{bmatrix}\nabla\Psi^{\mu_k}_{\rm NR}(x^k,y^k)\\ 0\end{bmatrix} = 0. \tag{3.18}$$

Since F is twice continuously differentiable, and F, $\nabla_{(x,y,z)}F$ and $\nabla^2_{(x,y,z)}F$ are continuous on the compact set $B\times\Omega$, in a similar way to the proof of Theorem 3.1, we have

$$\begin{aligned} & \lim_{k\to\infty}\frac{2}{N_k^2}\sum_{\xi^i\in\Omega_k}\nabla_{(x,y,z)}F(x^k,y^k,z^k,\xi^i)\sum_{\xi^i\in\Omega_k}F(x^k,y^k,z^k,\xi^i) \\ = \ & 2\mathbb{E}_\xi\left[\nabla_{(x,y,z)}F(\bar x,\bar y,\bar z,\xi)\right]\mathbb{E}_\xi\left[F(\bar x,\bar y,\bar z,\xi)\right] \\ = \ & 2\nabla_{(x,y,z)}(\mathbb{E}_\xi\left[F(\bar x,\bar y,\bar z,\xi)\right])\mathbb{E}_\xi\left[F(\bar x,\bar y,\bar z,\xi)\right] \\ = \ & \nabla_{(x,y,z)}(\|\mathbb{E}_\xi\left[F(\bar x,\bar y,\bar z,\xi)\right]\|^2) \end{aligned} \tag{3.19}$$

with probability one, where the second equality follows from by Theorem 16.8 of [20] due to the continuity of $\nabla_{(x,y,z)}F$ on the compact set $B\times\Omega$. We next show

$$\lim_{k\to\infty}\operatorname{dist}\big(\nabla\Psi^{\mu_k}_{\rm NR}(x^k,y^k), \partial\Psi_{\rm NR}(\bar x,\bar y)\big) = 0.$$

Denote by $x^k := (x^{k,1}, \ldots, x^{k,m}) \in \Re^{n_1} \times \cdots \times \Re^{n_m}$, $y^k := (y^{k,1}, \ldots, y^{k,m}) \in \Re^{n_1} \times \cdots \times \Re^{n_m}$ for each k and by $\bar{x} := (\bar{x}^1, \ldots, \bar{x}^m) \in \Re^{n_1} \times \cdots \times \Re^{n_m}$, $\bar{y} := (\bar{y}^1, \ldots, \bar{y}^m) \in \Re^{n_1} \times \cdots \times \Re^{n_m}$. From (3.16) and (3.17), it is sufficient to show

$$\lim_{k\to\infty} \operatorname{dist}\big(\nabla \Psi_{\rm NR}^{\mu_k,i}(x^{k,i}, y^{k,i}), \partial \Psi_{\rm NR}^i(\bar{x}^i, \bar{y}^i)\big) = 0 \tag{3.20}$$

for each i. In what follows, we let i be fixed and $\lambda_j^{k,i} := x_1^{k,i} - y_1^{k,i} + (-1)^j \|x_2^{k,i} - y_2^{k,i}\|$, $\bar{\lambda}_j^i := \bar{x}_1^i - \bar{y}_1^i + (-1)^j \|\bar{x}_2^i - \bar{y}_2^i\|$ for $j = 1, 2$. We consider six cases:

(I) Suppose that $\bar{\lambda}_1^i > 0$, that is, $\bar{x}_1^i - \bar{y}_1^i > \|\bar{x}_2^i - \bar{y}_2^i\|$. By Lemma 5.1 in [6], we have

$$\partial \Psi_{\rm NR}^i(\bar{x}^i, \bar{y}^i) = \left\{ 2 \begin{bmatrix} O \\ I \end{bmatrix} \phi_{\rm NR}(\bar{x}^i, \bar{y}^i) \right\}.$$

If $x_2^{k,i} - y_2^{k,i} = 0$ holds for infinitely many k, noting that $\lim_{k\to\infty} \frac{x_1^{k,i} - y_1^{k,i}}{\mu_k} = +\infty$ and taking a subsequence if necessary, we have from (3.4) that

$$\lim_{k\to\infty} M_{\mu_k}(x^{k,i}, y^{k,i}) = \lim_{k\to\infty} a_{\mu_k}(x^{k,i}, y^{k,i}) I = \lim_{k\to\infty} g'(\frac{x_1^{k,i} - y_1^{k,i}}{\mu_k}) I = I.$$

If $x_2^{k,i} - y_2^{k,i} \neq 0$ holds for every k sufficiently large, noting that $\lim_{k\to\infty} \frac{\lambda_1^{k,i}}{\mu_k} = +\infty$ and $\lim_{k\to\infty} \frac{\lambda_2^{k,i}}{\mu_k} = +\infty$, we obtain by calculating (3.5)–(3.7) that

$$\lim_{k\to\infty} b_{\mu_k}(x^{k,i}, y^{k,i}) = 1, \quad \lim_{k\to\infty} c_{\mu_k}(x^{k,i}, y^{k,i}) = 1, \quad \lim_{k\to\infty} d_{\mu_k}(x^{k,i}, y^{k,i}) = 0,$$

which means

$$\lim_{k\to\infty} M_{\mu_k}(x^{k,i}, y^{k,i}) = I.$$

Therefore, any accumulation point of $\{M_{\mu_k}(x^{k,i}, y^{k,i})\}$ must be I. On the other hand, we have $\lim_{k\to\infty} \phi_{\rm NR}^{\mu_k}(x^{k,i}, y^{k,i}) = \phi_{\rm NR}(\bar{x}^i, \bar{y}^i)$ from the proof of Theorem 3.1 by (3.15). Hence, we obtain (3.20) immediately.

(II) Suppose that $\bar{\lambda}_2^i < 0$, that is, $\bar{x}_1^i - \bar{y}_1^i < -\|\bar{x}_2^i - \bar{y}_2^i\|$. By Lemma 5.1 in [6], we have

$$\partial \Psi_{\rm NR}^i(\bar{x}^i, \bar{y}^i) = \left\{ 2 \begin{bmatrix} I \\ O \end{bmatrix} \phi_{\rm NR}(\bar{x}^i, \bar{y}^i) \right\}.$$

In a similar way to that in (I), we can show that any accumulation point of $\{M_{\mu_k}(x^{k,i}, y^{k,i})\}$ must be O and hence we have (3.20).

(III) Suppose that $\bar{\lambda}_1^i < 0$ and $\bar{\lambda}_2^i > 0$, that is, $|\bar{x}_1^i - \bar{y}_1^i| < \|\bar{x}_2^i - \bar{y}_2^i\|$. By Lemma 5.1 in [6], we have

$$\partial \Psi_{\rm NR}^i(\bar{x}^i, \bar{y}^i) = \left\{ 2 \begin{bmatrix} I - Z \\ Z \end{bmatrix} \phi_{\rm NR}(\bar{x}^i, \bar{y}^i) \right\},$$

where

$$Z := \frac{1}{2} \begin{bmatrix} 1 & \frac{(\bar{x}_2^i - \bar{y}_2^i)^T}{\|\bar{x}_2^i - \bar{y}_2^i\|} \\ \frac{\bar{x}_2^i - \bar{y}_2^i}{\|\bar{x}_2^i - \bar{y}_2^i\|} & I + \frac{\bar{x}_1^i - \bar{y}_1^i}{\|\bar{x}_2^i - \bar{y}_2^i\|}\left(I - \frac{(\bar{x}_2^i - \bar{y}_2^i)(\bar{x}_2^i - \bar{y}_2^i)^T}{\|\bar{x}_2^i - \bar{y}_2^i\|^2}\right) \end{bmatrix}.$$

Similarly to that in (I), we can show that any accumulation point of $\{M_{\mu_k}(x^{k,i}, y^{k,i})\}$ must be Z and hence we have (3.20).

(IV) Suppose that $\bar{\lambda}_1^i = 0$ and $\bar{\lambda}_2^i > 0$, that is, $\bar{x}_1^i - \bar{y}_1^i = \|\bar{x}_2^i - \bar{y}_2^i\| \neq 0$. By Lemma 5.1 in [6], we have

$$\partial \Psi_{\rm NR}^i(\bar{x}^i, \bar{y}^i) = \left\{ 2 \begin{bmatrix} I - V \\ V \end{bmatrix} \phi_{\rm NR}(\bar{x}^i, \bar{y}^i) \,\middle|\, V \in {\rm co}(I, Z) \right\},$$

where

$$Z := \frac{1}{2} \begin{bmatrix} 1 & \frac{(\bar{x}_2^i - \bar{y}_2^i)^T}{\|\bar{x}_2^i - \bar{y}_2^i\|} \\ \frac{\bar{x}_2^i - \bar{y}_2^i}{\|\bar{x}_2^i - \bar{y}_2^i\|} & 2I - \frac{(\bar{x}_2^i - \bar{y}_2^i)(\bar{x}_2^i - \bar{y}_2^i)^T}{\|\bar{x}_2^i - \bar{y}_2^i\|^2} \end{bmatrix}.$$

Note that $\lim_{k\to\infty} \frac{\lambda_2^{k,i}}{\mu_k} = +\infty$. Taking a subsequence if necessary, we assume that $\lim_{k\to\infty} \frac{\lambda_1^{k,i}}{\mu_k} = \alpha \in \Re \cup \{\pm\infty\}$. If $\lim_{k\to\infty} \frac{\lambda_1^{k,i}}{\mu_k} = +\infty$, we have

$$\lim_{k\to\infty} b_{\mu_k}(x^{k,i}, y^{k,i}) = 1, \quad \lim_{k\to\infty} c_{\mu_k}(x^{k,i}, y^{k,i}) = 1, \quad \lim_{k\to\infty} d_{\mu_k}(x^{k,i}, y^{k,i}) = 0,$$

which implies

$$\lim_{k\to\infty} M_{\mu_k}(x^{k,i}, y^{k,i}) = I.$$

If $\lim_{k\to\infty} \frac{\lambda_1^{k,i}}{\mu_k} = \alpha \in \Re$, by letting $\gamma := g'(\alpha) \in (0,1)$, we have

$$\lim_{k\to\infty} b_{\mu_k}(x^{k,i}, y^{k,i}) = \frac{1+\gamma}{2}, \quad \lim_{k\to\infty} c_{\mu_k}(x^{k,i}, y^{k,i}) = 1, \quad \lim_{k\to\infty} d_{\mu_k}(x^{k,i}, y^{k,i}) = \frac{1-\gamma}{2},$$

which implies

$$\lim_{k\to\infty} M_{\mu_k}(x^{k,i}, y^{k,i}) = \begin{bmatrix} \frac{1+\gamma}{2} & \frac{1-\gamma}{2} \frac{(\bar{x}_2^i - \bar{y}_2^i)^T}{\|\bar{x}_2^i - \bar{y}_2^i\|} \\ \frac{1-\gamma}{2} \frac{\bar{x}_2^i - \bar{y}_2^i}{\|\bar{x}_2^i - \bar{y}_2^i\|} & I + (\frac{\gamma-1}{2}) \frac{(\bar{x}_2^i - \bar{y}_2^i)(\bar{x}_2^i - \bar{y}_2^i)^T}{\|\bar{x}_2^i - \bar{y}_2^i\|^2} \end{bmatrix} = \gamma I + (1-\gamma) Z,$$

and hence the limit of $\{M_{\mu_k}(x^{k,i}, y^{k,i})\}$ is a convex combination of I and Z. If $\lim_{k\to\infty} \frac{\lambda_1^{k,i}}{\mu_k} = -\infty$, we have

$$\lim_{k\to\infty} b_{\mu_k}(x^{k,i}, y^{k,i}) = \frac{1}{2}, \quad \lim_{k\to\infty} c_{\mu_k}(x^{k,i}, y^{k,i}) = 1, \quad \lim_{k\to\infty} d_{\mu_k}(x^{k,i}, y^{k,i}) = \frac{1}{2},$$

which implies

$$\lim_{k\to\infty} M_{\mu_k}(x^{k,i}, y^{k,i}) = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \frac{(\bar{x}_2^i - \bar{y}_2^i)^T}{\|\bar{x}_2^i - \bar{y}_2^i\|} \\ \frac{1}{2} \frac{\bar{x}_2^i - \bar{y}_2^i}{\|\bar{x}_2^i - \bar{y}_2^i\|} & I - \frac{1}{2} \frac{(\bar{x}_2^i - \bar{y}_2^i)(\bar{x}_2^i - \bar{y}_2^i)^T}{\|\bar{x}_2^i - \bar{y}_2^i\|^2} \end{bmatrix}.$$

Therefore, in all cases, any limit of $\{M_{\mu_k}(x^{k,i}, y^{k,i})\}$ must belong to ${\rm co}(I, Z)$ and hence (3.20) holds.

(V) Suppose that $\bar{\lambda}_1^i < 0$ and $\bar{\lambda}_2^i = 0$, that is, $-\bar{x}_1^i + \bar{y}_1^i = \|\bar{x}_2^i - \bar{y}_2^i\| \neq 0$. By Lemma 5.1 in [6], we have

$$\partial \Psi_{\rm NR}^i(\bar{x}^i, \bar{y}^i) = \left\{ 2 \begin{bmatrix} I - V \\ V \end{bmatrix} \phi_{\rm NR}(\bar{x}^i, \bar{y}^i) \,\middle|\, V \in {\rm co}(O, Z) \right\},$$

where

$$Z := \frac{1}{2} \begin{bmatrix} 1 & \frac{(\bar{x}_2^i - \bar{y}_2^i)^T}{\|\bar{x}_2^i - \bar{y}_2^i\|} \\ \frac{\bar{x}_2^i - \bar{y}_2^i}{\|\bar{x}_2^i - \bar{y}_2^i\|} & \frac{(\bar{x}_2^i - \bar{y}_2^i)(\bar{x}_2^i - \bar{y}_2^i)^T}{\|\bar{x}_2^i - \bar{y}_2^i\|^2} \end{bmatrix}.$$

Similarly as in (IV), we can show that any accumulation point of $\{M_{\mu_k}(x^{k,i}, y^{k,i})\}$ must belong to ${\rm co}(O, Z)$ and hence (3.20) holds.

(VI) Suppose that $\bar{\lambda}_1^i = 0$ and $\bar{\lambda}_2^i = 0$, that is, $\bar{x}^i - \bar{y}^i = 0$. By Lemma 5.1 in [6], we have

$$\partial \Psi_{\rm NR}^i(\bar{x}^i, \bar{y}^i) = \left\{ 2 \begin{bmatrix} I - V \\ V \end{bmatrix} \phi_{\rm NR}(\bar{x}^i, \bar{y}^i) \,\middle|\, V \in {\rm co}\{O, I, S\} \right\},$$

where

$$S := \left\{ \frac{1}{2} \begin{bmatrix} 1 & w^T \\ w & (1+\beta)I - \beta w w^T \end{bmatrix} \,\middle|\, \beta \in [-1, 1],\ \|w\| = 1 \right\}.$$

We first consider the case where $x_2^{k,i} - y_2^{k,i} = 0$ for infinitely many k. If $\lim_{k\to\infty} \frac{x_1^{k,i} - y_1^{k,i}}{\mu_k} = +\infty$, from (3.4), we have

$$\lim_{k\to\infty} M_{\mu_k}(x^{k,i}, y^{k,i}) = \lim_{k\to\infty} a_{\mu_k}(x^{k,i}, y^{k,i}) I = \lim_{k\to\infty} g'(\frac{x_1^{k,i} - y_1^{k,i}}{\mu_k}) I = I.$$

If $\lim_{k\to\infty} \frac{x_1^{k,i} - y_1^{k,i}}{\mu_k} = \delta \in \Re$, from (3.4), we have

$$\lim_{k\to\infty} M_{\mu_k}(x^{k,i}, y^{k,i}) = \lim_{k\to\infty} a_{\mu_k}(x^{k,i}, y^{k,i}) I = \lim_{k\to\infty} g'(\frac{x_1^{k,i} - y_1^{k,i}}{\mu_k}) I = \tau I,$$

where $\tau := g'(\delta) \in (0, 1)$. If $\lim_{k\to\infty} \frac{x_1^{k,i} - y_1^{k,i}}{\mu_k} = -\infty$, from (3.4), we have

$$\lim_{k\to\infty} M_{\mu_k}(x^{k,i}, y^{k,i}) = \lim_{k\to\infty} a_{\mu_k}(x^{k,i}, y^{k,i}) I = \lim_{k\to\infty} g'(\frac{x_1^{k,i} - y_1^{k,i}}{\mu_k}) I = O.$$

In conclusion, any limit of $\{M_{\mu_k}(x^{k,i}, y^{k,i})\}$ must be a convex combination of O, I, and S.

We next consider the case where $x_2^{k,i} - y_2^{k,i} \neq 0$ holds for every k sufficiently large. From the mean-value theorem, there exists $\lambda^{k,i} \in [\lambda_1^{k,i}, \lambda_2^{k,i}]$ such that

$$g(\frac{\lambda_2^{k,i}}{\mu_k}) - g(\frac{\lambda_1^{k,i}}{\mu_k}) = g'(\frac{\lambda^{k,i}}{\mu_k})(\frac{\lambda_2^{k,i}}{\mu_k} - \frac{\lambda_1^{k,i}}{\mu_k}).$$

If $\lim_{k\to\infty} \frac{\lambda_1^{k,i}}{\mu_k} = +\infty$ and $\lim_{k\to\infty} \frac{\lambda_2^{k,i}}{\mu_k} = +\infty$, it is obvious that $\lim_{k\to\infty} \frac{\lambda^{k,i}}{\mu_k} = +\infty$. By calculating (3.5)–(3.7), we obtain

$$\lim_{k\to\infty} b_{\mu_k}(x^{k,i}, y^{k,i}) = 1, \quad \lim_{k\to\infty} c_{\mu_k}(x^{k,i}, y^{k,i}) = \lim_{k\to\infty} g'(\frac{\lambda^{k,i}}{\mu_k}) = 1, \quad \lim_{k\to\infty} d_{\mu_k}(x^{k,i}, y^{k,i}) = 0,$$

which means

$$\lim_{k\to\infty} M_{\mu_k}(x^{k,i}, y^{k,i}) = I.$$

If $\lim_{k\to\infty} \frac{\lambda_1^{k,i}}{\mu_k} = \alpha \in \Re$ and $\lim_{k\to\infty} \frac{\lambda_2^{k,i}}{\mu_k} = +\infty$, we have

$$\lim_{k\to\infty} b_{\mu_k}(x^{k,i}, y^{k,i}) = \frac{1+\gamma_1}{2}, \quad \lim_{k\to\infty} c_{\mu_k}(x^{k,i}, y^{k,i}) = 1, \quad \lim_{k\to\infty} d_{\mu_k}(x^{k,i}, y^{k,i}) = \frac{1-\gamma_1}{2},$$

where $\gamma_1 := g'(\alpha) \in (0, 1)$, and then

$$\begin{aligned} \lim_{k\to\infty} M_{\mu_k}(x^{k,i}, y^{k,i}) &= \begin{bmatrix} \frac{1+\gamma_1}{2} & \frac{1-\gamma_1}{2} w^T \\ \frac{1-\gamma_1}{2} w & I + \frac{\gamma_1 - 1}{2} w w^T \end{bmatrix} \\ &= \gamma_1 I + (1-\gamma_1) \frac{1}{2} \begin{bmatrix} 1 & w^T \\ w & (1+\beta_1) I - \beta_1 w w^T \end{bmatrix}, \end{aligned}$$

where $\beta_1 := 1$. This indicates that the limit of $\{M_{\mu_k}(x^{k,i}, y^{k,i})\}$ is a convex combination of O, I, S. Actually, the above conclusion can be shown in a similar way for the following four cases:
(i) $\lim_{k\to\infty} \frac{\lambda_1^{k,i}}{\mu_k} = -\infty$ and $\lim_{k\to\infty} \frac{\lambda_2^{k,i}}{\mu_k} = +\infty$;
(ii) $\lim_{k\to\infty} \frac{\lambda_1^{k,i}}{\mu_k} = \alpha \in \Re$ and $\lim_{k\to\infty} \frac{\lambda_2^{k,i}}{\mu_k} = \eta \in \Re$ with $\alpha \le \eta$;
(iii) $\lim_{k\to\infty} \frac{\lambda_1^{k,i}}{\mu_k} = -\infty$ and $\lim_{k\to\infty} \frac{\lambda_2^{k,i}}{\mu_k} = \eta \in \Re$;
(iv) $\lim_{k\to\infty} \frac{\lambda_1^{k,i}}{\mu_k} = -\infty$ and $\lim_{k\to\infty} \frac{\lambda_2^{k,i}}{\mu_k} = -\infty$.
Therefore, we can obtain (3.20) immediately.

As a result, we have (3.20) in all cases. Letting $k \to \infty$ in (3.18), we have from (3.19)–(3.20) that

$$0 \in \nabla \mathbb{E}_\xi \left[\, \|F(\bar{x}, \bar{y}, \bar{z}, \xi)\|^2 \,\right] + \partial \Psi_{\rm NR}(\bar{x}, \bar{y}) \times \{0\} \qquad \text{w.p.1},$$

which means that $(\bar{x}, \bar{y}, \bar{z})$ is a stationary point of (3.2) with probability one.

Now we consider the case of (3.10). For each k, since (x^k, y^k, z^k) is stationary to problem (3.10), we have

$$\frac{2}{N_k^2} \sum_{\xi^i \in \Omega_k} \nabla_{(x,y,z)} F(x^k, y^k, z^k, \xi^i) \sum_{\xi^i \in \Omega_k} F(x^k, y^k, z^k, \xi^i) + \begin{bmatrix} \nabla \Psi_{\rm FB}(x^k, y^k) \\ 0 \end{bmatrix} = 0. \qquad (3.21)$$

By Proposition 2 of [7], $\Psi_{\rm FB}$ is a smooth function, that is, $\nabla \Psi_{\rm FB}$ is continuous everywhere. Letting $k \to \infty$ in (3.21), we have from (3.19) that

$$\nabla \mathbb{E}_\xi \left[\, \|F(\bar{x}, \bar{y}, \bar{z}, \xi)\|^2 \,\right] + \nabla \Psi_{\rm FB}(\bar{x}, \bar{y}) = 0 \qquad \text{w.p.1},$$

which means that $(\bar{x}, \bar{y}, \bar{z})$ is a stationary point of (3.3) with probability one. □

4 Exponential Convergence Rate

In this section, we discuss the exponential convergence rate of the optimal solutions of the approximation problem (3.9) or (3.10). To this end, we need the following lemma.

Lemma 4.1 ([27])**.** *Let $\mathcal{X}$ be a compact set and $h : \mathcal{X} \times \Omega \to \Re$ be integrable everywhere. Suppose that the following conditions hold:*

(i) *For every $x \in \mathcal{X}$, the moment generating function $\mathbb{E}_\xi[e^{t(h(x,\xi) - \mathbb{E}_\xi[h(x,\xi)])}]$ is finite-valued for all t in a neighbourhood of zero.*

(ii) *There exist a measurable function $\kappa : \Omega \to \Re_+$ and a constant $\gamma > 0$ such that, for all $\xi \in \Omega$ and all $x', x \in \mathcal{X}$, $|h(x', \xi) - h(x, \xi)| \le \kappa(\xi)\|x' - x\|^\gamma$ holds.*

(iii) *The moment generating function $\mathbb{E}_\xi[e^{t\kappa(\xi)}]$ is finite-valued for all t in a neighbourhood of zero.*

Then, for every $\varepsilon > 0$, there exist positive constants $D(\varepsilon)$ and $\beta(\varepsilon)$, independent of N_k, such that

$$\text{Prob}\Big\{ \sup_{x \in \mathcal{X}} \Big| \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} h(x, \xi^i) - \mathbb{E}_\xi[h(x, \xi)] \Big| \ge \varepsilon \Big\} \le D(\varepsilon) e^{-N_k \beta(\varepsilon)}.$$

Applying this lemma, we can obtain the following result related to exponential convergence rate.

Theorem 4.2. *Let (x^k, y^k, z^k) be an optimal solution of (3.9) or (3.10) for each k and $(\bar{x}, \bar{y}, \bar{z})$ be an accumulation point of the sequence $\{(x^k, y^k, z^k)\}$. Then, for every $\varepsilon > 0$, there exist positive constants $D(\varepsilon)$ and $\beta(\varepsilon)$, independent of N_k, such that*

$$\text{Prob}\Big\{|\Theta_{\text{NR}}^k(x^k, y^k, z^k) - \Theta_{\text{NR}}(\bar{x}, \bar{y}, \bar{z})| \geq \varepsilon\Big\} \leq D(\varepsilon)e^{-N_k\beta(\varepsilon)} \tag{4.1}$$

or

$$\text{Prob}\Big\{|\Theta_{\text{FB}}^k(x^k, y^k, z^k) - \Theta_{\text{FB}}(\bar{x}, \bar{y}, \bar{z})| \geq \varepsilon\Big\} \leq D(\varepsilon)e^{-N_k\beta(\varepsilon)}. \tag{4.2}$$

Proof. Without loss of generality, we assume that $\{(x^k, y^k, z^k)\}$ itself converges to $(\bar{x}, \bar{y}, \bar{z})$. Let B be a compact set that contains the whole sequence $\{(x^k, y^k, z^k)\}$.

(1) Consider the case for (4.2). We first show that, for every $\varepsilon > 0$, there exist positive constants $D(\varepsilon)$ and $\beta(\varepsilon)$, independent of N_k, such that

$$\text{Prob}\Big\{\sup_{(x,y,z)\in B} |\Theta_{\text{FB}}^k(x, y, z) - \Theta_{\text{FB}}(x, y, z)| \geq \varepsilon\Big\} \leq D(\varepsilon)e^{-N_k\beta(\varepsilon)}. \tag{4.3}$$

In fact, from (3.3) and (3.10), it is sufficient to show

$$\text{Prob}\Big\{\sup_{(x,y,z)\in B} \Big|\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F(x, y, z, \xi^i)\|^2 - \|\mathbb{E}_\xi[F(x, y, z, \xi)]\|^2\Big| \geq \varepsilon\Big\} \leq D(\varepsilon)e^{-N_k\beta(\varepsilon)}. \tag{4.4}$$

Noting that F can be regarded as an $(n + l)$-dimensional vector, we obtain

$$\begin{aligned}
&\Big|\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F(x, y, z, \xi^i)\|^2 - \|\mathbb{E}_\xi[F(x, y, z, \xi)]\|^2\Big| \\
= \;& \Big|\sum_{i=1}^{n+l}\big(\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_i(x, y, z, \xi^i)\big)^2 - \sum_{i=1}^{n+l}\big(\mathbb{E}_\xi[F_i(x, y, z, \xi)]\big)^2\Big| \\
\leq \;& \sum_{i=1}^{n+l}\Big|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_i(x, y, z, \xi^i) + \mathbb{E}_\xi[F_i(x, y, z, \xi)]\Big| \cdot \\
& \Big|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_i(x, y, z, \xi^i) - \mathbb{E}_\xi[F_i(x, y, z, \xi)]\Big|.
\end{aligned}$$

Since $F(x, y, z, \xi)$ is continuously differentiable with respect to (x, y, z) and continuously integrable with respect to ξ over the compact set Ω, there exists a constant $M > 0$ such that, for all $i \in \{1, \ldots, n + l\}$,

$$\Big|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_i(x, y, z, \xi^i) + \mathbb{E}_\xi[F_i(x, y, z, \xi)]\Big| \leq M, \quad \forall (x, y, z, \xi) \in B \times \Omega.$$

Let

$$\Big|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_j(x, y, z, \xi^i) - \mathbb{E}_\xi[F_j(x, y, z, \xi)]\Big|$$

$$:= \max\Big\{\big|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_i(x,y,z,\xi^i) - \mathbb{E}_\xi[F_i(x,y,z,\xi)]\big|, \quad 1\le i\le n+l\Big\}.$$

We have

$$\begin{aligned}
&\Big|\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F(x,y,z,\xi^i)\|^2 - \|\mathbb{E}_\xi[F(x,y,z,\xi)]\|^2\Big| \\
\le\ & \sum_{i=1}^{n+l} M\Big|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_i(x,y,z,\xi^i) - \mathbb{E}_\xi[F_i(x,y,z,\xi)]\Big| \\
=\ & (n+l)M\Big|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_j(x,y,z,\xi^i) - \mathbb{E}_\xi[F_j(x,y,z,\xi)]\Big|,
\end{aligned}$$

which means

$$\begin{aligned}
&\sup_{(x,y,z)\in B}\big|\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F(x,y,z,\xi^i)\|^2 - \|\mathbb{E}_\xi[F(x,y,z,\xi)]\|^2\big| \ge \varepsilon \\
\Longrightarrow &\sup_{(x,y,z)\in B}\big|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_j(x,y,z,\xi^i) - \mathbb{E}_\xi[F_j(x,y,z,\xi)]\big| \ge \frac{\varepsilon}{(n+l)M}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\mathrm{Prob}\Big\{\sup_{(x,y,z)\in B}\big|\|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F(x,y,z,\xi^i)\|^2 - \|\mathbb{E}_\xi[F(x,y,z,\xi)]\|^2\big| \ge \varepsilon\Big\} \\
&\le \mathrm{Prob}\Big\{\sup_{(x,y,z)\in B}\big|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_j(x,y,z,\xi^i) - \mathbb{E}_\xi[F_j(x,y,z,\xi)]\big| \ge \frac{\varepsilon}{(n+l)M}\Big\}. \quad (4.5)
\end{aligned}$$

Let $\mathcal{X} := B$ and $h(x,y,z,\xi) := F_j(x,y,z,\xi)$. Since both B and Ω are compact sets, by the continuous differentiability assumption on $F(x,y,z,\xi)$ given in Section 1, it is not difficult to verify that $\mathcal{X}$ and the function $h(x,y,z,\xi)$ satisfy the conditions given in Lemma 4.1. Hence, for every $\varepsilon > 0$, there exist positive constants $\overline{D}(\frac{\varepsilon}{(n+l)M})$ and $\overline{\beta}(\frac{\varepsilon}{(n+l)M})$, independent of N_k, such that

$$\begin{aligned}
\mathrm{Prob}\Big\{&\sup_{(x,y,z)\in B}\big|\frac{1}{N_k}\sum_{\xi^i\in\Omega_k} F_j(x,y,z,\xi^i) - \mathbb{E}_\xi[F_j(x,y,z,\xi)]\big| \\
&\ge \frac{\varepsilon}{(n+l)M}\Big\} \le \overline{D}(\frac{\varepsilon}{(n+l)M})e^{-N_k\overline{\beta}(\frac{\varepsilon}{(n+l)M})}.
\end{aligned}$$

Noting that $(n+l)M$ is just a constant for every ε, we can use $D(\varepsilon)$ and $\beta(\varepsilon)$ to denote $\overline{D}(\frac{\varepsilon}{(n+l)M})$ and $\overline{\beta}(\frac{\varepsilon}{(n+l)M})$ respectively. Thus, combining with (4.5), we can obtain (4.4) easily.

Since (x^k,y^k,z^k) is an optimal solution of (3.10) for each k, by Theorem 3.1, $(\bar{x},\bar{y},\bar{z})$ must be an optimal solution of (3.3). It then follows that

$$\theta^k_{\mathrm{FB}}(x^k,y^k,z^k) \le \theta^k_{\mathrm{FB}}(\bar{x},\bar{y},\bar{z}), \qquad \theta_{\mathrm{FB}}(\bar{x},\bar{y},\bar{z}) \le \theta_{\mathrm{FB}}(x^k,y^k,z^k),$$

from which we have

$$\theta^k_{\mathrm{FB}}(x^k,y^k,z^k) - \theta_{\mathrm{FB}}(\bar{x},\bar{y},\bar{z})$$

$$
\begin{aligned}
&= \theta^k_{\rm FB}(x^k,y^k,z^k) - \theta^k_{\rm FB}(\bar x,\bar y,\bar z) + \theta^k_{\rm FB}(\bar x,\bar y,\bar z) - \theta_{\rm FB}(\bar x,\bar y,\bar z)\\
&\le \theta^k_{\rm FB}(\bar x,\bar y,\bar z) - \theta_{\rm FB}(\bar x,\bar y,\bar z)\\
&\le \sup_{(x,y,z)\in B} |\theta^k_{\rm FB}(x,y,z) - \theta_{\rm FB}(x,y,z)|
\end{aligned}
$$

and

$$
\begin{aligned}
&\theta^k_{\rm FB}(x^k,y^k,z^k) - \theta_{\rm FB}(\bar x,\bar y,\bar z)\\
=\ &\theta^k_{\rm FB}(x^k,y^k,z^k) - \theta_{\rm FB}(x^k,y^k,z^k) + \theta_{\rm FB}(x^k,y^k,z^k) - \theta_{\rm FB}(\bar x,\bar y,\bar z)\\
\ge\ &\theta^k_{\rm FB}(x^k,y^k,z^k) - \theta_{\rm FB}(x^k,y^k,z^k)\\
\ge\ &-\sup_{(x,y,z)\in B} |\theta^k_{\rm FB}(x,y,z) - \theta_{\rm FB}(x,y,z)|.
\end{aligned}
$$

It follows that

$$
|\theta^k_{\rm FB}(x^k,y^k,z^k) - \theta_{\rm FB}(\bar x,\bar y,\bar z)| \le \sup_{(x,y,z)\in B} |\theta^k_{\rm FB}(x,y,z) - \theta_{\rm FB}(x,y,z)|.
$$

This together with (4.3) implies (4.2).

(2) Consider the case for (3.9). The inequality (4.4) still holds from the proof of (1). This means that, for every $\varepsilon > 0$, there exist positive constants $D(\varepsilon)$ and $\beta(\varepsilon)$, independent of N_k, such that

$$
{\rm Prob}\Big\{ \sup_{(x,y,z)\in B} \big| \| \frac{1}{N_k} \sum_{\xi^i\in\Omega_k} F(x,y,z,\xi^i)\|^2 - \|\mathbb{E}_\xi[F(x,y,z,\xi)]\|^2 \big| \ge \varepsilon \Big\} \le D(\varepsilon) e^{-N_k\beta(\varepsilon)}.
$$

From (3.8), there exists a constant C_0 such that $\Big| \|\Phi^{\mu_k}_{\rm NR}(x,y)\|^2 - \|\Phi_{\rm NR}(x,y)\|^2 \Big| \le C_0\mu_k$. We have

$$
\begin{aligned}
&\big|\Theta^k_{\rm NR}(x,y,z) - \Theta_{\rm NR}(x,y,z)\big|\\
&\le \Big| \| \frac{1}{N_k} \sum_{\xi^i\in\Omega_k} F(x,y,z,\xi^i)\|^2 - \|\mathbb{E}_\xi[F(x,y,z,\xi)]\|^2 \Big| + \Big| \|\Phi^{\mu_k}_{\rm NR}(x,y)\|^2 - \|\Phi_{\rm NR}(x,y)\|^2 \Big|\\
&\le \Big| \| \frac{1}{N_k} \sum_{\xi^i\in\Omega_k} F(x,y,z,\xi^i)\|^2 - \|\mathbb{E}_\xi[F(x,y,z,\xi)]\|^2 \Big| + C_0\mu_k.
\end{aligned}
$$

Since $\mu_k \to 0$ as $k \to \infty$, there exists $K > 0$ such that $\mu_k \le \frac{\varepsilon}{C_0}$ for $k > K$ and then, from (4.4) and the above inequality,

$$
\begin{aligned}
&{\rm Prob}\Big\{ \sup_{(x,y,z)\in B} |\Theta^k_{\rm NR}(x,y,z) - \Theta_{\rm NR}(x,y,z)| \ge 2\varepsilon \Big\}\\
&\le {\rm Prob}\Big\{ \sup_{(x,y,z)\in B} \big| \| \frac{1}{N_k} \sum_{\xi^i\in\Omega_k} F(x,y,z,\xi^i)\|^2 - \|\mathbb{E}_\xi[F(x,y,z,\xi)]\|^2 \big| \ge \varepsilon \Big\}\\
&\le D(\varepsilon) e^{-N_k\beta(\varepsilon)}.
\end{aligned}
$$

Then, in a similar manner to (1), we can get

$$
|\theta^k_{\rm NR}(x^k,y^k,z^k) - \theta_{\rm NR}(\bar x,\bar y,\bar z)| \le \sup_{(x,y,z)\in B} |\theta^k_{\rm FB}(x,y,z) - \theta_{\rm FB}(x,y,z)|
$$

and hence (4.1) is true. This completes the proof. □

5 Error Bounds for SSOCCP

In this section, we devote to deriving some results related to error bound conditions, which bound the distance from the vectors in a test set to a goal set, for (1.1). Error bounds play important roles in variational inequalities, especially in convergence analysis for various numerical algorithms [11]. Although error bound conditions are not involved in the convergence analysis given in the last sections, they may be useful in the subsequent study on (1.1).

To this end, we rewrite (1.1) as the following equivalent constraint system with nonconvex cone:

$$\mathbb{E}_\xi[F(x,y,z,\xi)] = 0, \quad (x^i,y^i) \in \Lambda_i, \quad i = 1,\dots,m,$$

where $\Lambda_i \subseteq \Re^{2n_i}$ $(i = 1,\dots,m)$ denotes the complementarity set given by

$$\Lambda_i := \{(s,t) \,|\, s \in \mathcal{K}^{n_i}, t \in \mathcal{K}^{n_i}, s^T t = 0\} \;=\; \{(s,t) \,|\, -t \in \mathcal{N}_{\mathcal{K}^{n_i}}(s)\},$$

and we denote by $x := (x^1,\dots,x^m) \in \Re^{n_1} \times \cdots \times \Re^{n_m}$, $y := (y^1,\dots,y^m) \in \Re^{n_1} \times \cdots \times \Re^{n_m}$. We consider its perturbed system

$$\mathbb{E}_\xi[F(x,y,z,\xi)] + r = 0, \quad (x^i,y^i) + p_i \in \Lambda_i, \quad i = 1,\dots,m, \tag{5.1}$$

where the perturbation parameters $r \in \Re^{n+l}$ and $p_i \in \Lambda_i$ $(i = 1,\dots,m)$. In particular, we suppose that the system (1.1) has a solution (x^*,y^*,z^*) and denote by $\mathcal{S}(r,p)$ the solution set of the parametric system (5.1).

Definition 5.1. We say that the system (1.1) has a *local error bound* at (x^*,y^*,z^*) if there exist $\mu,\varepsilon > 0$ such that

$$\mathrm{dist}((x,y,z),\mathcal{S}(0,0)) \le \mu\|(r,p)\|$$

holds for each $(r,p) \in \varepsilon\mathbb{B}$ and each $(x,y,z) \in \mathcal{S}(r,p) \cap \mathbb{B}_\varepsilon(x^*,y^*,z^*)$.

From [33], the system (1.1) has a local error bound at (x^*,y^*,z^*) if and only if the solution map $\mathcal{S} : \Re^{n+l} \times \Re^{2n} \rightrightarrows \Re^{2n+l}$ is calm at $(0,0,x^*,y^*,z^*)$ or pseudo upper-Lipschitz continuous around $(0,0,x^*,y^*,z^*)$. This means that, if $\mathcal{S}$ is either pseudo-Lipschitz continuous around $(0,0,x^*,y^*,z^*)$ or upper-Lipschitz continuous at $(0,0)$, then (1.1) has a local error bound at (x^*,y^*,z^*).

In what follows, we assume the validity of switching orders of taking expectation with respect to ξ and taking gradient with respect to (x,y,z) in front of $F(x,y,z,\xi)$. We make this assumption for the sake of more elegant formal expression, with the understanding that it can be easily guaranteed under mild boundedness conditions. It is worth mentioning that the results given below remain true without this assumption.

Definition 5.2. We say that the *no nonzero abnormal multiplier constraint qualification* (NNAMCQ) holds at (x^*,y^*,z^*) for the system (5.1) if there is no nonzero vector $(\lambda^F,\lambda^x,\lambda^y)$ with $(\lambda^x,\lambda^y) \in \mathcal{N}_{\Lambda_1}(x^{*,1},y^{*,1}) \times \cdots \times \mathcal{N}_{\Lambda_m}(x^{*,m},y^{*,m})$ such that

$$\begin{cases} \mathbb{E}_\xi[\nabla_x F(x^*,y^*,z^*,\xi)]\lambda^F + \lambda^x = 0, \\ \mathbb{E}_\xi[\nabla_y F(x^*,y^*,z^*,\xi)]\lambda^F + \lambda^y = 0, \\ \mathbb{E}_\xi[\nabla_z F(x^*,y^*,z^*,\xi)]\lambda^F = 0. \end{cases}$$

Following the idea in [32], we can prove that the NNAMCQ condition is sufficient to guarantee the existence of local error bounds.

Theorem 5.3. *Assume that the NNAMCQ holds at a solution* (x^*, y^*, z^*). *Then the solution map* $\mathcal{S}$ *is pseudo-Lipschitz continuous around* $(0, 0, x^*, y^*, z^*)$ *and hence the SSOCCP* (1.1) *has a local error bound at* (x^*, y^*, z^*).

Proof. From the Mordukhovich criteria [19], it suffices to show

$$D^*\mathcal{S}(0, 0, x^*, y^*, z^*)(0, 0, 0) = \{(0, 0)\}.$$

In fact, suppose that $(\gamma, \eta) \in D^*\mathcal{S}(0, 0, x^*, y^*, z^*)(0, 0, 0)$, where $\gamma \in \Re^{n+l}$ and $\eta \in \mathcal{K} \times \mathcal{K}$. By Definition 2.2, there holds

$$(\gamma, \eta, 0, 0, 0) \in \mathcal{N}_{\mathrm{gph}\mathcal{S}}(0, 0, x^*, y^*, z^*).$$

By the definition of limiting normal cone, there are $\{(r^k, p^k, x^k, y^k, z^k)\}$ and $\{(\gamma^k, \eta^k, \alpha^k, \beta^k, \tau^k)\}$ converging to $(0, 0, x^*, y^*, z^*)$ and $(\gamma, \eta, 0, 0, 0)$, respectively, with

$$(\gamma^k, \eta^k, \alpha^k, \beta^k, \tau^k) \in \mathcal{N}^\pi_{\mathrm{gph}\mathcal{S}}(r^k, p^k, x^k, y^k, z^k).$$

By the definition of proximal normal cone, for each k, there exists $M_k > 0$ such that

$$\begin{aligned} & (\gamma^k, \eta^k, \alpha^k, \beta^k, \tau^k)^T (r - r^k, p - p^k, x - x^k, y - y^k, z - z^k) \\ \leq \;& M_k \|(r - r^k, p - p^k, x - x^k, y - y^k, z - z^k)\|^2. \end{aligned}$$

holds for any $(r, p, x, y, z) \in \mathrm{gph}\mathcal{S}$. We further observe that $(r^k, p^k, x^k, y^k, z^k)$ is the unique optimal solution to the optimization problem

$$\begin{aligned} \min_{(r,p,x,y,z)} \quad & M_k \|(r - r^k, p - p^k, x - x^k, y - y^k, z - z^k)\|^2 \\ & \quad -(\gamma^k, \eta^k, \alpha^k, \beta^k, \tau^k)^T (r - r^k, p - p^k, x - x^k, y - y^k, z - z^k) \\ \text{s.t.} \quad & \mathbb{E}_\xi[F(x, y, z, \xi)] + r = 0, \ (x^i, y^i) + p^i \in \Lambda_i, \ i = 1, \ldots, m. \end{aligned}$$

Here and in the subsequent analysis, we denote by $x^k := (x^{k,1}, \ldots, x^{k,m}) \in \Re^{n_1} \times \cdots \times \Re^{n_m}$, $y^k := (y^{k,1}, \ldots, y^{k,m}) \in \Re^{n_1} \times \cdots \times \Re^{n_m}$ and so on.

In the hope of that the NNAMCQ holds at $(r^k, p^k, x^k, y^k, z^k)$ for the above problem for each k, we verify the following system for NNAMCQ conditions:

$$\left\{ \begin{array}{l} \lambda^F = 0, \ (\lambda^x, \lambda^y) = 0, \\ \mathbb{E}_\xi[\nabla_x F(x^k, y^k, z^k, \xi)]\lambda^F + \lambda^x = 0, \\ \mathbb{E}_\xi[\nabla_y F(x^k, y^k, z^k, \xi)]\lambda^F + \lambda^y = 0, \\ \mathbb{E}_\xi[\nabla_z F(x^k, y^k, z^k, \xi)]\lambda^F = 0, \\ (\lambda^x, \lambda^y) \in \mathcal{N}_{\Lambda_1}((x^{k,1}, y^{k,1}) + p^{k,1}) \times \cdots \times \mathcal{N}_{\Lambda_m}((x^{k,m}, y^{k,m}) + p^{k,m}), \end{array} \right.$$

which is only satisfied by $(\lambda^F, \lambda^x, \lambda^y) = (0, 0, 0)$ and hence the NNAMCQ holds at $(r^k, p^k, x^k, y^k, z^k)$ for the above optimization problem. In consequence, there exist Lagrangian multipliers $\bar{\gamma}^k$ and $\bar{\eta}^k$ such that the KKT conditions hold, that is,

$$\left\{ \begin{array}{l} -\gamma^k + \bar{\gamma}^k = 0, \\ -\eta^k + \bar{\eta}^k = 0, \\ -(\alpha^k, \beta^k) + \mathbb{E}_\xi[\nabla_{x_i} F(x^k, y^k, z^k, \xi)]\bar{\gamma}^k + \bar{\eta}^k = 0, \\ -\tau^k + \mathbb{E}_\xi[\nabla_z F(x^k, y^k, z^k, \xi)]\bar{\gamma}^k = 0, \\ \bar{\eta}^k \in \mathcal{N}_{\Lambda_1}((x^{k,1}, y^{k,1}) + p^{k,1}) \times \cdots \times \mathcal{N}_{\Lambda_m}((x^{k,m}, y^{k,m}) + p^{k,m}). \end{array} \right.$$

Taking a limit in the above system, we have from the continuous differentiablity of F and the observation that $\gamma^k = \hat{\gamma}^k$ and $\eta^k = \hat{\eta}^k$ that

$$\begin{cases} \mathbb{E}_\xi[\nabla_{(x,y)} F(x^*, y^*, z^*, \xi)]\gamma + \eta = 0, \\ \mathbb{E}_\xi[\nabla_z F(x^*, y^*, z^*, \xi)]\gamma = 0, \\ \eta \in \mathcal{N}_{\Lambda_1}(x^{*,1}, y^{*,1}) \times \cdots \times \mathcal{N}_{\Lambda_m}(x^{*,m}, y^{*,m}). \end{cases}$$

Consequently, by the constraint qualification assumption, we have $(\gamma, \eta) = (0, 0)$ and hence $\mathcal{S}$ is pseudo-Lipschitz continuous around $(0, 0, x^*, y^*, z^*)$. This completes the proof. □

Recall that a set-valued mapping is called a *polyhedral multifunction* if its graph is a union of finitely many polyhedral convex sets. By [22], a polyhedral multifunction is upper-Lipschitz and hence calm, which implies the existence of local error bounds.

Theorem 5.4. *Assume that $F(x, y, z, \xi)$ is affine with respect to (x, y, z) and continuously integrable with respect to ξ over Ω and $n_i \leq 2$ for each i. Then the solution map $\mathcal{S}$ is upper-Lipschitz continuous around $(0, 0)$ and hence the SSOCCP (1.1) has a local error bound at an arbitrary solution (x^*, y^*, z^*).*

Proof. By the assumptions, $\mathbb{E}_\xi[F(x, y, z, \xi)]$ is affine. Note that each second-order-cone $\mathcal{K}^{n_i}$ is actually polyhedral and so is each complementarity set Λ_i. It is easy to see that the graph of the set-valued map $\mathcal{S}$ is a union of polyhedral convex sets and hence $\mathcal{S}$ is a polyhedral multifunction. By [22, Proposition 1], $\mathcal{S}$ is upper-Lipschitz continuous at $(0, 0)$. This completes the proof. □

6 Application in Chance-Constrained Optimal Power Flow

The theoretical results given in the previous sections indicate that the SSOCCP (1.1) can be solved via the minimization approach (1.6). In this section, as a further supplement, we consider a practical engineering problem.

Consider the transmission grids where the convey power is economical with minimal losses. Transmission systems balance consumption/load and generation using a complex strategy. Consider the case of a sudden load increase. In this case, generator frequency will start to drop. The so-called automatic generation control (AGC) undertakes the adjustment of generation levels to return frequency to nominal value. The optimal power flow (OPF) algorithm typically runs as frequently every 15 minutes and provides information for AGC, which ultimately resets generator output levels over a control area of the transmission grid.

Here, we employ power engineering terms such as “bus” and “line” to describe OPF [4, 15]. The set of all buses is denoted by $\mathcal{V}$, the set of lines by $\mathcal{E}$, and the set of buses that houses generators by $\mathcal{G}$ respectively. In what follows, we let $n := |\mathcal{V}|$. A line joining buses i and j is denoted by $\{i, j\}$. Then the standard DC-formulation for OPF can be expressed as the constrained optimization problem

$$\begin{aligned} \min_{(\theta,p)} \quad & c(p) \\ \text{s.t.} \quad & B\theta = p - d, \ p_i^{min} \leq p_i \leq p_i^{max}, \quad \forall i \in \mathcal{G}, \\ & |\beta_{ij}(\theta_i - \theta_j)| \leq f_{ij}^{max}, \quad \forall \{i, j\} \in \mathcal{E}. \end{aligned}$$

The goal is to determine the vector $p \in \Re^{\mathcal{G}}$, where p_i is the output of generator i for $i \in \mathcal{G}$, so as to minimize an objective function $c(p)$, which is usually a convex separable quadratic function of p. $\theta \in \Re^{|\mathcal{V}|}$ and $d \in \Re_+^{|\mathcal{V}|}$ represent the vector of phase angles and the vector of demands respectively. The $n \times n$ matrix B made up of line susceptance β_{ij} is a weighted-Laplacian defined as

$$B_{ij} := \begin{cases} -\beta_{ij} & \{i,j\} \in \mathcal{E}, \\ \sum_{k;\{k,j\}\in\mathcal{E}} \beta_{kj} & i = j, \\ 0 & \text{otherwise,} \end{cases}$$

The quantities p_i^{min} and p_i^{max} can be used to enforce the convention $p_i = 0$ for each $i \notin \mathcal{G}$ and, if $i \in \mathcal{G}$, p_i^{min} and p_i^{max} represent the lower and upper generation bounds which are generator-specific. f_{ij}^{max} is the transmission line limit (typically a thermal limit) for $\{i,j\}$ and is assumed to be strictly enforced.

The problem changes when renewable power sources (such as wind, etc.) are incorporated. We assume that a subset $\mathcal{W}$ of buses holds uncertain power sources (e.g. wind farms) and, for each $j \in \mathcal{W}$, we write the amount of power generated by source j at time t as $\mu_j + \xi_j(t)$, where μ_j is the forecast output of farm j in the time period of interest. The hazard embodied is that the uncertainty $\xi_j(t)$ can be large, resulting in stochastic changes in power flows significant enough to overload power lines. Lowering of the thermal limits (the quantities f_{ij}^{max} etc.) can prevent overloads, but it also forces excessively conservative choices of the generation redispatch, potentially causing extreme volatility of the electricity market.

Power lines do not fail (i.e., trip) instantly when their flows go beyond the thermal limits. A line carrying flow that exceeds the line's thermal limit will gradually heat up, possibly sag, and may trip or be disconnected by the operator through a variety of processes (such as a contact). Additionally, the rate at which a line overheats depends on its overload, which may dynamically change (or even temporarily disappear) as flows adjust due to external factors, such as fluctuations in renewable outputs. Ideally, the transmission constraint should be of the form "for each line, the fraction of the time that it exceeds its limit within a certain time window is small". Nevertheless, an exact representation of line tripping is not tractable. Instead, Bienstock et al. [5] propose a static proxy of the ideal model with a chance constraint, namely, they require the probability that a given line exceeds its limit to be small. In particular, they use boldface to indicate uncertain quantities. The chance constraint for line $\{i,j\}$ is

$$\text{Prob}\{\boldsymbol{f_{ij}} > f_{ij}^{max}\} < \epsilon_{ij}, \quad \forall\{i,j\},$$

where $\boldsymbol{f_{ij}}$ denotes the flow on a given line $\{i,j\}$. In the chance-constrained model, it is assumed in [5] that the wind power fluctuation is independent at different sites. For each $j \in \mathcal{W}$, the (stochastic) amount of power generated by source j is of the form $\mu_j + \boldsymbol{\xi_j}$, where μ_j is constant and known from the forecast, and $\boldsymbol{\xi_j}$ is a zero mean independent random variable with known standard deviation σ_j. Since the power injections at each bus are fluctuating, Bienstock et al. employ the affine control scheme so as to ensure that generation is equal to demand at all time within the time window of interest. Mathematical expression becomes $\boldsymbol{p_i} = \bar{p}_i - \alpha_i \sum_{j\in\mathcal{W}} \boldsymbol{\xi_j}$, $\forall i \in \mathcal{G}$, where the quantities $\bar{p}_i \geq 0$ is the design variable representing the average production of the generator i, and $\alpha_i \geq 0$ is the design variable satisfying $\sum_{i\in\mathcal{G}} \alpha_i = 1$. The affine control scheme creates the possibility of requiring a generator to produce power beyond its limit, which is inevitable with high penetration of wind power.

By applying the chance-constrained optimization ideas to the setting of OPF under uncertainty, [5] eventually provides the following generic DC-formulation for the chance-constrained OPF problem that is valid under the assumption of linear power flow and statistical independence of wind fluctuation at different buses, while using control law to specify standard generation responses to the wind fluctuations:

$$
\begin{aligned}
&\min_{(\bar{\theta},\bar{p},\alpha)} \mathbb{E}_{\boldsymbol{\xi}}[c(\bar{p}-(e^T\boldsymbol{\xi})\alpha)]\\
&\quad \text{s.t.} B\bar{\theta} = \bar{p}+\mu-d,\ \sum_{i\in\mathcal{G}}\alpha_i = 1,\ \alpha\geq 0,\ \bar{p}\geq 0,\ \sum_{i\in\mathcal{V}}(\bar{p}_i+\mu_i-d_i)=0,\\
&\qquad \text{Prob}\{\bar{p}_g-(e^T\boldsymbol{\xi})\alpha_i < p_g^{min}) < \epsilon_g,\ \text{Prob}(\bar{p}_g-(e^T\boldsymbol{\xi})\alpha_i > p_g^{max}\} < \epsilon_g, \forall g,\\
&\qquad \text{Prob}\{\beta_{ij}(\bar{\theta}_i-\bar{\theta}_j+[\check{B}(\boldsymbol{\xi}-(e^T\boldsymbol{\xi})\alpha)]_i-[\check{B}(\boldsymbol{\xi}-(e^T\boldsymbol{\xi})\alpha)]_j) > f_{ij}^{max}\} < \epsilon_{ij}, \forall\{i,j\}\in\mathcal{E},\\
&\qquad \text{Prob}\{\beta_{ij}(\bar{\theta}_i-\bar{\theta}_j+[\check{B}(\boldsymbol{\xi}-(e^T\boldsymbol{\xi})\alpha)]_i-[\check{B}(\boldsymbol{\xi}-(e^T\boldsymbol{\xi})\alpha)]_j) < -f_{ij}^{max}\} < \epsilon_{ij}, \forall\{i,j\}\in\mathcal{E}.
\end{aligned}
$$

The decision variables in this formulation include the affine control α, the standard $\bar{p}$ and $\bar{\theta}$ used in the standard OPF. For $i\notin\mathcal{W}$, there holds $\mu_i=\sigma_i=0$ and, for $i\notin\mathcal{G}$, $\bar{p}_i=\alpha_i=0$. The objective is the expected cost incurred by the stochastic generation over the varying wind power output $\boldsymbol{\xi}$, and $e\in\Re^n$ is the vector of all 1's. In standard power engineering practice, generation cost function $c(\cdot)$ is convex, quadratic and separable. $\hat{B}$ denotes the submatrix obtained from B by removing row and column $n\notin\mathcal{G}\cup\mathcal{W}$ for convenience and

$$
\check{B} := \begin{pmatrix} \hat{B}^{-1} & 0\\ 0 & 0\end{pmatrix}.
$$

Different from [5] which presents a fully convex deterministic formulation (unfortunately, their chain of reformulating contains an erroneous expression), we employ the second-order-cone representation to recast those constraints of the chance-constrained form, which leaves us the following stochastic programming problem with the second-order-cone constraints on variables $\{\bar{p},\alpha,\bar{\theta},\delta,s\}$:

$$
\begin{aligned}
\min_{(\bar{\theta},\bar{p},\alpha)} \quad & \mathbb{E}_{\boldsymbol{\xi}}\Big[\sum_{i\in\mathcal{G}} c_{i2}(\bar{p}_i^2+\alpha_i^2(\sum_{j\in\mathcal{W}}\boldsymbol{\xi_j})^2)+c_{i1}\bar{p}_i+c_{i0}\Big]\\
\text{s.t.} \quad & \sum_{j=1}^{n-1}\hat{B}_{ij}\delta_j=\alpha_i,\quad 1\leq i\leq n-1,\\
& \sum_{j=1}^{n-1}\hat{B}_{ij}\bar{\theta}_j=\bar{p}_i+\mu_i-d_i,\quad 1\leq i\leq n-1,\\
& \sum_{i\in\mathcal{G}}\alpha_i=1,\ \alpha\geq 0,\ \bar{p}\geq 0,\\
& \bar{p}_n=\alpha_n=\delta_n=\bar{\theta}_n=0,\\
& \phi^{-1}(1-\epsilon_g)\alpha_g\Big(\sum_{k\in\mathcal{W}}\sigma_k^2\Big)^{1/2}\leq p_g^{max}-\bar{p}_g,\quad \forall g,\\
& \phi^{-1}(1-\epsilon_g)\alpha_g\Big(\sum_{k\in\mathcal{W}}\sigma_k^2\Big)^{1/2}\leq \bar{p}_g-p_g^{min},\quad \forall g,\\
& \beta_{ij}|\bar{\theta}_i-\bar{\theta}_j|+\beta_{ij}\phi^{-1}(1-\epsilon_{ij})s_{ij}\leq f_{ij}^{max},\quad \forall\{i,j\}\in\mathcal{E},\\
& \Big[\sum_{k\in\mathcal{W}}\sigma_k^2(\check{B}_{ik}-\check{B}_{jk}-\delta_i+\delta_j)^2\Big]^{1/2}\leq s_{ij},\ \forall\{i,j\}\in\mathcal{E},
\end{aligned}
$$

where both $\delta:=\check{B}\alpha$ and s_{ij} are auxiliary variables and ϕ is the cumulative distribution function of a standard normally distributed random variable. In order to exploit the stochastic second-order-conic problem in a computationally approachable manner, we naturally adopt the prevailing KKT approach which reformulates the problem equivalently from a programming problem to an inequality system under suitable constraint qualifications. To do this, we rewrite the above program as

$$
\begin{aligned}
\min_{(\bar\theta,\bar p,\alpha)} \quad & \mathbb{E}_{\boldsymbol{\xi}}\Big[\sum_{i\in\mathcal{G}} c_{i2}(\bar p_i^2+\alpha_i^2(\sum_{j\in\mathcal{W}}\boldsymbol{\xi_j})^2)+c_{i1}\bar p_i+c_{i0}\Big] & & \\
\text{s.t.} \quad & \sum_{j=1}^{n-1}\hat B_{ij}\delta_j-\alpha_i=0, & & l_i,\quad \forall i\in\{1,\dots,n-1\},\\
& \sum_{j=1}^{n-1}\hat B_{ij}\bar\theta_j-\bar p_i-\mu_i+d_i=0, & & m_i,\quad \forall i\in\{1,\dots,n-1\},\\
& \sum_{i\in\mathcal{G}}\alpha_i-1=0, & & \mu,\\
& p_i^{max}-\bar p_i-\phi^{-1}(1-\epsilon_i)\alpha_i\Big(\sum_{k\in\mathcal{W}}\sigma_k^2\Big)^{1/2}\ge 0, & & u_i,\quad \forall i\in\mathcal{G},\\
& \bar p_i-p_i^{min}-\phi^{-1}(1-\epsilon_i)\alpha_i\Big(\sum_{k\in\mathcal{W}}\sigma_k^2\Big)^{1/2}\ge 0, & & v_i,\quad \forall i\in\mathcal{G},\\
& f_{ij}^{max}-\beta_{ij}\phi^{-1}(1-\epsilon_{ij})s_{ij}-\beta_{ij}(\bar\theta_i-\bar\theta_j)\ge 0, & & \eta_{ij},\quad \forall\{i,j\}\in\mathcal{E},\\
& \beta_{ij}(\bar\theta_i-\bar\theta_j)+f_{ij}^{max}-\beta_{ij}\phi^{-1}(1-\epsilon_{ij})s_{ij}\ge 0, & & \tau_{ij},\quad \forall\{i,j\}\in\mathcal{E},\\
& \begin{pmatrix} s_{ij}\\ \vdots \\ \sigma_k(\check B_{ik}-\check B_{jk}-\delta_i+\delta_j)\\ \vdots\end{pmatrix}\in\mathcal{K}^{|\mathcal{W}|+1}, & & \begin{pmatrix}\pi_{ij}\\ \vdots\\ \lambda_{ij}^k\\ \vdots\end{pmatrix}\in\mathcal{K}^{|\mathcal{W}|+1},\quad \forall\{i,j\}\in\mathcal{E},\\
& \alpha_i\ge 0, & & r_i,\quad \forall i\in\{1,\dots,n-1\},\\
& \bar p_i\ge 0, & & q_i,\quad \forall i\in\{1,\dots,n-1\},
\end{aligned}
$$

where $\{l_i, m_i, \mu, u_i, v_i, \eta_{ij}, \tau_{ij}, \pi_{ij}, \lambda_{ij}^k(k\in\mathcal{W}), r_i, q_i\}$ are Lagrangian multipliers. Under some mild bounded assumptions on wind output ξ, the KKT conditions can be written as

$$
\begin{aligned}
& \mathbb{E}_{\boldsymbol{\xi}}\Big[c_{i2}(2\bar p_i+\alpha_i^2(\sum_{j\in\mathcal{W}}\boldsymbol{\xi_j})^2)+c_{i1}\Big]-m_i+u_i-v_i-q_i=0, & & \forall i\in\mathcal{G},\\
& m_i+q_i=0, & & \forall i\in\{1,\dots,n-1\}\setminus\mathcal{G},\\
& \mathbb{E}_{\boldsymbol{\xi}}\Big[2c_{i2}\alpha_i(\sum_{j\in\mathcal{W}}\boldsymbol{\xi_j})^2\Big]-l_i+\mu+(u_i+v_i)\phi^{-1}(1-\epsilon_i)\Big(\sum_{k\in\mathcal{W}}\sigma_k^2\Big)^{1/2}-r_i=0, & & \forall i\in\mathcal{G},\\
& l_i+r_i=0, & & \forall i\in\{1,\dots,n-1\}\setminus\mathcal{G},\\
& m_i\hat B_{ii}+\eta_{ij}\beta_{ij}-\tau_{ij}\beta_{ij}=0, & & \forall\{i,j\}\in\mathcal{E},\\
& l_i\hat B_{ii}+\sum_{k\in\mathcal{W}}\lambda_{ij}^k\sigma_k=0, & & \forall\{i,j\}\in\mathcal{E},\\
& (\eta_{ij}+\tau_{ij})\beta_{ij}\phi^{-1}(1-\epsilon_{ij})-\pi_{ij}=0, & & \forall\{i,j\}\in\mathcal{E},\\
& \bar p_n=\alpha_n=\delta_n=\bar\theta_n=0,
\end{aligned}
$$

$$
\begin{aligned}
&\mathcal{K}^{|\mathcal{W}|+1} \ni \begin{pmatrix} \pi_{ij} \\ \vdots \\ \lambda_{ij}^k \\ \vdots \end{pmatrix} \perp \begin{pmatrix} s_{ij} \\ \vdots \\ \sigma_k(\check{B}_{ik} - \check{B}_{jk} - \delta_i + \delta_j) \\ \vdots \end{pmatrix} \in \mathcal{K}^{|\mathcal{W}|+1}, && \forall \{i,j\} \in \mathcal{E}, \\
&0 \le r_i \quad \perp \quad \alpha_i \ge 0, && \forall i \in \{1, \ldots, n-1\}, \\
&0 \le q_i \quad \perp \quad \bar{p}_i \ge 0, && \forall i \in \{1, \ldots, n-1\}.
\end{aligned}
$$

Denote $(\pi_{ij}, \ldots, \lambda_{ij}^k, \ldots, r_i, q_i)^{\mathrm{T}} \in \mathcal{K}^{|\mathcal{W}|+1} \times \mathcal{K}^1 \times \mathcal{K}^1$ by x, $(s_{ij}, \ldots, \sigma_k(\check{B}_{ik} - \check{B}_{jk} - \delta_i + \delta_j), \ldots, \alpha_i, \bar{p}_i)^{\mathrm{T}} \in \mathcal{K}^{|\mathcal{W}|+1} \times \mathcal{K}^1 \times \mathcal{K}^1$ by y, and the above system of equations by $\mathbb{E}_{\boldsymbol{\xi}}[F(x, y, z, \xi)] = 0$. Then, the induced KKT conditions are in the form of (1.1), which, as argued in Section 4, may be efficiently solvable.

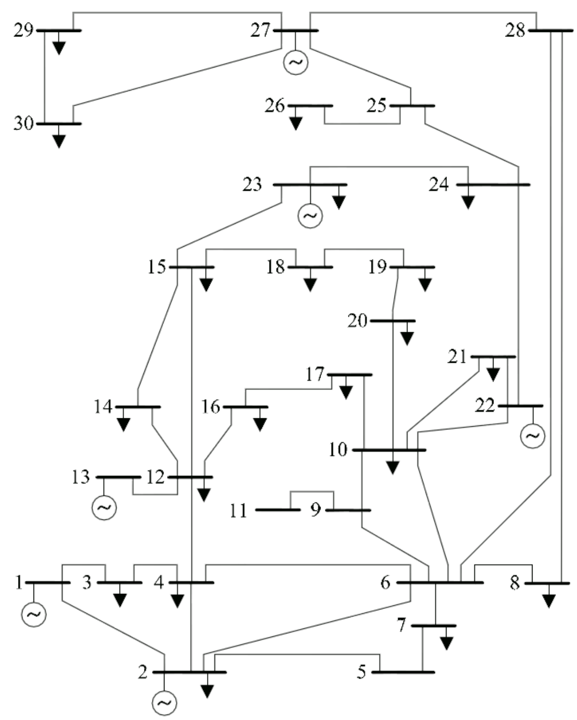

Figure 1: Single line diagram of IEEE 30-bus test system

We performed experiments involving the IEEE 30-bus system as in Figure 1. In our test, we used the standard quadratic cost functions and system data associated with this case (provided with [21]), $\epsilon_{ij} = 3.27\%$ for all lines, and two sources of wind power were added at arbitrary buses to meet 18% of demand in the case of average wind. Then we accounted for fluctuations in wind assuming Gaussian and site-independent fluctuations with standard derivation set as 40% of the respective means. The results, which are shown in Table 1, illustrate the control and dispatch decisions that the system regulator makes in the setting of chance-constrained OPF on those buses holding controllable generators. From the numerical results, we observed that, as the sample size increases, the convergence is stable, e.g., there

Table 1: Test results on IEEE 30-bus system

sample size	50	300	1000	2000
total cost	1567.715	1538.961	1539.061	1539.102
bus1	α_1=0.062	α_1=0.076	α_1=0.076	α_1=0.076
	$\bar{p}_1$=1.143	$\bar{p}_1$=1.407	$\bar{p}_1$=1.407	$\bar{p}_1$= 1.407
bus2	α_2=0.930	α_2=0.924	α_2=0.924	α_2=0.924
	$\bar{p}_2$=41.748	$\bar{p}_2$=40.737	$\bar{p}_2$=40.737	$\bar{p}_2$=40.737
bus13	α_{13}=0	α_{13}=0	α_{13}=0	α_{13}=0
	$\bar{p}_{13}$=22.294	$\bar{p}_{13}$=22.294	$\bar{p}_{13}$=22.294	$\bar{p}_{13}$=22.294
bus22	α_{22}=0	α_{22}=0	α_{22}=0	α_{22}=0
	$\bar{p}_{22}$=0	$\bar{p}_{22}$=0	$\bar{p}_{22}$=0	$\bar{p}_{22}$=0
bus23	α_{23}=0	α_{23}=0	α_{23}=0	α_{23}=0
	$\bar{p}_{23}$= 9.750	$\bar{p}_{23}$=9.297	$\bar{p}_{23}$=9.297	$\bar{p}_{23}$=9.297
bus27	α_{27}=0.008	α_{27}=0	α_{27}=0	α_{27}=0
	$\bar{p}_{27}$=0.143	$\bar{p}_{27}$=0	$\bar{p}_{27}$=0	$\bar{p}_{27}$=0

is clearly convergent trend toward the controlling and dispatching variables. Therefore, we believe that the proposed EV approach is promising.

7 Conclusions

To our best knowledge, this is the first attempt to study the SSOCCP (1.1) systematically. Motivated by the works on its deterministic version, and in order to develop numerical algorithms for solving (1.1), we first transform it into the minimization problem (1.6) equivalently and then present an approximation method based on the Monte Carlo approximation and some smoothing techniques. We have given a comprehensive convergence analysis for the approximation method. We have also derived some results related to error bounds. Furthermore, as an application of the above-mentioned theoretical results, we have performed experiments on the chance-constrained optimal power flow in the last section.

Note that, except the minimization approach (1.6), there are some other approaches suitable to dealing with the deterministic version of (1.1); see, e.g., [6] for details. To develop more approximation methods for (1.1) will be our next target. Moreover, the two-stage variational inequality problems have received much attention in the recent literature [9, 24]. Two-stage stochastic second-order-cone complementarity problems will be another target.

Acknowledgements.

The authors are grateful to Dr. Jin Zhang of Southern University of Science and Technology and two anonymous referees for their helpful comments and suggestions, which have led to much improvement of the paper.

References

[1] F. Alizadeh and D. Goldfarb, Second-order cone programming, *Math. Program.* 95 (2003) 3–51.

[2] J.P. Aubin, Lipschitz behavior of solutions to convex minimization problems, *Math. Oper. Res.* 9 (1994) 87–111.

[3] M. Baradar and M.R. Hesamzadeh, A stochastic SOCP optimal power flow with wind power uncertainty, *Proceedings of the IEEE PES General Meeting* (2014) 27–31.

[4] A. Bergen and V. Vittal, *Power System Analysis*, Prentice-Hall, New Jersey, 2000.

[5] D. Bienstock, M. Chertkov and S. Harnett, Chance-constrained optimal power flow: Risk-aware network control under uncertainty, *SIAM Rev.* 56 (2014) 461–495.

[6] J.S. Chen and S. Pan, A survey on SOC complementarity functions and solution methods for SOCPs and SOCCPs, *Pac. J. Optim.* 8 (2012) 33–74.

[7] J.S. Chen and P. Tseng, An unconstrained smooth minimization reformulation of the second-order cone complementarity problem, *Math. Program.* 104 (2005) 297–327.

[8] X. Chen and M. Fukushima, Expected residual minimization method for stochastic linear complementarity problems, *Math. Oper. Res.* 30 (2005) 1022–1038.

[9] X. Chen, H. Sun and H. Xu, Discrete approximation of two-stage stochastic and distributionally robust linear complementarity problems, *Math. Program.* 177 (2019) 255–289.

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

[11] F. Facchinei and J.S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems, I and II*, Springer, New York, 2003.

[12] M. Fukushima, Z.Q. Luo and P. Tseng, Smoothing functions for second-order-cone complementarity problems, *SIAM J. Optim.* 12 (2001) 436–460.

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

[14] S. Hayashi, N. Yamashita and M. Fukushima, A combined smoothing and regularization method for monotone second-order cone complementarity problems, *SIAM J. Optim.* 15 (2005) 593–615.

[15] P. Kundur, *Power System Stability and Control*, McGraw-Hill, New York, 1994.

[16] G.H. Lin, M.J. Luo, D. Zhang and J. Zhang, Stochastic second-order-cone complementarity problems: Expected residual minimization formulation and its applications, *Math. Program.* 165 (2017) 197–233.

[17] G.H. Lin and M. Fukushima, Stochastic equilibrium problems and stochastic mathematical programs with equilibrium constraints: A survey, *Pac. J. Optim.* 6 (2010) 455–482.

[18] B.S. Mordukhovich, Generalized differential calculus for nonsmooth and set-valued mappings, *J. Math. Anal. Appl.* 183 (1994) 250–288.

[19] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation: I and II*, Springer, Berlin, 2006.

[20] B. Patrick, *Probability and Measure*, A Wiley-Interscience Publication, John Wiley & Sons, New York, USA, 1995.

[21] Power System Test Case Archive, http://www.ee.washington.edu/research/pstca/.

[22] S.M. Robinson, Some continuity properties of polyhedral multifunctions, *Math. Program. Study.* 14 (1981) 206–214.

[23] S.M. Robinson, Stability theory for systems of inequalities, Part I: Linear systems, *SIAM J. Numer. Anal.* 12 (1975) 754–769.

[24] R.T. Rockafellar and R.J. Wets, Stochastic variational inequalities: Single-stage to multistage, *Math. Program.* 165 (2017) 331–360.

[25] R.T. Rockafellar and R.J. Wets, *Variational Analysis*, Springer, New York, 1998.

[26] U.V. Shanbhag, Stochastic variational inequality problems: Applications, analysis, and algorithms, *Tutorials Oper. Res.* (2013) 71–107.

[27] A. Shapiro and H. Xu, Stochastic mathematical programs with equilibrium constraints: Modelling and sample average approximation, *Optim.* 57 (2008) 395–418.

[28] H. Sun and H. Xu, A note on uniform exponential convergence of sample average approximation of random functions, *J. Math. Anal. Appl.* 385 (2012) 698–708.

[29] Y. Xie and U.V. Shanbhag, On robust solutions to uncertain monotone linear complementarity problems (LCPs) and their variants, *53rd IEEE Conference on Decision and Control* (2014) 2834–2839.

[30] H. Xu, Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, *J. Math. Anal. Appl.* 368 (2010) 692–710.

[31] X.M. Yang, Y. Zhao and G.H. Lin, Deterministic bicriteria model for stochastic variational inequalities, *J. Oper. Res. Soc. China.* 6 (2018) 507–527.

[32] J.J. Ye, Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints, *SIAM J. Optim.* 10 (2000) 943–962.

[33] J.J. Ye and X.Y. Ye, Necessary optimality conditions for optimization problems with variational inequality constraints, *Math. Oper. Res.* 22 (1997) 977–997.

Manuscript received 13 July 2018
revised 6 December 2018
accepted for publication 13 March 2019

Guo Sun
School of Management, Shanghai University
Shanghai 200444 and School of Management Science
Qufu Normal University, Shandong 276800, China
E-mail address: qrsunguo@163.com

Liying Yu
School of Management, Shanghai University
Shanghai 200444, China
E-mail address: yuliying@shu.edu.cn

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

Xiaodai Dong
Department of Electrical and Computer Engineering
University of Victoria, Victoria
BC V8P 5C2, Canada
E-mail address: xdong@ece.uvic.ca