

2017

Stochastic second-order-cone complementarity problems: Expected residual minimization formulation and its applications

Gui-Hua Lin

Shanghai University, guihualin@shu.edu.cn

Mei-Ju Luo

Liaoning University, meijuluolnu@126.com

Dali Zhang

Shanghai Jiao Tong University, zhangdl@sjtu.edu.cn

Jin Zhang

Hong Kong Baptist University, zhangjin@hkbu.edu.hk

This document is the authors' final version of the published article.

Link to published article: <http://dx.doi.org/10.1007/s10107-017-1121-z>

APA Citation

Lin, G., Luo, M., Zhang, D., & Zhang, J. (2017). Stochastic second-order-cone complementarity problems: Expected residual minimization formulation and its applications. *Mathematical Programming*, 165 (1), 197-233. <https://doi.org/10.1007/s10107-017-1121-z>

Stochastic Second-Order-Cone Complementarity Problems: Expected Residual Minimization Formulation and Its Applications

Gui-Hua Lin · Mei-Ju Luo · Dali Zhang · Jin Zhang

Received: February 2015 / Revised: November 2015, March 2016, and September 2016

Abstract This paper considers a class of stochastic second-order-cone complementarity problems (SSOCCP), which are generalizations of the noticeable stochastic complementarity problems and can be regarded as the Karush-Kuhn-Tucker conditions of some stochastic second-order-cone programming problems. Due to the existence of random variables, the SSOCCP may not have a common solution for almost every realization. In this paper, motivated by the works on stochastic complementarity problems, we present a deterministic formulation called the expected residual minimization (ERM) formulation for SSOCCP. We present an approximation method based on the Monte Carlo approximation techniques and investigate some properties related to existence of solutions of the ERM formulation. Furthermore, we experiment some practical applications, which include a stochastic natural gas transmission problem and a stochastic optimal power flow problem in radial network.

This work was supported in part by NSFC (No. 11501275, 71501127, 11671250, 11601458), Humanity and Social Science Foundation of Ministry of Education of China (No. 15Y-JA630034), Scientific Research Fund of Liaoning Provincial Education Department (No. L2015199), and Hong Kong Baptist University FRG1/15-16/027.

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

Mei-Ju Luo
School of Mathematics, Liaoning University, Liaoning 110036, China
E-mail: meijuluolu@126.com

Dali Zhang
Sino-US Global Logistics Institute, Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200030, China
E-mail: zhangdl@sjtu.edu.cn

Jin Zhang
Department of Mathematics, Hong Kong Baptist University, Hong Kong, China
E-mail: zhangjin@hkbu.edu.hk

Keywords SSOCCP · ERM formulation · Monte Carlo approximation · natural gas transmission · optimal power flow

1 Introduction

The second-order cone (SOC) in \mathfrak{R}^ν is a closed cone defined as

$$\mathcal{K}^\nu := \{(x_1, x_2) \in \mathfrak{R} \times \mathfrak{R}^{\nu-1} \mid \|x_2\| \leq x_1\},$$

where $\|\cdot\|$ denotes the Euclidean norm. The second-order-cone complementarity problem (SOCCP) is to find vectors $x, y \in \mathfrak{R}^n$ and $z \in \mathfrak{R}^l$ satisfying

$$x \in \mathcal{K}, \quad y \in \mathcal{K}, \quad x^T y = 0, \quad F(x, y, z) = 0, \quad (1)$$

where $F : \mathfrak{R}^n \times \mathfrak{R}^n \times \mathfrak{R}^l \rightarrow \mathfrak{R}^n \times \mathfrak{R}^l$ is continuously differentiable and

$$\mathcal{K} := \mathcal{K}^{n_1} \times \cdots \times \mathcal{K}^{n_m} \quad (2)$$

with $n_1 + \cdots + n_m = n$. This problem is clearly a generalization of the classical mixed complementarity problems and, especially, it includes the Karush-Kuhn-Tucker conditions of various second-order-cone programs (SOCP), which have lots of applications in engineering design and portfolio optimization etc. [2], as special cases. The SOCCP (1) has attracted much attention of many researchers and there have been proposed several methods for solving it; see, e.g., [4, 9, 15, 17].

In order to develop numerical algorithms for SOCCP, the so-called SOC complementarity function $\phi : \mathfrak{R}^\nu \times \mathfrak{R}^\nu \rightarrow \mathfrak{R}^\nu$ satisfying

$$s \in \mathcal{K}^\nu, \quad t \in \mathcal{K}^\nu, \quad s^T t = 0 \quad \iff \quad \phi(s, t) = 0 \quad (3)$$

has been studied extensively. Two such functions known in the literature are presented by Fukushima et al. in [15]: One is the vector-valued Fischer-Burmeister function associated with \mathcal{K}^ν defined as

$$\phi_{\text{FB}}(s, t) := s + t - (s^2 + t^2)^{1/2} \quad (4)$$

and the other is the natural residual function associated with \mathcal{K}^ν defined as

$$\phi_{\text{NR}}(s, t) := s - [s - t]_+, \quad (5)$$

where $[\cdot]_+$ denotes the projection operator onto the convex cone \mathcal{K}^ν .

To understand the above two functions, we need to review some basic concepts in Jordan algebras. For any $s = (s_1, s_2) \in \mathfrak{R} \times \mathfrak{R}^{\nu-1}$ and $t = (t_1, t_2) \in \mathfrak{R} \times \mathfrak{R}^{\nu-1}$, their Jordan product is defined as

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

The identity element under this product is $e := (1, 0, \dots, 0)^T \in \mathfrak{R}^\nu$ and, for simplicity, we denote by $s^2 = s \circ s$. See [12] for more details about the Jordan product associated with symmetric cones.

We next recall the spectral factorization of vectors in \mathfrak{R}^ν associated with the cone \mathcal{K}^ν . It is well-known that any vector $s = (s_1, s_2) \in \mathfrak{R} \times \mathfrak{R}^{\nu-1}$ can be decomposed as

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

where $\{\lambda_1, \lambda_2\}$ and $\{u^1, u^2\}$ are respectively the spectral values and the associated spectral vectors of s given by, for $i = 1, 2$,

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

with w being an arbitrary unit vector in $\mathfrak{R}^{\nu-1}$. It is obvious that, if $s_2 \neq 0$, the above decomposition is unique. It is shown in [15] that the projection of the vector s onto \mathcal{K}^ν can be written as

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

where $[\lambda]_+ := \max\{\lambda, 0\}$ for a scalar $\lambda \in \mathfrak{R}$. Moreover, if $s \in \mathcal{K}^\nu$, there exists a unique vector $s^{1/2} \in \mathcal{K}^\nu$ such that $(s^{1/2})^2 = s$. Thus, the functions defined in (4) and (5) can be represented as

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

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

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

and

$$\phi_{\text{NR}}(s, t) = s - [s - t]_+ = s - ([\lambda_1]_+ u^1 + [\lambda_2]_+ u^2), \quad (6)$$

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

$$\lambda_i := s_1 - t_1 + (-1)^i \|s_2 - t_2\|, \quad (7)$$

$$u^i := \begin{cases} \frac{1}{2} \left(1, (-1)^i \frac{s_2 - t_2}{\|s_2 - t_2\|}\right) & \text{if } s_2 \neq t_2, \\ \frac{1}{2} \left(1, (-1)^i w\right) & \text{if } s_2 = t_2, \end{cases} \quad (8)$$

with $w \in \mathfrak{R}^{\nu-1}$ being an arbitrary unit vector. Note that both ϕ_{FB} and ϕ_{NR} are locally Lipschitz continuous but not differentiable everywhere [15].

Denote by $x := (x^1, \dots, x^m) \in \mathfrak{R}^{n_1} \times \dots \times \mathfrak{R}^{n_m}$ and $y := (y^1, \dots, y^m) \in \mathfrak{R}^{n_1} \times \dots \times \mathfrak{R}^{n_m}$. By means of the SOC complementarity function (3), the SOCCP (1) can be easily reformulated as nonlinear equations

$$\Phi(x, y) := \begin{bmatrix} \phi(x^1, y^1) \\ \vdots \\ \phi(x^m, y^m) \end{bmatrix} = 0, \quad F(x, y, z) = 0 \quad (9)$$

and, along this approach, some Newton-type methods have been developed for solving (1) successively. Another approach is to reformulate (1) as an optimization problem

$$\min_{(x,y,z)} \|\Phi(x,y)\|^2 + \|F(x,y,z)\|^2$$

and some descent methods based on this approach are presented for solving (1). See [5] for more details about recent developments on SOCCP.

In this paper, we consider the following stochastic SOCCP (SSOCCP): Find vectors $x, y \in \mathfrak{R}^n$ and $z \in \mathfrak{R}^l$ such that

$$x \in \mathcal{K}, \quad y \in \mathcal{K}, \quad x^T y = 0, \quad F(x, y, z, \xi) = 0 \quad a.e. \xi \in \Omega, \quad (10)$$

where Ω denotes the support of the random variable ξ , $F : \mathfrak{R}^n \times \mathfrak{R}^n \times \mathfrak{R}^l \times \Omega \rightarrow \mathfrak{R}^n \times \mathfrak{R}^l$, and *a.e.* is the abbreviation for ‘‘almost every’’. This problem is obviously a generalization of the stochastic complementarity problem (SCP)

$$x \geq 0, \quad \mathbf{F}(x, \xi) \geq 0, \quad x^T \mathbf{F}(x, \xi) = 0 \quad a.e. \xi \in \Omega, \quad (11)$$

which has been extensively studied in the recent literature. See [18] for more details about the existing reformulations, numerical methods, and applications of (11).

Consider the stochastic optimization problem

$$\begin{aligned} \min \quad & f(u) \\ \text{s.t.} \quad & h(u, \xi) = 0, \quad g(u, \xi) \leq 0 \quad a.e. \xi \in \Omega, \end{aligned} \quad (12)$$

where the objective may involve expectations or variances. This problem has many practical applications such as water management in cooling-constrained power plants [25], homogeneous product market [30], etc. Note that the problems with stochastic dominance constraints and the two-stage stochastic programs with recourse can be rewritten in a generalized form of (12); see [11, 24]. In addition, methodologies for some special cases of (12) are also considered; see, e.g., [26, 30].

If some component functions of $g(\cdot, \xi)$ are SOC-representable [19], then (12) can be rewritten as a second-order-cone programming problem

$$\begin{aligned} \min \quad & f(u) \\ \text{s.t.} \quad & h(u, \xi) = 0, \quad H(x, u, \xi) = 0 \quad a.e. \xi \in \Omega, \\ & x \in \mathcal{K}. \end{aligned} \quad (13)$$

Recall that the SOC-representable functions include linear functions, convex quadratic functions, fractional quadratic functions, etc.; see, e.g., [2, 19] for more details. Note that the equality constraints in (13) can be rewritten as

$$\begin{aligned} \tilde{h}(u) &:= \mathbb{E}_\xi[h(u, \xi) \bullet h(u, \xi)] = 0, \\ \tilde{H}(x, u) &:= \mathbb{E}_\xi[H(x, u, \xi) \bullet H(x, u, \xi)] = 0, \end{aligned}$$

where \mathbb{E}_ξ denotes the expectation operator and \bullet denotes the Hadamard product. Then, the Karush-Kuhn-Tucker system of problem (13) is

$$\begin{aligned} \nabla f(u) + \nabla \tilde{h}(u)\lambda + \nabla_u \tilde{H}(x, u)\mu &= 0, \\ \nabla_x \tilde{H}(x, u)\mu - y &= 0, \\ h(u, \xi) = 0, H(x, u, \xi) &= 0 \quad a.e. \xi \in \Omega, \\ x \in \mathcal{K}, y \in \mathcal{K}, x^T y &= 0, \end{aligned}$$

which is in the form of (10). This is one motivation to study the SSOCCP (10) in this paper. Besides, instead of rewriting (12) into (13) by the above SOC representable approach, our research is also inspired by some well-established practical engineering SOC programming problems, as shown in Subsections 4.1 and 4.2. Specifically, engineers usually use the SOC convexification to attack malignant nonconvexity in practice and it is interesting that the SOC relaxation may be exact with physical explanation in some cases (for instance, in Subsection 4.2, we study a circuit network of tree topology in which the alternating current optimal power flow (AC-OPF) admits an exact SOC relaxation). When uncertainty occurs (for example, the renewable resource is involved; see Subsection 4.2), problem (13) appears naturally and motivates us to explore the SSOCCP (10) as well.

However, because of the existence of the random element ξ , we generally cannot expect that there exist vectors $\{x, y, z\}$ satisfying (10) for almost every $\xi \in \Omega$. This means that the SSOCCP (10) may not have a solution in general. Therefore, in order to get reasonable solutions in some senses, we need to present appropriately deterministic formulations for (10). In this paper, we mainly consider a deterministic formulation called expected residual minimization (ERM) formulation for (10), which is motivated by the work [7] on the SCP (11). An approximation method based on the Monte Carlo approximation techniques for solving the ERM model are proposed in Section 2 and some properties related to existence of solutions of the ERM model are discussed in Section 3. Then, in Section 4, we report the modeling effectiveness and computational efficiency of our investigations within the framework of two practical engineering settings, namely, a stochastic natural gas transmission problem and a stochastic optimal power flow problem in radial network.

Throughout this paper, we assume that the support set Ω is a compact set with infinite number of elements in a finite dimensional Euclidean space and $F(x, y, z, \xi)$ is twice continuously differentiable with respect to (x, y, z) and continuously integrable with respect to ξ . For a given differentiable function $H : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and a vector $x \in \mathbb{R}^n$, $\nabla H(x)$ denotes the transposed Jacobian of H at x . Given a vector $x \in \mathbb{R}^n$ and a set $X \subseteq \mathbb{R}^n$, $\text{dist}(x, X)$ denotes the distance from x to X under the Euclidean norm. For a given $m \times n$ matrix $A = (a_{ij})$, $\|A\|_{\mathcal{F}}$ denotes its Frobenius norm, that is, $\|A\|_{\mathcal{F}} := (\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2)^{1/2}$. Moreover, I and O denote the identity matrix and null matrix with suitable dimensions, respectively, and $\text{co}\{X\}$ denotes the convex hull of a set X . In addition, we use $\text{sign}(\cdot)$ to stand for the sign function.

2 ERM formulation for SSOCCP

As introduced in Section 1, by means of the SOC complementarity function (3), the SSOCCP (10) can be reformulated as the stochastic nonlinear equations

$$\Phi(x, y) = 0, \quad F(x, y, z, \xi) = 0 \quad a.e. \xi \in \Omega,$$

where Φ is given as in (9). Recall that the above stochastic equations may not have a common solution in general. Motivated by the work [7] on the SCP (11), we propose an ERM formulation for (10) as

$$\min_{(x, y, z)} \theta_{\text{ERM}}(x, y, z) := \mathbb{E}_{\xi} [\|F(x, y, z, \xi)\|^2] + \|\Phi(x, y)\|^2. \quad (14)$$

One main difficulty in dealing with (14) is that the problem contains an expectation, which may have no analytical expression in general. We can employ the Monte Carlo sampling techniques to approximate the expectation. Another possible difficulty is that the SOC complementarity function ϕ is generally not differentiable everywhere and so the objective function may be nonsmooth. But this is not always the case. For the two SOC complementarity functions introduced in Section 1, $\|\phi_{\text{FB}}\|^2$ is actually a smooth function although ϕ_{FB} is nonsmooth, while $\|\phi_{\text{NR}}\|^2$ and ϕ_{NR} are both nonsmooth functions.

As is known to us, the functions ϕ_{FB} and ϕ_{NR} are generalizations of the classical real-valued complementarity functions

$$\varphi_{\text{FB}}(a, b) := a + b - \sqrt{a^2 + b^2}, \quad (a, b) \in \mathfrak{R}^2$$

and

$$\varphi_{\text{min}}(a, b) := \min\{a, b\}, \quad (a, b) \in \mathfrak{R}^2$$

respectively. Similarly to their prototypes, compared with each other, ϕ_{FB} owns better smoothing property and ϕ_{NR} has better approximation property. In particular, since $\phi_{\text{NR}}(s, t) = s - [s - t]_+ = t - [t - s]_+$ for any s and t in \mathfrak{R}^{ν} , we have

$$\phi_{\text{NR}}(s, t) = \begin{cases} s & \text{if } t - s \in \mathcal{K}^{\nu}, \\ t & \text{if } s - t \in \mathcal{K}^{\nu}, \end{cases}$$

while there always exists a positive gap between $\phi_{\text{FB}}(s, t)$ and either s or t . This advantage that ϕ_{NR} possesses may be particularly useful in dealing with the SOCCP

$$G(x) \in \mathcal{K}, \quad H(x) \in \mathcal{K}, \quad G(x)^T H(x) = 0.$$

Our numerical experience reported in Section 4 also reveals that ϕ_{NR} may have better performance even though one fairly small smoothing parameter is involved. Further comparison between these two complementarity functions are given in Section 3.

In general, for an integrable function $\psi : \Omega \rightarrow \mathfrak{R}$, the Monte Carlo sampling estimate for $\mathbb{E}_{\xi}[\psi(\xi)]$ is obtained by taking independently and identically distributed random samples $\Omega_k := \{\xi^1, \dots, \xi^{N_k}\}$ 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,

$$\lim_{k \rightarrow \infty} \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \psi(\xi^i) = \mathbb{E}_\xi[\psi(\xi)] \quad \text{w.p.1.} \quad (15)$$

In what follows, we consider two cases where ϕ is taken to be ϕ_{FB} and ϕ_{NR} in (14) respectively.

2.1 The case of ϕ_{FB}

Consider the smooth ERM model

$$\min_{(x,y,z)} \theta_{\text{FB}}(x, y, z) := \mathbb{E}_\xi [\|F(x, y, z, \xi)\|^2] + \|\Phi_{\text{FB}}(x, y)\|^2, \quad (16)$$

where Φ_{FB} denotes the function Φ given in (9) by taking ϕ to be ϕ_{FB} . By generating independently and identically distributed random samples $\Omega_k = \{\xi^1, \dots, \xi^{N_k}\}$ from Ω , we can obtain the following approximation of (16):

$$\min_{(x,y,z)} \theta_{\text{FB}}^k(x, y, z) := \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \|F(x, y, z, \xi^i)\|^2 + \|\Phi_{\text{FB}}(x, y)\|^2. \quad (17)$$

We next study the convergence of the above sample average approximation method. Since (17) is a nonconvex optimization problem, we only investigate the limiting behavior of its stationary points here. Actually, similar convergence result for its optimal solutions can be obtained more easily.

Theorem 1 *Suppose that (x^k, y^k, z^k) is a stationary point of problem (17) 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 (16) with probability one.*

Proof Without loss of generality, we may assume $\lim_{k \rightarrow \infty} (x^k, y^k, z^k) = (\bar{x}, \bar{y}, \bar{z})$. Let B be a compact and convex set containing the whole sequence $\{(x^k, y^k, z^k)\}$. By the continuity of F , $\nabla_{(x,y,z)} F$, and $\nabla_{(x,y,z)}^2 F_j$ ($j = 1, \dots, n+l$) on the compact set $B \times \Omega$, there exists a constant $\bar{C} > 0$ such that

$$\|F(x, y, z, \xi)\| \leq \bar{C}, \quad \|\nabla_{(x,y,z)} F(x, y, z, \xi)\|_{\mathcal{F}} \leq \bar{C}, \quad (18)$$

$$\|\nabla_{(x,y,z)}^2 F_j(x, y, z, \xi)\|_{\mathcal{F}} \leq \bar{C} \quad (j = 1, \dots, n+l) \quad (19)$$

hold for every $(x, y, z, \xi) \in B \times \Omega$. Let

$$\Psi_{\text{FB}}(x, y) := \|\Phi_{\text{FB}}(x, y)\|^2.$$

By Proposition 2 of [6], Ψ_{FB} is smooth, that is, $\nabla \Psi_{\text{FB}}$ is continuous everywhere. For each k , since (x^k, y^k, z^k) is stationary to problem (17), we have

$$\frac{2}{N_k} \sum_{\xi^i \in \Omega_k} \nabla_{(x,y,z)} F(x^k, y^k, z^k, \xi^i) F(x^k, y^k, z^k, \xi^i) + \begin{bmatrix} \nabla \Psi_{\text{FB}}(x^k, y^k) \\ 0 \end{bmatrix} = 0. \quad (20)$$

Consider the first term in the left-hand of (20). For each k and each $j = 1, \dots, n+l$, we have

$$\begin{aligned}
& \left| \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \nabla_{(x,y,z)} F_j(x^k, y^k, z^k, \xi^i)^T F(x^k, y^k, z^k, \xi^i) \right. \\
& \quad \left. - \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \nabla_{(x,y,z)} F_j(\bar{x}, \bar{y}, \bar{z}, \xi^i)^T F(\bar{x}, \bar{y}, \bar{z}, \xi^i) \right| \\
& \leq \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \|\nabla_{(x,y,z)} F_j(x^k, y^k, z^k, \xi^i)\| \|F(x^k, y^k, z^k, \xi^i) - F(\bar{x}, \bar{y}, \bar{z}, \xi^i)\| \\
& \quad + \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \|\nabla_{(x,y,z)} F_j(x^k, y^k, z^k, \xi^i) - \nabla_{(x,y,z)} F_j(\bar{x}, \bar{y}, \bar{z}, \xi^i)\| \|F(\bar{x}, \bar{y}, \bar{z}, \xi^i)\| \\
& \leq \frac{\bar{C}}{N_k} \sum_{\xi^i \in \Omega_k} \int_0^1 \left(\|\nabla_{(x,y,z)} F(t x^k + (1-t)\bar{x}, t y^k + (1-t)\bar{y}, t z^k + (1-t)\bar{z}, \xi^i)\|_{\mathcal{F}} \right. \\
& \quad \left. + \|\nabla_{(x,y,z)}^2 F_j(t x^k + (1-t)\bar{x}, t y^k + (1-t)\bar{y}, t z^k + (1-t)\bar{z}, \xi^i)\|_{\mathcal{F}} \right) \\
& \quad \times \|(x^k, y^k, z^k) - (\bar{x}, \bar{y}, \bar{z})\| dt \\
& \leq 2\bar{C}^2 \|(x^k, y^k, z^k) - (\bar{x}, \bar{y}, \bar{z})\| \\
& \rightarrow 0 \quad \text{as } k \rightarrow +\infty,
\end{aligned}$$

where the second inequality follows from the mean-value theorem and (18), the third inequality follows from (18)–(19). We then have from (15) that

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

holds with probability one. Moreover, by (18), we have

$$\|\nabla_{(x,y,z)} F(x, y, z, \xi) F(x, y, z, \xi)\| \leq \bar{C}^2, \quad (x, y, z, \xi) \in B \times \Omega$$

and hence, by Theorem 16.8 of [21],

$$\lim_{k \rightarrow \infty} \frac{2}{N_k} \sum_{\xi^i \in \Omega_k} \nabla_{(x,y,z)} F(x^k, y^k, z^k, \xi^i) F(x^k, y^k, z^k, \xi^i) = \nabla \mathbb{E}_{\xi} [\|F(\bar{x}, \bar{y}, \bar{z}, \xi)\|^2]$$

holds with probability one. Thus, letting $k \rightarrow +\infty$ in (20), we have

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

which means that $(\bar{x}, \bar{y}, \bar{z})$ is a stationary point of (16) with probability one.

2.2 The case of ϕ_{NR}

Consider the nonsmooth ERM model

$$\min_{(x,y,z)} \theta_{\text{NR}}(x,y,z) := \mathbb{E}_{\xi} [\|F(x,y,z,\xi)\|^2] + \|\Phi_{\text{NR}}(x,y)\|^2, \quad (21)$$

where Φ_{NR} denotes the function Φ given in (9) by taking ϕ to be ϕ_{NR} . To deal with this model, besides the Monte Carlo sampling techniques, some smoothing skills are also necessary. Here, we employ the following smoothing approximation presented in [15] for the function ϕ_{NR} defined in (5): Given a scalar $\mu > 0$, let

$$\phi_{\text{NR}}^{\mu}(s,t) := s - \frac{1}{2} \left((\sqrt{\lambda_1^2 + 4\mu^2} + \lambda_1)u^1 + (\sqrt{\lambda_2^2 + 4\mu^2} + \lambda_2)u^2 \right),$$

where $\{\lambda_1, \lambda_2\}$ and $\{u^1, u^2\}$ are the same as in (7) and (8) respectively. It is shown in [15] that, for each $(s,t) \in \mathfrak{R}^{2\nu}$,

$$\lim_{\mu \rightarrow +0} \phi_{\text{NR}}^{\mu}(s,t) = \phi_{\text{NR}}(s,t)$$

and ϕ_{NR}^{μ} is a smooth function with

$$\nabla \phi_{\text{NR}}^{\mu}(s,t) = \begin{bmatrix} I - M_{\mu}(s,t) \\ M_{\mu}(s,t) \end{bmatrix}, \quad (22)$$

where

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

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

$$a_{\mu}(s,t) := \frac{s_1-t_1}{2\sqrt{(s_1-t_1)^2+4\mu^2}} + \frac{1}{2}, \quad (23)$$

$$b_{\mu}(s,t) := \frac{s_1-t_1-\|s_2-t_2\|}{4\sqrt{(s_1-t_1-\|s_2-t_2\|)^2+4\mu^2}} + \frac{s_1-t_1+\|s_2-t_2\|}{4\sqrt{(s_1-t_1+\|s_2-t_2\|)^2+4\mu^2}} + \frac{1}{2}, \quad (24)$$

$$c_{\mu}(s,t) := \frac{s_1-t_1}{\sqrt{(s_1-t_1-\|s_2-t_2\|)^2+4\mu^2} + \sqrt{(s_1-t_1+\|s_2-t_2\|)^2+4\mu^2}} + \frac{1}{2}, \quad (25)$$

$$d_{\mu}(s,t) := -\frac{s_1-t_1-\|s_2-t_2\|}{4\sqrt{(s_1-t_1-\|s_2-t_2\|)^2+4\mu^2}} + \frac{s_1-t_1+\|s_2-t_2\|}{4\sqrt{(s_1-t_1+\|s_2-t_2\|)^2+4\mu^2}}. \quad (26)$$

In addition, from the proof of Proposition 5.1 in [15], it is not difficult to see that there exists a positive constant C such that

$$\|\phi_{\text{NR}}^{\mu}(s,t) - \phi_{\text{NR}}(s,t)\| \leq C\mu \quad (27)$$

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

Taking a smoothing parameter $\mu_k > 0$ and independently and identically distributed random samples $\Omega_k := \{\xi^1, \dots, \xi^{N_k}\}$ from Ω , we can obtain the following smooth approximation of (21):

$$\min_{(x,y,z)} \theta_{\text{NR}}^k(x, y, z) := \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \|F(x, y, z, \xi^i)\|^2 + \|\Phi_{\text{NR}}^{\mu_k}(x, y)\|^2, \quad (28)$$

where

$$\Phi_{\text{NR}}^{\mu}(x, y) := \begin{bmatrix} \phi_{\text{NR}}^{\mu}(x^1, y^1) \\ \vdots \\ \phi_{\text{NR}}^{\mu}(x^m, y^m) \end{bmatrix}$$

for $x = (x^1, \dots, x^m) \in \mathfrak{R}^{n_1} \times \dots \times \mathfrak{R}^{n_m}$ and $y = (y^1, \dots, y^m) \in \mathfrak{R}^{n_1} \times \dots \times \mathfrak{R}^{n_m}$. Suppose that $\mu_k \rightarrow +0$ as $k \rightarrow +\infty$. We next study the convergence of the above sample average approximation method. As mentioned in the last subsection, we only need to investigate the limiting behavior of stationary points of problem (28) because similar convergence result for optimal solutions can be obtained more easily. To this end, the following definitions are useful.

Definition 1 [10] Let $H : \mathfrak{R}^p \rightarrow \mathfrak{R}^q$ be locally Lipschitz continuous. The *Clarke generalized gradient* of H at w is defined as

$$\partial H(w) := \text{co} \left\{ \lim_{w' \rightarrow w, w' \in D_H} \nabla H(w') \right\},$$

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

Definition 2 [17] Let $H : \mathfrak{R}^p \rightarrow \mathfrak{R}^q$ be locally Lipschitz continuous and $H^{\mu} : \mathfrak{R}^p \rightarrow \mathfrak{R}^q$ be a function such that H^{μ} is continuously differentiable everywhere for any $\mu > 0$ and $\lim_{\mu \rightarrow +0} H^{\mu}(w) = H(w)$ for any $w \in \mathfrak{R}^p$. We say that H^{μ} satisfies the *Jacobian consistency* with H if

$$\lim_{\mu \rightarrow +0} \text{dist}(\nabla H^{\mu}(w), \partial H(w)) = 0$$

holds for any $w \in \mathfrak{R}^p$.

For simplicity, we denote by

$$\Psi_{\text{NR}}(x, y) := \|\Phi_{\text{NR}}(x, y)\|^2, \quad \Psi_{\text{NR}}^{\mu}(x, y) := \|\Phi_{\text{NR}}^{\mu}(x, y)\|^2,$$

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

$$\Psi_{\text{NR}}^i(x^i, y^i) := \|\phi_{\text{NR}}(x^i, y^i)\|^2, \quad \Psi_{\text{NR}}^{\mu, i}(x^i, y^i) := \|\phi_{\text{NR}}^{\mu}(x^i, y^i)\|^2$$

for each i . Then we have

$$\partial \Psi_{\text{NR}}(x, y) = \partial \Psi_{\text{NR}}^1(x^1, y^1) \times \dots \times \partial \Psi_{\text{NR}}^m(x^m, y^m) \quad (29)$$

and

$$\nabla \Psi_{\text{NR}}^\mu(x, y) = \begin{bmatrix} \nabla \Psi_{\text{NR}}^{\mu,1}(x^1, y^1) \\ \vdots \\ \nabla \Psi_{\text{NR}}^{\mu,m}(x^m, y^m) \end{bmatrix}. \quad (30)$$

By Theorem 4.9 of [17], we have the following lemma immediately.

Lemma 1 *The function Ψ_{NR}^μ satisfies the Jacobian consistency with Ψ_{NR} .*

We next show the main convergence result of this subsection.

Theorem 2 *Suppose that (x^k, y^k, z^k) is a stationary point of problem (28) 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 (21) with probability one.*

Proof Without loss of generality, we may assume $\lim_{k \rightarrow \infty} (x^k, y^k, z^k) = (\bar{x}, \bar{y}, \bar{z})$. Let B and $\bar{C} > 0$ be the same as in the proof of Theorem 1. For each k , since (x^k, y^k, z^k) is stationary to problem (28), we have

$$\frac{2}{N_k} \sum_{\xi^i \in \Omega_k} \nabla_{(x,y,z)} F(x^k, y^k, z^k, \xi^i) F(x^k, y^k, z^k, \xi^i) + \begin{bmatrix} \nabla \Psi_{\text{NR}}^{\mu_k}(x^k, y^k) \\ 0 \end{bmatrix} = 0. \quad (31)$$

In a similar way to the proof of Theorem 1, we can show that

$$\begin{aligned} & \lim_{k \rightarrow \infty} \frac{2}{N_k} \sum_{\xi^i \in \Omega_k} \nabla_{(x,y,z)} F(x^k, y^k, z^k, \xi^i) F(x^k, y^k, z^k, \xi^i) \\ &= \nabla \mathbb{E}_\xi [\|F(\bar{x}, \bar{y}, \bar{z}, \xi)\|^2] \end{aligned} \quad (32)$$

holds with probability one. We next show

$$\lim_{k \rightarrow \infty} \text{dist}(\nabla \Psi_{\text{NR}}^{\mu_k}(x^k, y^k), \partial \Psi_{\text{NR}}(\bar{x}, \bar{y})) = 0.$$

Denote by $x^k := (x^{k,1}, \dots, x^{k,m}) \in \mathfrak{R}^{n_1} \times \dots \times \mathfrak{R}^{n_m}$, $y^k := (y^{k,1}, \dots, y^{k,m}) \in \mathfrak{R}^{n_1} \times \dots \times \mathfrak{R}^{n_m}$ for each k and by $\bar{x} := (\bar{x}^1, \dots, \bar{x}^m) \in \mathfrak{R}^{n_1} \times \dots \times \mathfrak{R}^{n_m}$, $\bar{y} := (\bar{y}^1, \dots, \bar{y}^m) \in \mathfrak{R}^{n_1} \times \dots \times \mathfrak{R}^{n_m}$. From (29) and (30), it is sufficient to show that, for each i ,

$$\lim_{k \rightarrow \infty} \text{dist}(\nabla \Psi_{\text{NR}}^{\mu_k,i}(x^{k,i}, y^{k,i}), \partial \Psi_{\text{NR}}^i(\bar{x}^i, \bar{y}^i)) = 0. \quad (33)$$

First of all, for any given i , we have

$$\begin{aligned} & \|\phi_{\text{NR}}^{\mu_k}(x^{k,i}, y^{k,i}) - \phi_{\text{NR}}^{\mu_k}(\bar{x}^i, \bar{y}^i)\| \\ & \leq \|\phi_{\text{NR}}^{\mu_k}(x^{k,i}, y^{k,i}) - \phi_{\text{NR}}(x^{k,i}, y^{k,i})\| + \|\phi_{\text{NR}}(x^{k,i}, y^{k,i}) - \phi_{\text{NR}}(\bar{x}^i, \bar{y}^i)\| \\ & \quad + \|\phi_{\text{NR}}(\bar{x}^i, \bar{y}^i) - \phi_{\text{NR}}^{\mu_k}(\bar{x}^i, \bar{y}^i)\| \\ & \leq 2C\mu_k + \|\phi_{\text{NR}}(x^{k,i}, y^{k,i}) - \phi_{\text{NR}}(\bar{x}^i, \bar{y}^i)\| \\ & \rightarrow 0 \quad \text{as } k \rightarrow +\infty, \end{aligned} \quad (34)$$

where the second inequality follows from (27). We consider five cases:

- (I) Suppose that $\bar{x}^i = \bar{y}^i$. It follows that $\bar{\lambda}_j = \bar{x}_1^i - \bar{y}_1^i + (-1)^j \|\bar{x}_2^i - \bar{y}_2^i\| = 0$ for $j = 1, 2$. By Theorem 4.6 and Proposition 4.8 of [17], we have

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

where

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

It is not difficult to see from (23)–(26) that any accumulation points of the sequences

$$\{a_{\mu_k}(x^{k,i}, y^{k,i})\}, \quad \{b_{\mu_k}(x^{k,i}, y^{k,i})\}, \quad \{c_{\mu_k}(x^{k,i}, y^{k,i})\}, \quad \{d_{\mu_k}(x^{k,i}, y^{k,i})\}$$

belong to the intervals $[0, 1], [0, 1], [0, 1], [0, \frac{1}{2}]$ respectively. Therefore, any accumulation point of $\{M_{\mu_k}(x^{k,i}, y^{k,i})\}$ must belong to $\text{co}(O, I, S)$. On the other hand, it is easy to see from (27) and (34) that $\lim_{k \rightarrow \infty} \phi_{\text{NR}}^{\mu_k}(x^{k,i}, y^{k,i}) = \phi_{\text{NR}}(\bar{x}^i, \bar{y}^i)$. Therefore, by (22), we can obtain (33) easily.

- (II) Suppose that $\bar{x}^i \neq \bar{y}^i$ and $\bar{x}_2^i - \bar{y}_2^i = 0$. Since $\{\nabla \phi_{\text{NR}}^{\mu_k}(x^{k,i}, y^{k,i})\}$ is bounded, we have from (34) that

$$\lim_{k \rightarrow \infty} \nabla \phi_{\text{NR}}^{\mu_k}(x^{k,i}, y^{k,i}) (\phi_{\text{NR}}^{\mu_k}(x^{k,i}, y^{k,i}) - \phi_{\text{NR}}^{\mu_k}(\bar{x}^i, \bar{y}^i)) = 0. \quad (35)$$

On the other hand, it is easy to see from (23)–(26) that

$$\begin{aligned} \lim_{k \rightarrow \infty} a_{\mu_k}(x^{k,i}, y^{k,i}) &= \lim_{k \rightarrow \infty} b_{\mu_k}(x^{k,i}, y^{k,i}) = \lim_{k \rightarrow \infty} c_{\mu_k}(x^{k,i}, y^{k,i}) \\ &= \frac{1}{2} \text{sign}(\bar{x}_1^i - \bar{y}_1^i) + \frac{1}{2} \end{aligned}$$

and $\lim_{k \rightarrow \infty} d_{\mu_k}(x^{k,i}, y^{k,i}) = 0$, which means

$$\lim_{k \rightarrow \infty} (M_{\mu_k}(x^{k,i}, y^{k,i}) - M_{\mu_k}(\bar{x}^i, \bar{y}^i)) = 0 \quad (36)$$

and hence

$$\lim_{k \rightarrow \infty} (\nabla \phi_{\text{NR}}^{\mu_k}(x^{k,i}, y^{k,i}) - \nabla \phi_{\text{NR}}^{\mu_k}(\bar{x}^i, \bar{y}^i)) = 0. \quad (37)$$

Since $\{\phi_{\text{NR}}^{\mu_k}(\bar{x}^i, \bar{y}^i)\}$ is bounded, we have

$$\lim_{k \rightarrow \infty} (\nabla \phi_{\text{NR}}^{\mu_k}(x^{k,i}, y^{k,i}) - \nabla \phi_{\text{NR}}^{\mu_k}(\bar{x}^i, \bar{y}^i)) \phi_{\text{NR}}^{\mu_k}(\bar{x}^i, \bar{y}^i) = 0. \quad (38)$$

Therefore, it follows from (35) and (38) that

$$\begin{aligned} & \frac{1}{2} \|\nabla \Psi_{\text{NR}}^{\mu_k, i}(x^{k,i}, y^{k,i}) - \nabla \Psi_{\text{NR}}^{\mu_k, i}(\bar{x}^i, \bar{y}^i)\| \\ &= \|\nabla \phi_{\text{NR}}^{\mu_k}(x^{k,i}, y^{k,i}) \phi_{\text{NR}}^{\mu_k}(x^{k,i}, y^{k,i}) - \nabla \phi_{\text{NR}}^{\mu_k}(\bar{x}^i, \bar{y}^i) \phi_{\text{NR}}^{\mu_k}(\bar{x}^i, \bar{y}^i)\| \\ &\leq \|\nabla \phi_{\text{NR}}^{\mu_k}(x^{k,i}, y^{k,i}) (\phi_{\text{NR}}^{\mu_k}(x^{k,i}, y^{k,i}) - \phi_{\text{NR}}^{\mu_k}(\bar{x}^i, \bar{y}^i))\| \\ &\quad + \|(\nabla \phi_{\text{NR}}^{\mu_k}(x^{k,i}, y^{k,i}) - \nabla \phi_{\text{NR}}^{\mu_k}(\bar{x}^i, \bar{y}^i)) \phi_{\text{NR}}^{\mu_k}(\bar{x}^i, \bar{y}^i)\| \\ &\rightarrow 0 \quad \text{as } k \rightarrow +\infty. \end{aligned}$$

By Lemma 1, we have

$$\lim_{k \rightarrow \infty} \text{dist}(\nabla \Psi_{\text{NR}}^{\mu_k, i}(\bar{x}^i, \bar{y}^i), \partial \Psi_{\text{NR}}^i(\bar{x}^i, \bar{y}^i)) = 0$$

and hence (33) holds.

(III) Suppose that $\bar{x}_2^i - \bar{y}_2^i \neq 0$ and $|\bar{x}_1^i - \bar{y}_1^i| \neq \|\bar{x}_2^i - \bar{y}_2^i\|$. From (23)–(26), we have

$$\begin{aligned} \lim_{k \rightarrow \infty} (a_{\mu_k}(x^{k,i}, y^{k,i}) - a_{\mu_k}(\bar{x}^i, \bar{y}^i)) &= \lim_{k \rightarrow \infty} (b_{\mu_k}(x^{k,i}, y^{k,i}) - b_{\mu_k}(\bar{x}^i, \bar{y}^i)) = 0, \\ \lim_{k \rightarrow \infty} (c_{\mu_k}(x^{k,i}, y^{k,i}) - c_{\mu_k}(\bar{x}^i, \bar{y}^i)) &= \lim_{k \rightarrow \infty} (d_{\mu_k}(x^{k,i}, y^{k,i}) - d_{\mu_k}(\bar{x}^i, \bar{y}^i)) = 0, \end{aligned}$$

which implies (36) and hence (37). In a similar way to (II), we can get (33).

(IV) Suppose that $\bar{x}_2^i - \bar{y}_2^i \neq 0$ and $\bar{x}_1^i - \bar{y}_1^i = \|\bar{x}_2^i - \bar{y}_2^i\|$. It follows that $\bar{\lambda}_1 = 0$ and $\bar{\lambda}_2 > 0$, where $\{\bar{\lambda}_1, \bar{\lambda}_2\}$ are the same as in (I). By Theorem 4.6 and Proposition 4.8 of [17], we have

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

where

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

Note that, by (23)–(26), any accumulation points of the sequences

$$\{b_{\mu_k}(x^{k,i}, y^{k,i})\}, \quad \{c_{\mu_k}(x^{k,i}, y^{k,i})\}, \quad \{d_{\mu_k}(x^{k,i}, y^{k,i})\}$$

belong to $[\frac{1}{2}, 1]$, $[\frac{1}{2}, 1]$, $[0, \frac{1}{2}]$ respectively. It is not difficult to see that any accumulation point of $\{M_{\mu_k}(x^{k,i}, y^{k,i})\}$ must belong to $\text{co}(I, I + Z)$. Since $\lim_{k \rightarrow \infty} \phi_{\text{NR}}^{\mu_k}(x^{k,i}, y^{k,i}) = \phi_{\text{NR}}(\bar{x}^i, \bar{y}^i)$, by (22), we can obtain (33) easily.

(V) Suppose that $\bar{x}_2^i - \bar{y}_2^i \neq 0$ and $\bar{x}_1^i - \bar{y}_1^i = -\|\bar{x}_2^i - \bar{y}_2^i\|$. It follows that $\bar{\lambda}_1 < 0$ and $\bar{\lambda}_2 = 0$. By Theorem 4.6 and Proposition 4.8 of [17], we have

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

where

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

In a similar way to (IV), we can show (33).

As a result, (33) holds in all cases. Letting $k \rightarrow +\infty$ in (31), we have from (32) and (33) that

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

which means that $(\bar{x}, \bar{y}, \bar{z})$ is a stationary point of (21) with probability one.

2.3 Exponential convergence rate

In this subsection, we show that, with the increase of sample size, the optimal solutions of the approximation problem (17) or (28) converge exponentially to a solution of the ERM formulation (16) or (21) with probability approaching one under suitable conditions. To this end, we first introduce a lemma.

Lemma 2 [26] *Let W be a compact set and $h : W \times \Omega \rightarrow \mathfrak{R}$ be integrable everywhere. Suppose that the following conditions hold:*

(i) *For every $w \in W$, the moment generating function*

$$\mathbb{E}_\xi \left[e^{t(h(w, \xi) - \mathbb{E}_\xi[h(w, \xi)])} \right]$$

is finite-valued for all t in a neighbourhood of zero.

(ii) *There exist a measurable function $\kappa : \Omega \rightarrow \mathfrak{R}_+$ and a constant $\gamma > 0$ such that*

$$|h(w', \xi) - h(w, \xi)| \leq \kappa(\xi) \|w' - w\|^\gamma$$

for all $\xi \in \Omega$ and all $w', w \in W$.

(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

$$\text{Prob} \left\{ \sup_{w \in W} \left| \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} h(w, \xi^i) - \mathbb{E}_\xi[h(w, \xi)] \right| \geq \varepsilon \right\} \leq D(\varepsilon) e^{-N_k \beta(\varepsilon)}.$$

Applying the above lemma, we can obtain the following result related to the exponential convergence of the above approximation methods.

Theorem 3 *Let (x^k, y^k, z^k) be an optimal solution of (17) or (28) 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*

$$\text{Prob} \left\{ |\theta_{\text{FB}}^k(x^k, y^k, z^k) - \theta_{\text{FB}}(\bar{x}, \bar{y}, \bar{z})| \geq \varepsilon \right\} \leq D(\varepsilon) e^{-N_k \beta(\varepsilon)} \quad (39)$$

or

$$\text{Prob} \left\{ |\theta_{\text{NR}}^k(x^k, y^k, z^k) - \theta_{\text{NR}}(\bar{x}, \bar{y}, \bar{z})| \geq \varepsilon \right\} \leq D(\varepsilon) e^{-N_k \beta(\varepsilon)}. \quad (40)$$

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 (17). We first show that there exist positive constants $D(\varepsilon)$ and $\beta(\varepsilon)$ such that

$$\text{Prob} \left\{ \sup_{(x, y, z) \in B} |\theta_{\text{FB}}^k(x, y, z) - \theta_{\text{FB}}(x, y, z)| \geq \varepsilon \right\} \leq D(\varepsilon) e^{-N_k \beta(\varepsilon)}, \quad (41)$$

which is equivalent to show

$$\begin{aligned} & \text{Prob} \left\{ \sup_{(x,y,z) \in B} \left| \frac{1}{N_k} \sum_{\xi^i \in \Omega_k} \|F(x, y, z, \xi^i)\|^2 - \mathbb{E}_\xi [\|F(x, y, z, \xi)\|^2] \right| \geq \varepsilon \right\} \\ & \leq D(\varepsilon) e^{-N_k \beta(\varepsilon)} \end{aligned}$$

by (16) and (17). To do this, it is sufficient to prove that the set $W := B$ and the function $h(x, y, z, \xi) := \|F(x, y, z, \xi)\|^2$ satisfy the conditions given in Lemma 2. In fact, since both Ω and B are compact, by the continuous differentiability assumption on F given in Section 1, these conditions hold.

Since (x^k, y^k, z^k) is an optimal solution of (17) for each k , in a similar way to Theorem 1, we can show that $(\bar{x}, \bar{y}, \bar{z})$ is an optimal solution of (14). It then follows that

$$\theta_{\text{FB}}^k(x^k, y^k, z^k) \leq \theta_{\text{FB}}^k(\bar{x}, \bar{y}, \bar{z}), \quad \theta_{\text{FB}}(\bar{x}, \bar{y}, \bar{z}) \leq \theta_{\text{FB}}(x^k, y^k, z^k),$$

from which we have

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

and

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

It follows that

$$|\theta_{\text{FB}}^k(x^k, y^k, z^k) - \theta_{\text{FB}}(\bar{x}, \bar{y}, \bar{z})| \leq \sup_{(x,y,z) \in B} |\theta_{\text{FB}}^k(x, y, z) - \theta_{\text{FB}}(x, y, z)|.$$

This together with (41) implies (39).

(2) Consider the case for (28). In this case, it is sufficient to show that the set $W := B \times [0, \mu_0]$ and the function $h(x, y, z, \mu, \xi) := \|F(x, y, z, \xi)\|^2 + \Phi_{\text{NR}}^\mu(x, y)$ satisfy the conditions given in Lemma 2. In fact, this can be guaranteed by the assumptions given at the end of Section 1 and (27). Thus, in a similar manner to (1), we can show (40) easily. This completes the proof.

2.4 Boundedness of level sets

It is well known that the boundedness of the iteration sequences is a desired requirement for various optimization methods. In order to ensure the boundedness of the iteration sequence, we generally investigate the boundedness of level sets. Given a nonnegative number γ , the level set of the ERM model (14) is given as

$$\mathcal{L}_{\text{ERM}}(\gamma) := \{(x, y, z) \in \mathfrak{R}^{2n+l} \mid \theta_{\text{ERM}}(x, y, z) \leq \gamma\}.$$

Note that, if there is some $n_i > 1$ and we take

$$x^k := (\cdots, \underbrace{k, k, 0, \cdots, 0}_{\mathcal{K}^{n_i}}, \cdots)^T, \quad y^k := (\cdots, \underbrace{k, -k, 0, \cdots, 0}_{\mathcal{K}^{n_i}}, \cdots)^T$$

for every k , then both x^k and y^k belong to the cone \mathcal{K} and they are perpendicular to each other. This means $\Phi(x^k, y^k) = 0$ for each k , where Φ is given in (9) with ϕ to be an SOC complementarity function, regardless of whether it takes ϕ_{FB} or ϕ_{NR} . As a result, we can only investigate the conditions on the mapping F to guarantee the boundedness of the level set $\mathcal{L}_{\text{ERM}}(\gamma)$.

Definition 3 [28] The mapping $H : \mathfrak{R}^p \times \Omega \rightarrow \mathfrak{R}^q$ is *locally coercive* if, for any $\{w^k\} \subseteq \mathfrak{R}^p$ satisfying $\|w^k\| \rightarrow +\infty$, we have

$$\text{Prob}\{\lim_{k \rightarrow \infty} \|H(w^k, \xi)\| = +\infty\} > 0.$$

The main result of this subsection can be stated as follows.

Theorem 4 Suppose that F is locally coercive. Then, for any $\gamma \geq 0$, the level set $\mathcal{L}_{\text{ERM}}(\gamma)$ is bounded.

Proof Suppose on the contrary that there exist a constant $\gamma > 0$ and a sequence $\{(x^k, y^k, z^k)\} \subseteq \mathfrak{R}^{2n+l}$ with $\|(x^k, y^k, z^k)\| \rightarrow +\infty$ such that

$$\theta_{\text{ERM}}(x^k, y^k, z^k) \leq \gamma$$

for each k . It follows that

$$\gamma \geq \mathbb{E}_\xi [\|F(x^k, y^k, z^k, \xi)\|^2] \geq (\mathbb{E}_\xi [\|F(x^k, y^k, z^k, \xi)\|])^2 \quad (42)$$

for each k . By the Fatou's lemma [21], we have

$$\mathbb{E}_\xi [\liminf_{k \rightarrow \infty} \|F(x^k, y^k, z^k, \xi)\|] \leq \liminf_{k \rightarrow \infty} \mathbb{E}_\xi [\|F(x^k, y^k, z^k, \xi)\|]. \quad (43)$$

Note that F is locally coercive, which means that

$$\text{Prob}\{\lim_{k \rightarrow \infty} \|F(x^k, y^k, z^k, \xi)\| = +\infty\} > 0.$$

Therefore, the left-hand side in (43) is infinity and hence

$$\liminf_{k \rightarrow \infty} \mathbb{E}_\xi [\|F(x^k, y^k, z^k, \xi)\|] = +\infty,$$

which contradicts (42) and so the level set $\mathcal{L}_{\text{ERM}}(\gamma)$ is bounded for any $\gamma \geq 0$.

3 Comparison between ϕ_{NR} and ϕ_{FB}

As stated at the beginning of Section 2, the function ϕ_{FB} generally owns better smoothing property, while the function ϕ_{NR} has better approximation property. On the other hand, Chen and Fukushima [7] show that, in the one-dimensional case, ϕ_{NR} has some property that ϕ_{FB} does not possess; see Lemma 2.2 and Example 1 given in [7] for details. An interesting and natural question is whether the result can be extended to multi-dimensional cases. We devote to answer this question in this section.

Consider the following special affine SSOCCP:

$$x \in \mathcal{K}, \quad M(\xi)x + q(\xi) \in \mathcal{K}, \quad x^T(M(\xi)x + q(\xi)) = 0 \quad a.e. \xi \in \Omega, \quad (44)$$

where $M : \Omega \rightarrow \mathfrak{R}^{n \times n}$ and $q : \Omega \rightarrow \mathfrak{R}^n$. For (44), it is natural to suggest its ERM formulation as follows:

$$\min_x \theta(x) := \mathbb{E}_\xi [\|\Phi(x, M(\xi)x + q(\xi))\|^2], \quad (45)$$

where Φ is the same as in (9). Then we have the following result for ϕ_{NR} .

Theorem 5 *Let $\Omega := \{\xi^1, \xi^2, \dots, \xi^N\}$ and each $n_i \leq 2$ in (2). Then the solution set of problem (45) with ϕ to be ϕ_{NR} is nonempty.*

Proof First of all, we show that $\|\phi_{\text{NR}}(s, t)\|^2$ is a piecewise quadratic function over polyhedral partitions in $(s, t) \in \mathfrak{R}^{n_i} \times \mathfrak{R}^{n_i}$ when $n_i \leq 2$. In fact, it is evident that $\phi_{\text{NR}}(s, t) = \min\{s, t\}$ is piecewise quadratic when $n_i = 1$. Next, we suppose $n_i = 2$ and consider the following five cases:

(I) Suppose that $s_2 - t_2 = 0$. It is easy to see that

$$\|\phi_{\text{NR}}(s, t)\|^2 = \begin{cases} \|s\|^2 & \text{if } s_1 \leq t_1, s_2 = t_2, \\ \|t\|^2 & \text{if } s_1 \geq t_1, s_2 = t_2. \end{cases}$$

(II) Suppose that $s_1 - t_1 \geq |s_2 - t_2| > 0$. This means $s - t \in \mathcal{K}^2$ and hence

$$\|\phi_{\text{NR}}(s, t)\|^2 = \|t\|^2$$

from the discussion at the beginning of Section 2.

(III) Suppose that $s_1 - t_1 \leq -|s_2 - t_2| < 0$. This means $t - s \in \mathcal{K}^2$ and, similarly to (II), we have

$$\|\phi_{\text{NR}}(s, t)\|^2 = \|s\|^2.$$

(IV) Suppose that $s_2 - t_2 > 0$ and $t_2 - s_2 \leq s_1 - t_1 \leq s_2 - t_2$. By (6)–(8), we have

$$\begin{aligned} \|\phi_{\text{NR}}(s, t)\|^2 &= \left\| s - [s_1 - t_1 - |s_2 - t_2|]_+ \left(\frac{1}{2}, -\frac{s_2 - t_2}{2|s_2 - t_2|} \right)^T \right. \\ &\quad \left. + [s_1 - t_1 + |s_2 - t_2|]_+ \left(\frac{1}{2}, \frac{s_2 - t_2}{2|s_2 - t_2|} \right)^T \right\|^2 \\ &= \frac{1}{4}(s_1 + t_1 + s_2 - t_2)^2 + \frac{1}{4}(-s_1 + t_1 + 3s_2 - t_2)^2. \end{aligned}$$

(V) Suppose that $s_2 - t_2 < 0$ and $s_2 - t_2 \leq s_1 - t_1 \leq t_2 - s_2$. By (6)–(8), we have

$$\begin{aligned} \|\phi_{\text{NR}}(s, t)\|^2 &= \left\| s - [s_1 - t_1 - |s_2 - t_2|]_+ \left(\frac{1}{2}, -\frac{s_2 - t_2}{2|s_2 - t_2|} \right)^T \right. \\ &\quad \left. + [s_1 - t_1 + |s_2 - t_2|]_+ \left(\frac{1}{2}, \frac{s_2 - t_2}{2|s_2 - t_2|} \right)^T \right\|^2 \\ &= \frac{1}{4}(s_1 + t_1 + s_2 - t_2)^2 + \frac{1}{4}(s_1 - t_1 - 3s_2 + t_2)^2. \end{aligned}$$

In summary, from the continuity of the function ϕ_{NR} , we have

$$\|\phi_{\text{NR}}(s, t)\|^2 = \begin{cases} t_1^2 + t_2^2 & \text{if } s_1 - t_1 \geq s_2 - t_2, s_1 - t_1 \geq t_2 - s_2, \\ s_1^2 + s_2^2 & \text{if } t_1 - s_1 \geq s_2 - t_2, t_1 - s_1 \geq t_2 - s_2, \\ \frac{1}{4}(s_1 + t_1 + s_2 - t_2)^2 + \frac{1}{4}(s_1 - t_1 - 3s_2 + t_2)^2 & \text{otherwise,} \end{cases}$$

which indicates that $\|\phi_{\text{NR}}(s, t)\|^2$ is indeed a piecewise quadratic function over polyhedral partitions in (s, t) .

Note that the objective function of (45) can be written as

$$\theta(x) = \sum_{i=1}^m \sum_{j=1}^N p_j \|\phi_{\text{NR}}(x^i, y_j^i)\|^2,$$

where $x := (x^1, \dots, x^m) \in \mathfrak{R}^{n_1} \times \dots \times \mathfrak{R}^{n_m}$, $M(\xi^j)x + q(\xi^j) := (y_j^1, \dots, y_j^m) \in \mathfrak{R}^{n_1} \times \dots \times \mathfrak{R}^{n_m}$ for each $j = 1, \dots, N$, and p_j denotes the probability of the sample ξ^j for each $j = 1, \dots, N$. From the above discussion, it is not difficult to see that the objective function θ is piecewise quadratic function over polyhedral partitions. Since this function is bounded below on whole space, by the Frank-Wolfe theorem given in [14], problem (45) has an optimal solution at least. This completes the proof.

Theorem 5 indicates that Lemma 2.2 in [7] can be extended from one-dimensional cases to two-dimensional cases. It is still an open question whether the lemma can be extended to general cases. Note that, by Example 1 given in [7], the conclusion in Theorem 5 does not hold for the case of ϕ_{FB} .

Remark 1 If we rewrite the affine SSOCCP (44) by introducing an additional variable as the form

$$x \in \mathcal{K}, \quad y \in \mathcal{K}, \quad x^T y = 0, \quad y - M(\xi)x + q(\xi) = 0 \quad \text{a.e. } \xi \in \Omega$$

and consider its ERM model

$$\min_{(x, y)} \mathbb{E}_\xi [\|y - M(\xi)x - q(\xi)\|^2] + \|\Phi(x, y)\|^2 \quad (46)$$

instead of problem (45), then we can get an interesting and understandable result, that is, Theorem 5 remains true without the discrete assumption related to the sample space Ω (we can show this result in a very similar way and so omit its proof here). Our next question is how about the case of ϕ_{FB} ? Note

that Example 1 in [7] can not serve as a counterexample in this case. In fact, for this example, the model (46) becomes

$$\min_{(x,y)} y^2 + 1 + (x + y - \sqrt{x^2 + y^2})^2,$$

which attains its infimum at $x^* = 0$ and $y^* = 0$ obviously. We are not sure whether the model (46) with ϕ to be ϕ_{FB} has similar conclusion as ϕ_{NR} . We would like to leave it as a future work.

4 Applications

The theoretical results given in the previous sections indicate that the SSOC-CP (10) can be solved via the variation scheme (14). In this section, as a further supplement, we consider two practical engineering problems.

4.1 Natural gas production and transportation

Natural gas is one of the fastest-growing energy sources. How to plan the generation and distribution of natural gas in transmission networks is becoming a key issue in gas production and transportation. Recently, a substantial number of studies focus on natural gas transmission optimization problems [22, 23, 27]. The most difficult issue in optimization is from the nonlinear relationship between the transmission flow in a pipeline and the pressures at two ends of the pipeline, and the uncertainty of losses at compressor stations. In this subsection, we implement the ERM framework to solve this type of problems.

Consider an optimization model for a natural gas network¹ consisted by a number of hub-and-spoke subnetworks where every hub is a compressor station with a number of fields and markets connected to and only to it. We denote by $\mathcal{N} := \{1, 2, \dots, N\}$ the set of nodes in the transmission network, which is classified into three subsets: $\mathcal{N}_g \subset \mathcal{N}$ of all field nodes with producers, $\mathcal{N}_s \subset \mathcal{N}$ of all station nodes with compressors by which every gas flow is split into two or more pipelines, and $\mathcal{N}_m \subset \mathcal{N}$ of all market nodes with consumptions. Without loss of generality, we assume that the intersections between any two of the above three subsets are empty. This assumption can be achieved by introducing virtual pipelines. For every node $i \in \mathcal{N}$, we have two sets of nodes connected to it: $\mathcal{I}(i)$, the set of nodes with pipelines going into node i , and $\mathcal{O}(i)$, the set of nodes with pipelines going out node i . It is worth emphasizing that, due to the hub-and-spoke structure, for any $i \in \mathcal{N}_g$, $\mathcal{I}(i)$ is empty and $\mathcal{O}(i)$ is a singleton with its element being the station node to which node i is connected. Similarly, for any $i \in \mathcal{N}_m$, $\mathcal{O}(i)$ is empty and $\mathcal{I}(i)$ is a singleton with its element being the station node from which node i is connected.

¹ We refer readers to [3, 23] for detailed analysis on how to model natural gas production and transmission.

Uncertainties often occur in transmission operations (for example, a small proportion of the gas flow at a station which is tapped off transmission pipelines to provide fuel for the compressors at the station; see, e.g., Section V in [29]). Here, we take the losses at stations as uncertainties and denote by ξ_i the loss at node $i \in \mathcal{N}_s$. We also denote the losses at all compressor stations by a random vector $\xi := (\xi_1, \dots, \xi_S)$, where S is the number of stations in the network.

The transmission network is centrally operated to deliver natural gas for each market. We denote by $p_j(\mathbf{Q}_j)$ the unit price of natural gas in market $j \in \mathcal{N}_m$, where $p_j(\mathbf{Q}_j)$ is a decreasing function with respect to the supply quantity \mathbf{Q}_j into market j . In our study, we consider an extensively investigated price function $p_j(\mathbf{Q}_j) := a_j - b_j \mathbf{Q}_j$ for every $j \in \mathcal{N}_m$. At field nodes, before real transmissions, the producers are contracted for their minimum generations. We use G_i to denote the minimum generations at node i and C_i to denote the unit generation cost of natural gas at field $i \in \mathcal{N}_g$.

Now we introduce the parameters for pipelines. We denote by P_{ij}^{\max} and P_{ij}^{\min} (\bar{P}_{ij}^{\max} and \bar{P}_{ij}^{\min}) the maximum and minimum pressures at the inlet (outlet) end of the pipeline from node i to node j for $i, j \in \mathcal{N}$. Here, we define $P_{ij}^{\max} = P_{ij}^{\min} = 0$ and $\bar{P}_{ij}^{\max} = \bar{P}_{ij}^{\min} = 0$ when i and j are not connected. To generate one unit inlet pressure for the pipeline from i to j , a cost c_{ij} is incurred. In our model, the pressures at the inlet ends of the pipelines connecting from fields to stations are fixed by the production contract and denoted by \hat{P}_{ij} for every $i \in \mathcal{N}_g$ and the corresponding station node $j \in \mathcal{O}(i)$, which is enforced by the production standards. On the other hand, the pressures at the inlet ends of the pipelines connecting to markets from stations are fixed by the supply contracts and denoted by \check{P}_{ji} for every $i \in \mathcal{N}_m$ and the corresponding station node $j \in \mathcal{I}(i)$, which is predetermined based on the market standards.

The nonlinear relationship between the pressure difference at two ends of a pipeline and the quantity of natural gas transmitted through this pipeline can be described via the Weymouth equation. In particular, the flow \mathbf{q}_{ij} for the pipeline connecting nodes i and j is upper bounded by

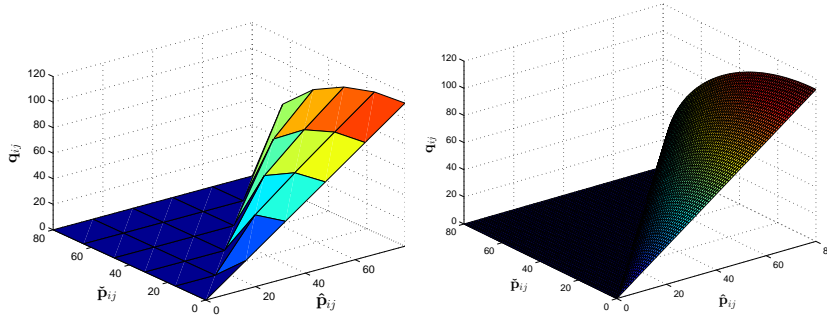
$$q_{ij}^c(\hat{\mathbf{p}}_{ij}, \check{\mathbf{p}}_{ij}) := K_{ij} \sqrt{(\hat{\mathbf{p}}_{ij})^2 - (\check{\mathbf{p}}_{ij})^2}, \quad (47)$$

where $\hat{\mathbf{p}}_{ij}$ and $\check{\mathbf{p}}_{ij}$ are the inlet and outlet pressures, respectively, and K_{ij} is a constant computed from the physical properties of this pipeline such as length, dimension, friction factor, etc.²

² For a particular pipeline, the Weymouth equation can be expressed as (see (1) and (2) in [1] with the adjustment on parameter units)

$$q_{ij}^c(\hat{\mathbf{p}}_{ij}, \check{\mathbf{p}}_{ij}) = 77.52 \times 10^{-6} \times \left(\frac{T_b}{p_b} \right) D^{5/2} \sqrt{\frac{(\hat{\mathbf{p}}_{ij})^2 - (\check{\mathbf{p}}_{ij})^2}{\gamma_g Z T L f}},$$

where T_b is the base temperature (unit R), p_b is the base pressure (unit psia), D is the inside diameter (unit in), γ_g is the gas specific gravity (air = 1), Z is the gas deviation factor at average flowing temperature and average pressure, T is the average flowing temperature (unit R), L is the length of pipe (unit mile), and f is the Moody friction factor.

(a) Linear approximation for Weymouth constraints with $T = T' = 5$

(b) Weymouth constraints

Fig. 1 Weymouth constraint and its linear approximation

To deal with the nonlinearity introduced by (47), a linear approximation framework is proposed in [23], where the Weymouth equation (47) is approximated by a set of linear forms (see Figure 1). For a pipeline connecting nodes i and j , the approximation is performed through the following two steps: Firstly, two finite sets of points $\mathcal{T} := \{\hat{p}_{ij}^t, t = 1, 2, \dots, T\}$ and $\mathcal{T}' := \{\check{p}_{ij}^{t'}, t' = 1, 2, \dots, T'\}$ are selected from the feasible sets of $\hat{\mathbf{p}}_{ij}$ and $\check{\mathbf{p}}_{ij}$ respectively. Without loss of generality, we let $P_{ij}^{\min} \leq \hat{p}_{ij}^1 \leq \hat{p}_{ij}^2 \leq \dots \leq \hat{p}_{ij}^T \leq P_{ij}^{\max}$ and $\bar{P}_{ij}^{\min} \leq \check{p}_{ij}^1 \leq \check{p}_{ij}^2 \leq \dots \leq \check{p}_{ij}^{T'} \leq \bar{P}_{ij}^{\max}$, where \hat{p}_{ij}^t and $\check{p}_{ij}^{t'}$ are selected as 'breakpoints' like in [23]. In the second step, the linear approximations are used to substitute the Weymouth constraints $\mathbf{q}_{ij} \leq q_{ij}^c(\hat{\mathbf{p}}_{ij}, \check{\mathbf{p}}_{ij})$ from (47) for all $i, j \in \mathcal{N}$ as

$$\mathbf{q}_{ij} \leq K_{ij} \left(\frac{\hat{p}_{ij}^t}{\sqrt{(\hat{p}_{ij}^t)^2 - (\check{p}_{ij}^{t'})^2}} \hat{\mathbf{p}}_{ij} - \frac{\check{p}_{ij}^{t'}}{\sqrt{(\hat{p}_{ij}^t)^2 - (\check{p}_{ij}^{t'})^2}} \check{\mathbf{p}}_{ij} \right) \quad (48)$$

for $t \in \mathcal{T}$ and $t' \in \mathcal{T}'$. We refer readers to [23] for more details about the linear approximation of the Weymouth constraints. For ease of notation, we define

$$\hat{P}_{ij}^{tt'} := \frac{\hat{p}_{ij}^t}{\sqrt{(\hat{p}_{ij}^t)^2 - (\check{p}_{ij}^{t'})^2}}, \quad \check{P}_{ij}^{t't} := \frac{\check{p}_{ij}^{t'}}{\sqrt{(\hat{p}_{ij}^t)^2 - (\check{p}_{ij}^{t'})^2}}.$$

Now let us look into the maximization of the overall profit from gas supply in the entire network, subject to production, transmission, and compressor constraints. Because the network is consisted by a number of hub-and-spoke subnetworks (see Figure 2 given below), the problem can be reduced to the

following second-order cone program:

$$\begin{aligned}
\max \quad & \sum_{i \in \mathcal{N}_m} p_i(\mathbf{Q}_i^+) \mathbf{Q}_i^+ - \sum_{j \in \mathcal{N}_g} C_j \mathbf{Q}_j^- - \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{I}(i)} c_{ji} \hat{\mathbf{P}}_{ji} \\
\text{s.t.} \quad & \left\{ \begin{aligned} & \sum_{j \in \mathcal{O}(i)} \mathbf{q}_{ij} \geq G_i, \quad \forall i \in \mathcal{N}_g, \\ & \mathbf{q}_{ij}^2 + K_{ij}^2 \check{\mathbf{p}}_{ij}^2 \leq K_{ij}^2 \hat{P}_{ij}^2, \quad \forall i \in \mathcal{N}_g, j \in \mathcal{O}(i), \\ & \mathbf{q}_{ji}^2 + K_{ji}^2 \check{\mathbf{p}}_{ji}^2 \leq K_{ji}^2 \hat{P}_{ji}^2, \quad \forall i \in \mathcal{N}_m, j \in \mathcal{I}(i), \\ & \mathbf{q}_{ij} \leq K_{ij} \left(\hat{P}_{ij}^{tt'} \hat{\mathbf{p}}_{ij} - \check{P}_{ij}^{tt'} \check{\mathbf{p}}_{ij} \right), \quad \forall t \in \mathcal{T}, t' \in \mathcal{T}', i \in \mathcal{N}_s, j \in \mathcal{O}(i) \cap \mathcal{N}_s, \\ & \mathbf{Q}_i^+ = \sum_{j \in \mathcal{I}(i)} \mathbf{q}_{ji}, \quad \forall i \in \mathcal{N}, \\ & \mathbf{Q}_i^- = \sum_{j \in \mathcal{O}(i)} \mathbf{q}_{ij}, \quad \forall i \in \mathcal{N}, \\ & \mathbf{Q}_i^- - \mathbf{Q}_i^+ \leq -\xi_i, \quad \forall i \in \mathcal{N}_s, \\ & \bar{P}_{ij}^{\min} \leq \check{\mathbf{p}}_{ij} \leq \bar{P}_{ij}^{\max}, \quad \forall j \in \mathcal{N}, i \in \mathcal{I}(j), \\ & P_{ij}^{\min} \leq \hat{\mathbf{p}}_{ij} \leq P_{ij}^{\max}, \quad \forall j \in \mathcal{N}, i \in \mathcal{I}(j) \end{aligned} \right\} \quad \forall \xi_i, i \in \mathcal{N}_s.
\end{aligned}$$

Here, the decision variables include the vectors of inlet pressures, outlet pressures, and flow quantities transmitted in the pipelines and are denoted respectively by

$$\begin{cases} \check{\mathbf{p}} := (\check{\mathbf{p}}_{12}, \dots, \check{\mathbf{p}}_{1N}, \dots, \check{\mathbf{p}}_{ij}, \dots, \check{\mathbf{p}}_{N1}, \dots, \check{\mathbf{p}}_{N(N-1)}), \\ \hat{\mathbf{p}} := (\hat{\mathbf{p}}_{12}, \dots, \hat{\mathbf{p}}_{1N}, \dots, \hat{\mathbf{p}}_{ij}, \dots, \hat{\mathbf{p}}_{N1}, \dots, \hat{\mathbf{p}}_{N(N-1)}), \\ \mathbf{q} := (\mathbf{q}_{12}, \dots, \mathbf{q}_{1N}, \dots, \mathbf{q}_{ij}, \dots, \mathbf{q}_{N1}, \dots, \mathbf{q}_{N(N-1)}), \end{cases}$$

the intermediate variables \mathbf{Q}_i^+ and \mathbf{Q}_i^- are the total quantities transmitted into and out node i respectively for each $i \in \mathcal{N}$. The objective function includes three terms: the total revenue with $p_i(\mathbf{Q}_i^+) \mathbf{Q}_i^+$ to be the revenue obtained from market $i \in \mathcal{N}_m$, the production costs in the fields with \mathbf{Q}_i^- to be the quantity of gas produced in field $i \in \mathcal{N}_g$, and the operation costs spent for generating pressures. It is straightforward to see that the objective function is concave and quadratic.

Now let us go through the constraints. The first set of constraints is on the productions of natural gas at field nodes, which means that the gas quantity transmitted out a field must be no less than the minimum contracted production. The second set of constraints is on the flow quantities transmitted in a pipeline from field nodes. By setting $\hat{P}_{ij} = 0$ for any unconnected pair $i \in \mathcal{N}_g$ and $j \in \mathcal{N}$, we have that \mathbf{q}_{ij} and $\check{\mathbf{p}}_{ij}$ are both zeros. The third set of constraints is on the flow quantities transmitted in a pipeline to market nodes. By setting $\hat{P}_{ji} = 0$ for any unconnected pair $i \in \mathcal{N}_m$ and $j \in \mathcal{N}$, we have that \mathbf{q}_{ji} and $\check{\mathbf{p}}_{ji}$ are both zeros.

Remark 2 Denote by $\mathbf{x} := (\check{\mathbf{p}}, \hat{\mathbf{p}}, \mathbf{q})$. Noting that, for every $i \in \mathcal{N}_g$, there exists a unique $j \in \mathcal{O}(i)$, we can write the second set of constraints as

$$\|A_i \mathbf{x}\| \leq \beta_i^T \mathbf{x}, \quad \forall i \in \mathcal{N}_g.$$

Here, A_i is an $N_g \times 3N$ dimensional matrix with the (i, j_1) -th element being K_{ij} , the (i, j_2) -th element being 1, and the other elements are all zeros, where N_g is the number of field nodes, $j_1 := (N-1)(i-1) + j$ and $j_2 := (2N+i-1)(N-1) + j$ are the positions of $\check{\mathbf{p}}_{ij}$ and \mathbf{q}_{ij} in \mathbf{x} respectively. In addition, β_i is a $3N$ dimensional vector with the j_3 -th element being K_{ij} , where $j_3 := (N+i-1)(N-1) + j$ is the position of $\hat{\mathbf{p}}_{ij}$ in \mathbf{x} . We can write the third set of constraints in a similar form as

$$\|A'_i \mathbf{x}\| \leq \beta'_i{}^T \mathbf{x}, \quad \forall i \in \mathcal{N}_m,$$

where A'_i is an $N_m \times 3N$ dimensional matrix, β'_i is a $3N$ dimensional vector, and N_m is the number of market nodes.

The fourth set of constraints is on the flow in a pipeline connecting two station nodes, where the linear approximations are implemented to reformulate the constraints $\mathbf{q}_{ij} \leq q_{ij}^c(\hat{\mathbf{p}}_{ij}, \check{\mathbf{p}}_{ij})$, $i \neq j$ and $i, j \in \mathcal{N}_s$, into linear constraints, where $q_{ij}^c(\hat{\mathbf{p}}_{ij}, \check{\mathbf{p}}_{ij})$ is given as in (47).

Remark 3 In [23], this type of linear approximations has been used for every pipeline in the transmission network, where the number of constraints resulted from linear approximation is $N(N-1) \times T \times T'$. For the network with hub-and-spoke substructure, the second, third and fourth sets of constraints show that we only need to implement the linear approximations to the pipelines connecting stations and so the number of constraints on pipeline flows can be reduced in our problem.

The fifth and sixth sets of constraints define \mathbf{Q}_i^+ and \mathbf{Q}_i^- as the quantities of flows transmitted into and out node $i \in \mathcal{N}$. The seventh set of constraints is on the random losses at station nodes, where ξ_i is the quantity of the natural gas tapped off at station $i \in \mathcal{N}_s$. The eighth and ninth sets of constraints are from the upper and lower bounds of inlet pressures and outlet pressures of pipelines. Recall that $P_{ij}^{\max} = P_{ij}^{\min} = 0$ and $\bar{P}_{ij}^{\max} = \bar{P}_{ij}^{\min} = 0$ when i and j are not connected. Recall also that the pressures at inlet ends of pipelines connecting from fields are fixed by the contracts and thus we define $\bar{P}_{ij}^{\max} = \bar{P}_{ij}^{\min} = \hat{P}_{ij}$ for all $i \in \mathcal{N}_g$, and the inlet ends of pipelines connecting to the markets from stations are fixed by the supply contracts and thus we define $\bar{P}_{ji}^{\max} = \bar{P}_{ji}^{\min} = \hat{P}_{ji}$ for all $i \in \mathcal{N}_m$.

It is worth mentioning that, in the real world, the natural gases produced at different fields are usually with different percentages of their chemical components, where these percentages are often used to measure the quality of natural gas produced at fields. Accordingly, another set of constraints has to be made on inlet and outlet pressures to control the mixed gas at each station to achieve the contracted quality standard. To concentrate our focus on

applying theoretical framework, we assume that the natural gas qualities at different fields are the same and never changed in the transmission. This assumption can be relaxed by introducing a set of linear equality constraints for mass balance on each component; see, e.g., [23].

Before proceeding to the numerical tests, we summarize the notation on parameters used in the model in the following table.

Table 1 Notation

G_i	minimum production at field $i \in \mathcal{N}_g$
K_{ij}	Weymouth constant for pipeline from node i to $j \in \mathcal{O}(i)$
$\hat{P}_{ij}^{tt'}$, $\check{P}_{ij}^{tt'}$	coefficients in the linear approximation of Weymouth equations
P_{ij}^{\min} , P_{ij}^{\max}	lower and upper bounds on inlet end of pipeline from i to $j \in \mathcal{O}(i)$
\bar{P}_{ij}^{\min} , \bar{P}_{ij}^{\max}	lower and upper bounds on outlet end of pipeline from i to $j \in \mathcal{O}(i)$
$a_i - b_i Q_i^+$	unit price for gas in market $i \in \mathcal{N}_m$
C_i	unit cost for gas production in field $i \in \mathcal{N}_g$
c_{ij}	unit cost for generating pressure at an end of pipeline from $i \in \mathcal{N}$ to $j \in \mathcal{O}(i)$

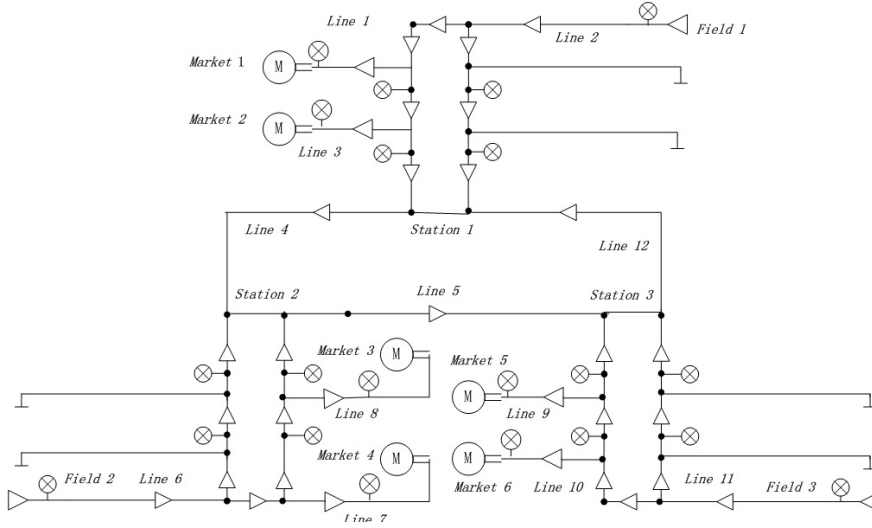


Fig. 2 Structure of the natural gas network

Let us consider an example on a natural gas transmission network with three compressor stations. The network is described in Figure 2. In this example, the set of market nodes $\mathcal{N}_m = \{1, 2, \dots, 6\}$, the set of field nodes $\mathcal{N}_g = \{1', 2', 3'\}$, and the set of station nodes $\mathcal{N}_s = \{7, 8, 9\}$. Moreover, markets 1 and 2 (nodes $\{1, 2\}$) and field 1 (node $\{1'\}$) are connected to station 1 (node 7), markets 3 and 4 (nodes $\{3, 4\}$) and field 2 (node $\{2'\}$) are connected

to station 2 (node 8), and markets 5 and 6 (nodes $\{5, 6\}$) and field 3 (node $\{3'\}$) are connected to station 3 (node 9).

Table 2 Network of Figure 2: Weymouth equation parameters, production costs, price co-efficiencies

Network Data														
Pipeline										Market Data				
From node	To node	p_b (psia)	T_b (R)	D (in)	γ_g (air = 1)	Z (-)	Len (mile)	f (-)	T (R)	Node	a_m (kCNY/mmcf)	b_m (kCNY/mmcf ²)	$p_{m'm}^{\min}$ (psia)	$p_{m'm}^{\max}$ (psia)
1'	7	15	560	20	0.87	0.93	33	0.01	560	1	29	0.04	15	15
7	1	15	560	30	0.87	0.93	11	0.01	560	2	26	0.04	15	15
7	2	15	560	20	0.87	0.93	12	0.01	560	3	22	0.03	15	15
7	8	15	560	25	0.87	0.93	75	0.01	560	4	21	0.03	15	15
2'	8	15	560	30	0.85	0.93	28	0.01	546	5	30	0.02	15	15
8	3	15	560	25	0.85	0.93	12	0.01	546	6	25	0.02	15	15
8	4	15	560	20	0.85	0.93	12	0.01	546	Field Data				
8	9	15	560	20	0.85	0.93	80	0.01	546					
3'	9	15	560	20	0.90	0.93	23	0.01	564	Node	C_g (mmcdf)	C_g (kCNY/mmcf)	$p_{gg'}^{\max}$ (psia)	$p_{gg'}^{\min}$ (psia)
9	5	15	560	20	0.90	0.93	13	0.01	564	1'	450	19	50	0
9	6	15	560	25	0.90	0.93	12	0.01	564	2'	800	15	50	0
9	7	15	560	30	0.90	0.93	72	0.01	564	3'	275	18	50	0

The values of the parameters in the model are given in Table 2. In the model, we fix the outlet pressures of pipelines to each compressor station at 15 (unit: psia), and cost c_{ij} for generating one unit inlet pressure to be 87 (unit: k×CNY/psia) for all pipelines. In addition, for uncertainties in the model, we set that the losses ξ_i at stations (i.e., nodes $i = 7, 8, 9$) follow normal distribution with means being 15, 20, 10 and variances being 5.5, 9.0, 3.5, respectively.

In our numerical tests, we used the linear approximation (48) with $T = T' = 10$ for the Weymouth equation (47). Note that the linear approximation does not change the objective function and other constraints. We solved our optimization problems and the problems in [23] with the same samples of ξ . In Table 3, we denote our optimization problem and the problem in [23] by SOC(n) and LC(n) respectively, where n is the sample size in each test. In our tests, we varied n from 100, 300 to 1000 and compared the results of SOC(n) and LC(n).

We implemented the ERM scheme to solve this model and analyzed the impact of loss uncertainties at stations to the overall profit and optimal operations in the transmission network. The results given in Table 3 were obtained by using the function ϕ_{NR} . The problems were solved in the GAMS platform by NLP solver. Table 3 lists the sample averaged solutions. Here, to avoid redundancy, we only report the main results in approximation solutions for inlet pressures in pipelines, gas quantities transmitted by pipelines, and the optimal values of objective function under different sample sizes.

Notice that, in our method, the linear approximation is only incorporated for the constraints of gas quantities in pipelines connecting stations rather than all pipelines in 'LC(n)'. Therefore, we can take the results solved from the 'LC(n)' model as approximation to the results in 'SOC(n)'. From another

Table 3 Numerical results for optimal flows, optimal injection pressures, and optimal values

Optimal Solutions										
sample size	Flows to market m						sample size	Inlet Pressures at field g		
	q_{71}	q_{72}	q_{83}	q_{84}	q_{95}	q_{96}		$q_{1'7}$	$q_{2'8}$	$q_{3'9}$
(mmcf)						(mmcf)				
SOC(100)	297.13	242.36	200.00	131.94	764.05	368.92	SOC(100)	589.28	1136.50	378.83
LC(100)	283.92	244.73	196.78	136.54	742.03	339.83	LC(100)	559.71	1207.92	359.06
SOC(300)	303.59	254.21	200.00	137.75	753.24	361.26	SOC(300)	608.49	1045.08	402.18
LC(300)	312.71	262.43	194.41	139.05	724.55	312.90	LC(300)	589.28	1112.05	412.03
SOC(1000)	305.02	250.39	200.00	142.54	748.76	352.07	SOC(1000)	615.08	1008.24	437.90
LC(1000)	317.85	265.92	195.57	140.31	735.61	331.45	LC(1000)	601.44	1052.43	441.12
sample size	Inlet pressures to market m						sample size	Inlet pressures at field g		
	\bar{p}_{71}	\bar{p}_{72}	\bar{p}_{83}	\bar{p}_{84}	\bar{p}_{95}	\bar{p}_{96}		$\bar{p}_{1'7}$	$\bar{p}_{2'8}$	$\bar{p}_{3'9}$
(psia)						(psia)				
SOC(100)	26.87	39.71	24.73	21.91	43.33	31.01	100	19.56	43.64	26.98
LC(100)	25.17	40.25	23.16	22.71	40.00	29.02	LC(100)	17.02	45.92	26.01
SOC(300)	28.10	40.54	24.73	22.16	43.06	30.76	300	19.69	42.06	27.55
LC(300)	29.74	42.98	23.01	23.22	39.55	28.32	LC(300)	18.03	43.53	28.31
SOC(1000)	28.15	40.27	24.73	22.37	42.94	30.56	1000	19.74	41.40	28.40
LC(1000)	30.29	43.12	23.11	23.90	39.11	27.88	LC(1000)	18.81	42.01	29.82
sample size			100	300	1000	sample size	100	300	1000	
Optimal Value (mCNY)			1.315	1.367	1.392	Time (sec)	129	217	289	
Optimal Value in LC (mCNY)			1.233	1.172	1.216	Time in LC (sec)	107	194	296	

point of view, both models can be taken as approximations to the true natural gas transmission problem, where ‘SOC(n)’ is closer with less number of linear approximation of the Weymouth constraints. For computational time, the results show that solving the model with second-order cone constraint is almost at the same level as ‘LC(n)’ with a little longer computational time.

4.2 Stochastic SOCP optimal power flow

In this subsection, we consider a stochastic optimal power flow (OPF) model in radial network connected to the wind farms and recast it as a stochastic SOCP problem. The proposed stochastic SOCP-OPF model can be adopted as a tool for different stochastic analysis of power systems. We implemented the preceding ERM scheme to solve the SOCP-OPF model and observed the impact of power injection uncertainty on the total generation cost in the power systems.

A radial network is composed of buses and lines connecting these buses and has a tree topology. The root of the tree is a substation bus that is connected to the transmission network. It has a fixed voltage and redistributes the bulk power receiving from the transmission network to other buses. We index the substation bus by 0 and the other buses by $1, \dots, N$. Denote by $\mathcal{N} := \{0, 1, \dots, N\}$ the collection of all buses and let $\mathcal{N}^+ := \mathcal{N} \setminus \{0\}$. Each line connects an ordered pair (i, j) of buses where bus j lies on the unique path from bus i to bus 0. Let \mathcal{E} denote the collection of all lines and abbreviate $(i, j) \in \mathcal{E}$ by $i \rightarrow j$ for convenience.

Uncertainty often occurs when renewable power resources such as wind are incorporated. We assume that a subset \mathcal{W} of buses holds uncertain power sources (wind farms) and, for each bus i which holds uncertain power sources, the stochastic amount of power generated by source i is of the form

$\xi_i := \xi_i^p + \mathbf{i}\xi_i^q$, where ξ_i is an independent random variable with known mean and deviation. In particular, we investigate both the Gaussian and Weibull distributions of ξ_i . In our model, we further assume the independence of wind power fluctuations at different sites, which is justified by the fact that the wind farms are sufficiently far away from each other. For the typical OPF time span 15 minutes and typical wind speed of $10m/s$, fluctuations of wind at the farms more than 10km apart are not correlated.

For each bus $i \in \mathcal{N}$, let \mathbf{V}_i denote its complex voltage and define $\mathbf{v}_i := |\mathbf{V}_i|^2$. Specifically, the substation voltage v_0 is given and fixed. Let $\mathbf{s}_i := \mathbf{p}_i + \mathbf{i}\mathbf{q}_i$ denote the power injection of bus i , where p_i and q_i denote the real and reactive power injections respectively. For each line $(i, j) \in \mathcal{E}$, let $z_{ij} := r_{ij} + \mathbf{i}x_{ij}$ denote its impedance. Denote by \mathbf{I}_{ij} the complex current from bus i to bus j and let $\mathbf{l}_{ij} := |\mathbf{I}_{ij}|^2$. Denote further by $\mathbf{S}_{ij} := \mathbf{P}_{ij} + \mathbf{i}\mathbf{Q}_{ij}$ the sending-end power flow from bus i to bus j , where \mathbf{P}_{ij} and \mathbf{Q}_{ij} stand for the real and reactive power flow respectively. As customary, we assume that the complex voltage V_0 on the substation bus is given. We summarize the notation in Table 4 and Figure 3.

Table 4 Notation

$\mathbf{V}_i, \mathbf{v}_i$	complex voltages on bus i with $\mathbf{v}_i = \mathbf{V}_i ^2$
$\mathbf{s}_i = \mathbf{p}_i + \mathbf{i}\mathbf{q}_i$	complex net injection on bus i
$\mathbf{I}_{ij}, \mathbf{l}_{ij}$	complex currents from buses i to j with $\mathbf{l}_{ij} = \mathbf{I}_{ij} ^2$
$\mathbf{S}_{ij} = \mathbf{P}_{ij} + \mathbf{i}\mathbf{Q}_{ij}$	complex power flowing out from buses i to bus j
$z_{ij} = r_{ij} + \mathbf{i}x_{ij}$	impedance on line (i, j)

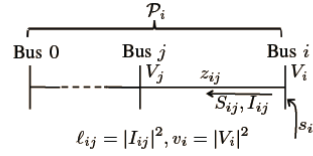


Fig. 3 Illustration of the notation in network

The OPF problem seeks to minimize the generation cost, subject to power flow constraint, power injection constraint, and voltage constraint. Following [13,16], in case of a radial network, the stochastic OPF using branch flow model is reduced to the following convex optimization problem with exact relaxation or, more precisely, a second-order cone program, interpreted as operations in

real variables:

$$\begin{aligned}
& \min \sum_{i \in \mathcal{N}} f_i(\mathbf{p}_i^g) \\
& \text{s.t. } \left\{ \begin{aligned}
& \mathbf{P}_{ij} = (\mathbf{p}_i^g + \boldsymbol{\xi}_i^p) - \mathbf{p}_i^c + \sum_{h:h \rightarrow i} (\mathbf{P}_{hi} - r_{hi} \mathbf{l}_{hi}), \quad \forall (i, j) \in \mathcal{E}, \\
& \mathbf{Q}_{ij} = (\mathbf{q}_i^g + \boldsymbol{\xi}_i^q) - \mathbf{q}_i^c + \sum_{h:h \rightarrow i} (\mathbf{Q}_{hi} - x_{hi} \mathbf{l}_{hi}), \quad \forall (i, j) \in \mathcal{E}, \\
& \mathbf{p}_0^g - \mathbf{p}_0^c + \sum_{h:h \rightarrow 0} (\mathbf{P}_{h0} - r_{h0} \mathbf{l}_{h0}) = 0, \\
& \mathbf{q}_0^g - \mathbf{q}_0^c + \sum_{h:h \rightarrow 0} (\mathbf{Q}_{h0} - x_{h0} \mathbf{l}_{h0}) = 0, \\
& \mathbf{v}_i - \mathbf{v}_j = 2(r_{ij} \mathbf{P}_{ij} + x_{ij} \mathbf{Q}_{ij}) - \mathbf{l}_{ij}(r_{ij}^2 + x_{ij}^2), \quad \forall (i, j) \in \mathcal{E}, \\
& \frac{\mathbf{P}_{ij}^2 + \mathbf{Q}_{ij}^2}{\mathbf{v}_i} \leq \mathbf{l}_{ij}, \quad \forall (i, j) \in \mathcal{E}, \\
& \mathbf{s}_i \in S_i, \quad i \in \mathcal{N}^+, \\
& \underline{v}_i \leq \mathbf{v}_i \leq \bar{v}_i, \quad i \in \mathcal{N}^+ \} \quad \forall \boldsymbol{\xi}_i, \quad i \in \mathcal{N}^+,
\end{aligned} \right.
\end{aligned}$$

where $\boldsymbol{\xi}_i := \boldsymbol{\xi}_i^p + \mathbf{i}\boldsymbol{\xi}_i^q$, $\mathbf{p}_i := \mathbf{p}_i^c - \mathbf{p}_i^g$ and $\mathbf{q}_i := \mathbf{q}_i^c - \mathbf{q}_i^g$ are the real and reactive net loads at node i . In particular, \mathbf{p}_i^c and \mathbf{q}_i^c are the real and reactive power consumption at node i , \mathbf{p}_i^g and \mathbf{q}_i^g are the real and reactive conventional power generation at node i . We use $h : h \rightarrow i$ to denote a collection of buses in \mathcal{N} prior to i in the tree network topology with $(h, i) \in \mathcal{E}$ and use $h : h \rightarrow 0$ to denote the collection of buses in \mathcal{N} prior to the substation root with $(h, 0) \in \mathcal{E}$. In addition, we assume that each f_i is convex quadratic and, particularly, $f_i(\mathbf{p}_i^g) := c_{i2}(\mathbf{p}_i^g)^2 + c_{i1}\mathbf{p}_i^g + c_{i0}$ in our model.

Notice that the convex relaxed power flow equation $\mathbf{l}_{ij} \geq \frac{\mathbf{P}_{ij}^2 + \mathbf{Q}_{ij}^2}{\mathbf{v}_i}$, $(i, j) \in \mathcal{E}$, is exactly in the form of a rotated second-order cone in \Re^4 , which is a convex set given as

$$\mathcal{K}_r^2 := \{(x_1, x_2, x_3) \in \Re \times \Re \times \Re^2 \mid x_1 x_2 \geq x_3^T x_3, x_1 \geq 0, x_2 \geq 0\}.$$

Trivially, the rotated second-order cone in \Re^4 can be expressed as a linear transformation (actually, a rotation) of the (plain) second-order cone in \Re^4 , due to

$$x_1 x_2 \geq x_3^T x_3, x_1 \geq 0, x_2 \geq 0 \iff \left\| \begin{bmatrix} x_1 - x_2 \\ 2x_3 \end{bmatrix} \right\| \leq x_1 + x_2.$$

Recall that $(x_1, x_2, x_3) \in \mathcal{K}_r^2$ if and only if $(x_1 + x_2, x_4) \in \mathcal{K}^4$, where $x_4 := (x_1 - x_2, 2x_3)$. For a clear presentation in the subsequent discussion, we recast the above distribution network stochastic dispatch problem in the following

compact SOCP form:

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s. t.} \quad & \left\{ \begin{array}{l} \mathbf{x} \in \mathcal{X}, \\ \mathbf{A}\mathbf{x} + B\xi = 0, \\ \|G_i\mathbf{x}\| \leq g_i^T \mathbf{x}, \quad \forall i \text{ with } (i, j) \in \mathcal{E} \end{array} \right\} \quad \forall \xi, \end{aligned} \quad (49)$$

where the vector \mathbf{x} represents all dispatching variables related to optimal power flow of the distribution network, $f(\mathbf{x})$ represents the total generation cost, and \mathcal{X} is the feasible set of radial distribution network.

Notice that (49) actually coincides with the linear form of (13), which is not computationally trackable by any commonly-used commercial SOCP solvers such as MOSEK or Gurobi. Our key step toward numerical solution of problem (49) is to reformulate it into the SSOCCP (10) by its KKT system, as discussed previously. Hence, as in convex programming, the KKT conditions are necessary and sufficient for global optimality [20]. As each KKT point is a global minimizer of (49), we may consider to solve (49) by searching its KKT points, which is exactly what we have discussed in the preceding sections for SSOCCP.

In order to evaluate the proposed SSOCP-OPF, a slightly modified version of a real-world 47-bus network in the service territory of Southern California Edison (SCE) with two wind farms connected to bus 5 and bus 20 are considered. The SCE network is shown in Figure 4 with parameters given in Table 5. We suppose that the system regulator is to optimize the total generation cost in an hourly basis in the presence of variable wind power generations by assuming that the forecasted distribution for wind speed is available for the next hour interval. The dispatch results from the day ahead market, which are obtained by using ϕ_{FB} and ϕ_{NR} , based on the wind power assuming Gaussianity and Weibull distribution injected to bus 5 and bus 20 in GAMS platform using default NLP solvers, are given in Tables 6 and 7 respectively. Tables 6 and 7 also depict the expected residual and the total generation cost, as the sample size increases.

Table 5 Network of Figure 1: Line impedances, peak spot load KVA, capacitors and PV generation's nameplate ratings

Network Data																	
Line Data				Line Data				Line Data				Load Data		Load Data		PV Generators	
From Bus.	To Bus.	R (Ω)	X (Ω)	From Bus.	To Bus.	R (Ω)	X (Ω)	From Bus.	To Bus.	R (Ω)	X (Ω)	Bus No.	Peak MVA	Bus No.	Peak MVA	Bus No.	Nameplate Capacity
1	2	0.250	0.808	8	41	0.107	0.031	21	22	0.198	0.046	1	10	34	0.2		
2	13	0	0	8	35	0.076	0.015	22	23	0	0	11	0.67	36	0.27	13	1.5MW
2	3	0.031	0.092	8	9	0.031	0.031	27	31	0.046	0.015	12	0.45	38	0.45	17	0.4MW
3	4	0.46	0.092	9	10	0.015	0.015	27	28	0.107	0.031	14	0.89	39	1.34	19	1.5MW
3	14	0.092	0.031	9	42	0.153	0.046	28	29	0.107	0.031	16	0.07	40	0.13	23	1MW
3	15	0.214	0.046	10	11	0.107	0.076	29	30	0.061	0.015	18	0.67	41	0.67	24	2MW
4	20	0.336	0.061	10	46	0.229	0.122	32	33	0.046	0.015	21	0.45	42	0.13		
4	5	0.107	0.183	11	47	0.031	0.015	33	34	0.031	0	22	2.23	44	0.45		
5	26	0.061	0.015	11	12	0.076	0.046	35	36	0.076	0.015	25	0.45	45	0.2		
5	6	0.015	0.031	15	18	0.046	0.015	35	37	0.076	0.046	26	0.2	46	0.45		
6	27	0.168	0.061	15	16	0.107	0.015	35	38	0.107	0.015	28	0.13				
6	7	0.031	0.046	16	17	0	0	42	43	0.061	0.015	29	0.13				
7	32	0.076	0.015	18	19	0	0	43	44	0.061	0.015	30	0.2				
7	8	0.015	0.015	20	21	0.122	0.092	43	45	0.061	0.015	31	0.07				
8	40	0.046	0.015	20	25	0.214	0.046					32	0.13				
8	39	0.224	0.046	21	24	0	0					33	0.27				

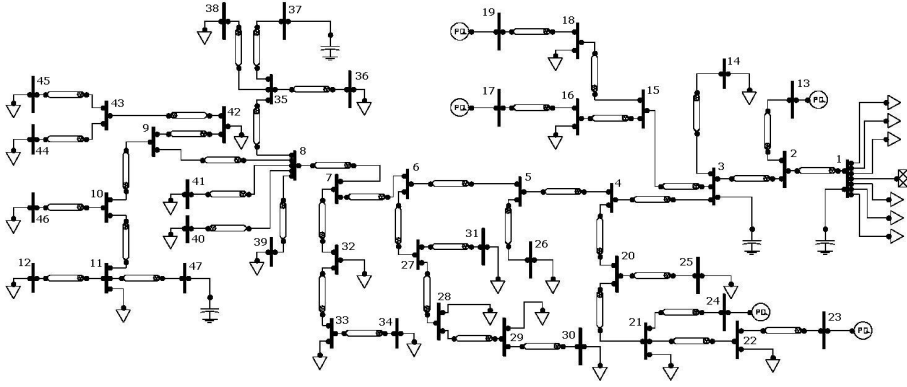


Fig. 4 Schematic diagram of SCE 47-bus distribution system

Table 6 OPF results for SCE 47-bus obtaining from SSOCP-OPF using ϕ_{FB}

Power Uncertainty	Gaussian Wind Power Uncertainty				Weibull Wind Power Uncertainty			
	100	500	1000	2000	100	500	1000	2000
Sample Size	100	500	1000	2000	100	500	1000	2000
Expected Residual	2.849	3.433	3.397	3.407	8.625	8.840	8.504	8.660
Generation Cost (MW)	24.272	24.255	25.104	25.052	41.879	42.466	42.594	42.267
Bus13 (MW)	$p_{13}^w = 0.242$ (MV Ar)	$p_{13}^w = 0.194$ $q_{13}^w = 1.500$	$p_{13}^w = 0.244$ $q_{13}^w = 1.500$	$p_{13}^w = 0.248$ $q_{13}^w = 1.500$	$p_{13}^w = 0.437$ $q_{13}^w = 1.500$	$p_{13}^w = 0.444$ $q_{13}^w = 1.500$	$p_{13}^w = 0.437$ $q_{13}^w = 1.500$	$p_{13}^w = 0.437$ $q_{13}^w = 1.500$
Bus17 (MW)	$p_{17}^w = 0.180$ (MV Ar)	$p_{17}^w = 0.194$ $q_{17}^w = 0.400$	$p_{17}^w = 0.244$ $q_{17}^w = 0.400$	$p_{17}^w = 0.237$ $q_{17}^w = 0.400$	$p_{17}^w = 0.400$ $q_{17}^w = 0.400$	$p_{17}^w = 0.400$ $q_{17}^w = 0.400$	$p_{17}^w = 0.400$ $q_{17}^w = 0.400$	$p_{17}^w = 0.400$ $q_{17}^w = 0.400$
Bus19 (MW)	$p_{19}^w = 0.185$ (MV Ar)	$p_{19}^w = 0.200$ $q_{19}^w = 1.500$	$p_{19}^w = 0.244$ $q_{19}^w = 1.500$	$p_{19}^w = 0.233$ $q_{19}^w = 1.500$	$p_{19}^w = 0.459$ $q_{19}^w = 1.500$	$p_{19}^w = 0.466$ $q_{19}^w = 1.500$	$p_{19}^w = 0.437$ $q_{19}^w = 1.500$	$p_{19}^w = 0.443$ $q_{19}^w = 1.500$
Bus23 (MW)	$p_{23}^w = 0.189$ (MV Ar)	$p_{23}^w = 0.205$ $q_{23}^w = 1.000$	$p_{23}^w = 0.239$ $q_{23}^w = 1.000$	$p_{23}^w = 0.230$ $q_{23}^w = 1.000$	$p_{23}^w = 0.474$ $q_{23}^w = 1.000$	$p_{23}^w = 0.482$ $q_{23}^w = 1.000$	$p_{23}^w = 0.419$ $q_{23}^w = 1.000$	$p_{23}^w = 0.428$ $q_{23}^w = 1.000$
Bus24 (MW)	$p_{24}^w = 0.187$ (MV Ar)	$p_{24}^w = 0.202$ $q_{24}^w = 2.000$	$p_{24}^w = 0.241$ $q_{24}^w = 2.000$	$p_{24}^w = 0.228$ $q_{24}^w = 2.000$	$p_{24}^w = 0.466$ $q_{24}^w = 2.000$	$p_{24}^w = 0.474$ $q_{24}^w = 2.000$	$p_{24}^w = 0.428$ $q_{24}^w = 2.000$	$p_{24}^w = 0.440$ $q_{24}^w = 2.000$
Times(s)	168.753	280.493	316.686	431.546	192.221	321.740	240.382	454.491

Table 7 OPF results for SCE 47-bus obtaining from SSOCP-OPF using ϕ_{NR}^μ with $\mu = 10^{-4}$

Power Uncertainty	Gaussian				Weibull			
	100	500	1000	2000	100	500	1000	2000
Sample Size	100	500	1000	2000	100	500	1000	2000
Expected Residual	1.310	1.514	1.500	1.551	3.801	3.841	3.692	3.487
Generation Cost (MW)	11.534	11.504	11.807	11.667	18.742	18.906	18.923	18.086
Bus13 (MW)	$p_{13}^w = 0.115$ (MV Ar)	$p_{13}^w = 0.115$ $q_{13}^w = 1.500$	$p_{13}^w = 0.118$ $q_{13}^w = 1.500$	$p_{13}^w = 0.117$ $q_{13}^w = 1.500$	$p_{13}^w = 0.187$ $q_{13}^w = 1.500$	$p_{13}^w = 0.189$ $q_{13}^w = 1.500$	$p_{13}^w = 0.189$ $q_{13}^w = 1.500$	$p_{13}^w = 0.181$ $q_{13}^w = 1.500$
Bus17 (MW)	$p_{17}^w = 0.115$ (MV Ar)	$p_{17}^w = 0.115$ $q_{17}^w = 0.400$	$p_{17}^w = 0.118$ $q_{17}^w = 0.400$	$p_{17}^w = 0.117$ $q_{17}^w = 0.400$	$p_{17}^w = 0.187$ $q_{17}^w = 0.400$	$p_{17}^w = 0.189$ $q_{17}^w = 0.400$	$p_{17}^w = 0.189$ $q_{17}^w = 0.400$	$p_{17}^w = 0.181$ $q_{17}^w = 0.400$
Bus19 (MW)	$p_{19}^w = 0.115$ (MV Ar)	$p_{19}^w = 0.115$ $q_{19}^w = 1.500$	$p_{19}^w = 0.118$ $q_{19}^w = 1.500$	$p_{19}^w = 0.117$ $q_{19}^w = 1.500$	$p_{19}^w = 0.187$ $q_{19}^w = 1.500$	$p_{19}^w = 0.189$ $q_{19}^w = 1.500$	$p_{19}^w = 0.189$ $q_{19}^w = 1.500$	$p_{19}^w = 0.181$ $q_{19}^w = 1.500$
Bus23 (MW)	$p_{23}^w = 0.115$ (MV Ar)	$p_{23}^w = 0.115$ $q_{23}^w = 1.000$	$p_{23}^w = 0.118$ $q_{23}^w = 1.000$	$p_{23}^w = 0.117$ $q_{23}^w = 1.000$	$p_{23}^w = 0.187$ $q_{23}^w = 1.000$	$p_{23}^w = 0.189$ $q_{23}^w = 1.000$	$p_{23}^w = 0.189$ $q_{23}^w = 1.000$	$p_{23}^w = 0.181$ $q_{23}^w = 1.000$
Bus24 (MW)	$p_{24}^w = 0.115$ (MV Ar)	$p_{24}^w = 0.115$ $q_{24}^w = 2.000$	$p_{24}^w = 0.118$ $q_{24}^w = 2.000$	$p_{24}^w = 0.117$ $q_{24}^w = 2.000$	$p_{24}^w = 0.187$ $q_{24}^w = 2.000$	$p_{24}^w = 0.189$ $q_{24}^w = 2.000$	$p_{24}^w = 0.189$ $q_{24}^w = 2.000$	$p_{24}^w = 0.181$ $q_{24}^w = 2.000$
Times (s)	10.278	18.773	20.925	18.909	9.944	10.064	34.614	34.946

From the numerical results, we observed that, as the sample size increases, the convergence in terms of both ϕ_{FB} and ϕ_{NR} are stable, e.g., there are clearly convergent trends toward the expected residual, as well as the dispatching

variables. However, as one may observe immediately from the above tables, even with a fairly small smoothing parameter $\mu = 10^{-4}$, ϕ_{NR} results in smaller and better expected residuals in some sense. Another observation of interest is that, from the time consumptions, the ERM scheme using ϕ_{NR} performs computationally much powerful than the case of ϕ_{FB} . Nevertheless, this phenomenon comes as no surprise to us because, as mentioned at the beginning of Section 2, compared with ϕ_{FB} , ϕ_{NR} usually preserves more information of the complementary structure.

5 Conclusions

We have proposed the ERM formulation (14) for the SSOCCP (10). Some properties related to existence of solutions have been given and an approximation method based on the Monte Carlo techniques and some smoothing techniques have been presented. Further applications including the natural gas production and transportation problems and the stochastic optimal power flow problem in radial network have also been discussed.

Note that we may use the constrained optimization problem

$$\begin{aligned} \min \quad & \mathbb{E}_{\xi} [\|F(x, y, z, \xi)\|^2] + \|\Phi(x, y)\|^2 \\ \text{s.t.} \quad & x \in \mathcal{K}, y \in \mathcal{K} \end{aligned}$$

as a new ERM formulation, instead of the unconstrained formulation (14). It is possible to extend the convergence results shown in Section 2 to this case because the feasible region of the above problem is convex and the Slater's constraint qualification holds for it obviously. One advantage of this constrained formulation is to guarantee better feasibility of approximation solutions than the unconstrained formulation. Alternatively, as is known to us, the constrained optimization problem is generally more difficult to solve than the unconstrained optimization problem.

Acknowledgements The authors are grateful to two anonymous referees and Prof. Xiaojun Chen for their helpful comments and suggestions, which have led to much improvement of the paper.

References

1. T.A. Adeosun, O.A. Olatunde, J.O. Aderohunmu and T.O. Ogunjare, *Development of unsteady-state Weymouth equations for gas volumetric flow rate in horizontal and inclined pipes*, Journal of Natural Gas Science and Engineering, 1 (2009), 113–117.
2. F. Alizadeh and D. Goldfarb, *Second-order cone programming*, Mathematical Programming, 95 (2003), 3–51.
3. F. Babonneau, Y. Nesterov and J.P. Vial, *Design and operations of gas transmission networks*, Operations Research, 60 (2012), 34–47.
4. J.S. Chen, X. Chen and P. Tseng, *Analysis of nonsmooth vector-valued functions associated with second-order cones*, Mathematical Programming, 101 (2004), 95–117.

5. J.S. Chen and S. Pan, *A survey on SOC complementarity functions and solution methods for SOCPs and SOCCPs*, Pacific Journal of Optimization, 8 (2012), 33–74.
6. J.S. Chen and P. Tseng, *An unconstrained smooth minimization reformulation of the second-order cone complementarity problem*, Mathematical Programming, 104 (2005), 297–327.
7. X. Chen and M. Fukushima, *Expected residual minimization method for stochastic linear complementarity problems*, Mathematics of Operations Research, 30 (2005), 1022–1038.
8. X. Chen, C. Zhang and M. Fukushima, *Robust solution of monotone stochastic linear complementarity problems*, Mathematical Programming, 117 (2009), 51–80.
9. X.D. Chen, D. Sun and J. Sun, *Complementarity functions and numerical experiments on some smoothing Newton methods for second-order-cone complementarity problems*, Computational Optimization and Applications, 25 (2003), 39–56.
10. F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
11. D. Dentcheva and A. Ruszczyński, *Optimization with stochastic dominance constraints*, SIAM Journal on Optimization, 14 (2003), 548–566.
12. U. Faraut and A. Koranyi, *Analysis on Symmetric Cones*, Oxford Mathematical Monographs, Oxford University Press, 1994.
13. M. Farivar and S.H. Low, *Branch flow model: Relaxations and convexification*, IEEE 51st Annual Conference on Decision and Control, 2012, 3672–3679.
14. M. Frank and P. Wolfe, *An algorithm for quadratic programming*, Naval Research Logistics, 3 (1956), 95–110.
15. M. Fukushima, Z.Q. Luo and P. Tseng, *Smoothing functions for second-order-cone complementarity problems*, SIAM Journal on Optimization, 12 (2001), 436–460.
16. L. Gan, N. Li, U. Topcu and S.H. Low, *Exact convex relaxation of optimal power flow in radial networks*, IEEE Transactions on Automatic Control, 60 (2015), 72–87.
17. S. Hayashi, N. Yamashita and M. Fukushima, *A combined smoothing and regularization method for monotone second-order cone complementarity problems*, SIAM Journal on Optimization, 15 (2005), 593–615.
18. G.H. Lin and M. Fukushima, *Stochastic equilibrium problems and stochastic mathematical programs with equilibrium constraints: A survey*, Pacific Journal of Optimization, 6 (2010), 455–482.
19. M.S. Lobo, L. Vandenberghe, S. Boyd and H. Lebret, *Applications of second order cone programming*, Linear Algebra and Its Applications, 284 (1998), 193–228.
20. M. Lopez and G. Still, *Semi-infinite programming*, European Journal of Operational Research, 180 (2007), 491–518.
21. B. Patrick, *Probability and Measure*, A Wiley-Interscience Publication, John Wiley & Sons, New York, USA, 1995.
22. R.Z. Rios-Mercado, S. Wu, L.R. Scott and E.A. Boyd, *A reduction technique for natural gas transmission network optimization problems*, Annals of Operations Research, 117 (2002), 217–234.
23. F. Rømo, A. Tomasgard, L. Hellemo, M. Fodstad, B.H. Eidesen and B. Pedersen, *Optimizing Norwegian natural gas production and transport*, Interface, 39 (2009), 46–56.
24. A. Ruszczyński and A. Shapiro, *Stochastic Programming*, Handbooks in Operations Research and Management Science, Elsevier, 2003.
25. J.M. Salazar, U. Diwekar, E. Constantinescu and V.M. Zavala, *Stochastic optimization approach to water management in cooling-constrained power plants*, Applied Energy, 112 (2013), 12–22.
26. A. Shapiro and H. Xu, *Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation*, Optimization, 57 (2008), 395–418.
27. M. Thompson, M. Davison and H. Rasmussen, *Natural gas storage valuation and optimization: A real options application*, Naval Research Logistics, 56 (2009), 226–238.
28. C.L. Tang and X.P. Wu, *Periodic solutions for second order systems with not uniformly coercive potential*, Journal of Mathematical Analysis and Applications, 259 (2001), 386–397.
29. P.J. Wong and R.E. Larson, *Optimization of natural gas pipeline systems via dynamic programming*, IEEE Transactions on Automatic Control, 13 (1968), 475–481.
30. H. Xu, *An implicit programming approach for a class of stochastic mathematical programs with complementarity constraints*, SIAM Journal on Optimization, 16 (2006), 670–696.