Hindawi
Mathematical Problems in Engineering
Volume 2017, Article ID 1564642, 15 pages
https://doi.org/10.1155/2017/1564642

Research Article

A Quasi-Monte-Carlo-Based Feasible Sequential System of Linear Equations Method for Stochastic Programs with Recourse

Changyin Zhou, Rui Su, and Zhihui Jiang

College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao 266590, China

Correspondence should be addressed to Changyin Zhou; zhoucy123@163.com

Received 5 April 2017; Revised 14 July 2017; Accepted 24 July 2017; Published 24 August 2017

Academic Editor: Huanqing Wang

A two-stage stochastic quadratic programming problem with inequality constraints is considered. By quasi-Monte-Carlo-based approximations of the objective function and its first derivative, a feasible sequential system of linear equations method is proposed. A new technique to update the active constraint set is suggested. We show that the sequence generated by the proposed algorithm converges globally to a Karush-Kuhn-Tucker (KKT) point of the problem. In particular, the convergence rate is locally superlinear under some additional conditions.

1. Introduction

Stochastic programming is a framework for modeling optimization problems that involve uncertainty. It has applications in a broad range of areas ranging between finance, transportation, and energy optimization [1, 2]. In the field of industrial production, stochastic programming is also widely used in stochastic control [3–7].

We consider the following two-stage stochastic quadratic programming problem:

$$\min \quad f(x) = P(x) + Q(x), \tag{1a}$$

$$\text{subject to} \quad c(x) \le 0, \tag{1b}$$

where

$$Q(x) = \int_{\Omega} Q(x, \omega)\, p(\omega)\, d\omega, \tag{1c}$$

$$Q(x,\omega) = \max\left\{-\frac{1}{2}y^T G y + y^T(h(\omega) - Tx) \mid Wy \le q,\ y \in R^s\right\}. \tag{1d}$$

$P(\cdot): R^n \to R$ and $c(\cdot): R^n \to R^m$ are twice continuously differentiable. $G \in R^{s\times s}$ is symmetric positive definite. $T \in R^{s\times n}$, $q \in R^t$, and $W \in R^{t\times s}$ are fixed matrices or vectors. $\omega \in R^r$ and $h(\cdot)$ are random vectors. $p(\cdot): R^r \to R_+$ is a continuously differentiable probability density function.

Let $\mathcal{F} = \{x \in R^n \mid c(x) \le 0\}$ and $\mathcal{Z} = \{y \in R^s \mid Wy \le q\}$. We denote the active constraint by $I_0(x) = \{i \in I \mid c_i(x) = 0\}$, where $I = \{1, \ldots, m\}$. Throughout the paper, the following hypotheses hold.

Assumption 1. $\mathcal{F}$ and $\mathcal{Z}$ are bounded.

Assumption 2. At every $x \in \mathcal{F}$, the vectors $\nabla c_i(x)$, $i \in I_0(x)$ are linearly independent.

A basic difficulty of solving stochastic optimization problem (1a), (1b), (1c), and (1d) is that the objective function with uncertainty can be complicated or difficult to compute even approximately. The aim of this paper is to give computational approaches based on quasi-Monte-Carlo sampling techniques. To solve stochastic programming problems, one usually resorts to deterministic optimization methods. This idea is a natural one and was used by many authors over the years [8–12]. Deterministic methods were also applied to stochastic programming problems which involve quadratic programming in a vast literature. The extended linear quadratic programming (ELQP) model was introduced by Rockafellar and Wets [13, 14]. Qi and Womersley [15] proposed an sequence quadratic programming (SQP) algorithm for ELQP problems. To solve ELQP, Chen et al. [16] suggested a Newton-type approach and showed that this method is globally convergent and locally superlinear convergent. At the same time, Birge et al. [17] investigated a stochastic Newton method for ELQP with inequality constraint $Ax \le b$. Global convergence and local superlinear convergence of the method were established.

In order to get a numerical solution of (1a), (1b), (1c), and (1d) based on quasi-Monte-Carlo techniques, consider the following approximation of (1c):

$$Q_{n_k}(x) = \frac{1}{n_k}\sum_{i=0}^{n_k-1} \xi^i Q\left(x, \omega^i\right), \tag{2}$$

where $\omega^i \in \Omega$ and ξ^i is generated by lattice rules [18, 19]. Consequently problem (1a), (1b), (1c), and (1d) is approximated by

$$\min \quad f_{n_k}(x) = P(x) + Q_{n_k}(x), \tag{3a}$$

$$\text{subject to} \quad c(x) \le 0. \tag{3b}$$

Since $\mathscr{Z}$ is bounded, it follows from [17] that $f(x)$ is twice continuously differentiable. Moreover, from [16], the approximated objective function $f_{n_k}(x)$ has the following continuous first derivative in R^n:

$$g_{n_k}(x) = \nabla P(x) - \frac{1}{n_k} T^T \left(\sum_{i=0}^{n_k-1} \xi^i z^*\left(x, \omega^i\right)\right), \tag{4}$$

where $z^*(x,\omega^i) = \arg\max\{-(1/2)z^T G z + z^T(h(\omega^i) - Tx) \mid z \in \mathscr{Z}\}$.

Let $\{n_k\}_{k=1}^{\infty}$ be an integer sequence satisfying $1 \le n_1 \le \cdots \le n_k \le n_{k+1} \le \cdots$ and $n_k \to \infty$ as $k \to \infty$. Generate observations $\{\omega^i, i = 1, \ldots, n_k\}$ on the unit hypercube according to an integration rule. Here, we choose quasi-Monte-Carlo sequences [20]. Since $\mathscr{F}$ and $\mathscr{Z}$ are compact, it follows from [20] (or [21]) that there exists a constant $C > 0$ such that, for any $x \in \mathscr{F}$,

$$\left|f(x) - f_{n_k}(x)\right| \le \frac{C\left(\log n_k\right)^{m-1}}{n_k}, \tag{5}$$

$$\left|\nabla f(x) - g_{n_k}(x)\right| \le \frac{C\left(\log n_k\right)^{m-1}}{n_k}. \tag{6}$$

The paper addresses a feasible sequential system of linear equations (SSLE) approach to solve (1a), (1b), (1c), and (1d). This study is strongly motivated by recent successful development of various SSLE algorithms for deterministic optimization problems and quasi-Monte-Carlo simulation techniques. SSLE methods for deterministic optimization problems have been proposed by many authors over the years. An interested reader is referred to the literature [22–26] for excellent surveys. Our algorithm has the following interesting features.

(a) Without assuming isolatedness of the accumulation point or boundedness of the Lagrange multiplier approximation sequence, every accumulation point of the iterative sequence generated by the proposed algorithm converges to a KKT point of problem (1a), (1b), (1c), and (1d).

(b) At each iteration, we only to solve four symmetric systems of linear equations with a common coefficient matrix and a simple structure. In the proposed algorithm the last system of linear equation only needs to be solved for achieving a local one-step superlinear convergence rate.

(c) In order to achieve the "working set," the multiplier function $\lambda(x) := (\nabla c(x)^T \nabla c(x) + \mathrm{diag}(c_i^2(x)))^{-1} \nabla c(x)^T \nabla f(x)$ is needed to be obtained firstly in [27]. The multiplier function also is suggested by Facchinei et al. [28], while our algorithm provides a new technique to update the "working set," consequently, without calculating the multiplier function.

(d) In order to find a search direction, a quadratic programming subproblem needs to be solved at each iteration in [17]. Consequently, the Hessian of objective function needs to be approximated by Monte Carlo (or quasi-Monte-Carlo) rule, while for the SSLE methods the approximation is not necessary. Our algorithm solves four linear systems of equations with only the first-order derivative of objective function involved.

The remainder of this paper is organized as follows. Section 2 gives the algorithm of (1a), (1b), (1c), and (1d) and shows the proposed algorithm is well defined. In Section 3 we discuss the convergence of algorithm in detail. We proceed in Section 4 by showing the local superlinear convergence. Finally, our conclusions are presented in Section 5.

2. Algorithm

The Lagrangian function associated with problem (1a), (1b), (1c), and (1d) is defined by

$$L(x, \lambda) = f(x) + \lambda^T c(x). \tag{7}$$

A point x^* in $\mathscr{F}$ is called a KKT point of problem (1a), (1b), (1c), and (1d), if there exits λ^* such that the following KKT conditions hold:

$$\begin{aligned} \nabla_x L\left(x^*, \lambda^*\right) = 0, &\quad \lambda^* \ge 0, \\ c_i\left(x^*\right)\lambda_i^* = 0, &\quad \forall i \in I, \end{aligned} \tag{8}$$

where

$$\nabla_x L(x, \lambda) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla c_i(x). \tag{9}$$

For $x, y \in \mathscr{F}$, let

$$I(x, y, \lambda, \varepsilon) := \left\{i \in I \mid c_i(x) + \varepsilon\rho(y, \lambda) > 0\right\}, \tag{10}$$

where ε is a nonnegative parameter and $\rho(y,\lambda) = \sqrt{\|\Phi(y,\lambda)\|}$ with

$$\Phi(y, \lambda) := \begin{pmatrix} \nabla_y L(y, \lambda) \\ \min\{-c(y), \lambda\} \end{pmatrix}. \tag{11}$$

From the definition of $\rho(x,\lambda)$, $\rho(x^*,\lambda^*) = 0$ if and only if (x^*,λ^*) satisfies KKT conditions (8). In order to achieve the active constraint set in our algorithm, the estimate of set $I(x,y,\lambda,\varepsilon)$ is defined by

$$I_{n_k}(x,y,\lambda,\varepsilon) := \{i \in I \mid c_i(x) + \varepsilon\psi_{n_k}(y,\lambda) > 0\}, \quad (12)$$

where

$$\psi_{n_k}(y,\lambda) = \left\{2\left[\left(\frac{1}{n_k^{\delta_2}}\right)^2 + \left(\rho_{n_k}(y,\lambda)\right)^4\right]\right\}^{1/4},$$
$$\rho_{n_k}(y,\lambda) = \sqrt{\left\|\Phi_{n_k}(y,\lambda)\right\|},$$
$$\Phi_{n_k}(y,\lambda) := \begin{pmatrix} \nabla_y L_{n_k}(y,\lambda) \\ \min\{-c(y),\lambda\} \end{pmatrix}, \quad (13)$$
$$\nabla_y L_{n_k}(y,\lambda) = g_{n_k}(y) + \sum_{i=1}^{m}\lambda_i \nabla c_i(y),$$

and δ_2 is a positive parameter in $(1/2,1)$. Since $f(x)$ and $c(x)$ are continuously differentiable, it follows from Theorem 3.15 in [28] that $\rho_{n_k}(x,\lambda)$ is nonnegative and continuous on R^{n+m}. Hence, from (6) and continuous differentiability of $f_{n_k}(x)$, we have that $\rho_{n_k}(y,\lambda) \to \rho(x^*,\lambda^*)$, as $n_k \to \infty$, $(y,\lambda) \to (x^*,\lambda^*)$.

For simplicity, let $I(n_k,k+1,\varepsilon) := I_{n_k}(x^{k+1},x^k,\lambda^k,\varepsilon)$, and

$$M(n_k,k+1) := \begin{pmatrix} H_{k+1} & \nabla c(n_k,x^{k+1}) \\ \nabla c(n_k,x^{k+1})^T & 0 \end{pmatrix}, \quad (14)$$

where

$$\nabla c(n_k,x^{k+1}) := \left(\nabla c_i(x^{k+1}) \mid i \in I(n_k,k+1,\varepsilon)\right). \quad (15)$$

Now we formally state our algorithm.

Algorithm 3.

(S.0) (Initialization)

Parameters: $\sigma \in (0,1)$, $\sigma_1 \in (0,1)$, $\theta \in (0,1)$, $\eta \in (2,3)$, $u \in (0,1/2)$, $\beta \in (0,1)$, $\delta_1 \in (0,1/2)$, $\delta_2 = 1-\delta_1$;

Data: $0 < w_0 < 1$, $M > 0$, $\lambda^0 = 0 \in R^m$, $\varepsilon_0 > 0$, symmetric positive define matrix $H_0 \in R^{n\times n}$, $x^0 = x^1 \in \mathcal{F}$, and $c_i(x^0) < 0$, for every $i \in I$. Sequence $\{\alpha_k\}$ satisfies $\alpha_k > 0$, for all k, and $\sum_{k=0}^{\infty}\alpha_k < +\infty$;

Choose n_0, n_1 such that $1/n_0^{\delta_2} \in (0,\alpha_0)$, and $1/n_1^{\delta_2} \in (0,\alpha_1)$;

Generate observations $\{\omega^i: i = 0,\ldots,n_0\}$ by quasi-Monte-Carlo rules and calculate $g_{n_0}(x^0)$;

Set $k := 1$.

(S.1) (Choose Working Set)

(S1.1) If $\psi_{n_{k-1}}(x^{k-1},\lambda^{k-1}) > M$, then set $\psi_{n_{k-1}}(x^{k-1},\lambda^{k-1}) = M$.

(S1.2) Set $\varepsilon := \varepsilon_k$, $w := w_k$.

(S1.3) Calculate $I(n_{k-1},k,\varepsilon_k)$ and $M(n_{k-1},k)$.

(S1.4) If $\|(M(n_{k-1},k))\| < w$, then set $\varepsilon = \sigma\varepsilon$, $w = \sigma_1 w$, and go to (S1.3).

(S1.5) Set $\varepsilon_{k+1} = \varepsilon$, $w_{k+1} = w$.

(S.2) (Computation of Search Direction)

If $I(n_{k-1},k,\varepsilon_k) = \emptyset$, then run the following step (S2.1)–(S2.4); otherwise go to (S2.5).

(S2.1) Set $l := 0$.

(S2.2) Generate observations $\{\omega^i: i = 0,\ldots,n_k+l\}$ by quasi-Monte-Carlo rules and calculate $g_{n_k+l}(x^k)$.

(S2.3) Set $d^{k_0} = d^{k_1} = d^{k_2} = d^{k_3} = -H_k g_{n_k+l}(x^k)$.

(S2.4) If $1/(n_k+l)^{\delta_2} > M\|d^{k_2}\|$, then set $l = l+1$ and go to (S2.2); otherwise set $\overline{n}_k = n_k + l$, $\lambda^k = 0$, $l_k = l$, and go to (S.3).

(S2.5) Set $l := 0$.

(S2.6) Generate observations $\{\omega^i: i = 0,\ldots,n_k+l\}$ by quasi-Monte-Carlo rules and calculate $g_{n_k+l}(x^k)$.

(S2.7) Compute (d^{k_0},λ^{k_0}) by solving the system of linear equation in (d,λ)

$$M(n_{k-1},k)\begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -g_{n_k+l}(x^k) \\ 0 \end{pmatrix}. \quad (16)$$

Set $\lambda^k := \binom{\lambda^{k_0}}{0} \in R^m$.

(S2.8) Let

$$v_i^k = \begin{cases} \lambda_i^{k_0}, & i \in \Gamma_{k_0}^-, \\ \min\{-c_i(x^k),\lambda_i^{k_0}\}, & i \in I(n_{k-1},k,\varepsilon_k) \setminus \Gamma_{k_0}^-, \end{cases} \quad (17)$$

where $\Gamma_{k_0}^- = \{i \in I(n_{k-1},k,\varepsilon) \mid \lambda_i^{k_0} < 0\}$.

Compute (d^{k_1},λ^{k_1}) by solving the system of linear equation in (d,λ)

$$M(n_{k-1},k)\begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -g_{n_k+l}(x^k) \\ v^k \end{pmatrix}. \quad (18)$$

If $1/(n_k+l)^{\delta_2} > M\|d^{k_1}\|$, then set $l = l+1$ and go to (S2.6); otherwise set $\overline{n}_k = n_k + l$, and $l_k = l$.

(S2.9) Compute (d^{k_2},λ^{k_2}) by solving the system of linear equation in (d,λ)

$$M(n_{k-1},k)\begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -g_{n_k+l}(x^k) \\ \mu^k \end{pmatrix}, \quad (19)$$

where

$$\mu^k = \upsilon^k - \rho_k e,$$

$$\rho_k = \frac{(\theta - 1)\, g_{n_k+l}\left(x^k\right)^T d^{k_1}}{1 + \sum_{i\in I(n_{k-1},k,\varepsilon_k)} \left|\lambda_i^{k_0}\right| \left\|d^{k_1}\right\|^{\eta}} \left\|d^{k_1}\right\|^{\eta}. \quad (20)$$

(S2.10) Compute (d^{k_3}, λ^{k_3}) by solving the system of linear equation in (d, λ)

$$M\left(n_{k-1}, k\right)\begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -g_{n_k+l}\left(x^k\right) \\ \omega^k \end{pmatrix}, \quad (21)$$

where

$$\omega^k = \nabla c\left(n_{k-1}, x^k\right)^T d^{k_2} - c\left(n_{k-1}, x^k + d^{k_2}\right) - \left\|d^{k_2}\right\|^{\eta} e. \quad (22)$$

(S.3) If $\|d^{k_3} - d^{k_2}\| > \|d^{k_2}\|$, then set $d^{k_3} = d^{k_2}$.

Choose t_k, the first number t in the sequence $\{1, \beta, \beta^2, \ldots\}$ satisfying

$$f_{\bar{n}_k}\left(x^k + td^{k_2} + t^2\left(d^{k_3} - d^{k_2}\right)\right) - f_{\bar{n}_k}\left(x^k\right) \leq ut g_{\bar{n}_k}\left(x^k\right)^T d^{k_2} + \alpha_k, \quad (23)$$

$$c_i\left(x^k + td^{k_2} + t^2\left(d^{k_3} - d^{k_2}\right)\right) < 0, \quad i \in I. \quad (24)$$

(S.4) Compute N_k such that $1/N_k^{\delta_2} \in (0, \alpha_{k+1})$.

Set $n_{k+1} = \max\{N_k, \bar{n}_k\}$, and $x^{k+1} := x^k + t_k d^{k_2} + t_k^2(d^{k_3} - d^{k_2})$. Generate a new symmetric positive define matrix H_{k+1}. Set $k := k + 1$ and go to (S.1).

Remarks

(a) The main purpose of (S.1) is to generate a working set and ensure that the matrix $M(n_{k-1}, k)$ is nonsingular, for every k. Hence, (d^{k_i}, λ^{k_i}) is well defined, for all $i \in \{0, 1, 2, 3\}$. The calculation of set $I(n_{k-1}, k, \varepsilon_k)$ specially is different from the one proposed in [27]. We use the solution λ^{k_1} of system (18) as a substitute for the multiplier function proposed in [27]. Moreover, $M(n_{k-1}, k)$ is also uniformly bounded. Details will subsequently be given.

(b) From the construction of the algorithm, four linear systems need to be solved at each iteration. To ensure the iterate sequence globally converges to KKT point of (1a), (1b), (1c), and (1d), we only need to solve the previous three linear systems (16), (18), and (19). The linear systems (16) and (19) play important roles in proving the global convergence. The main aim of the linear system (21) is to guarantee the one-step superlinear convergence rate of the algorithm under mild conditions.

(c) It is not difficult to show that there exists t_k, the first number of the sequence $\{1, \beta, \beta^2, \ldots\}$, which satisfies the linear search (23) and (24). In Section 4 we will show that $t_k = 1$, for sufficiently large k. Hence, the Maratos effect will be avoided.

(d) In numerical experiments H_k is usually updated by the Broyden-Fletcher-Goldfarb-Shanno (BFGS) formula [27, 29]. At any iteration k Algorithm 3 stops as the following termination criteria, with $\varepsilon_{\text{stop}} \in (0, 1)$ and maximum iterations $N_{\max}$:

$$\text{(i)} \quad \left\|\Phi_{n_k}\left(x^k, \lambda^k\right)\right\| < \varepsilon_{\text{stop}}, \qquad \text{or (ii)} \quad k = N_{\max}. \quad (25)$$

The rest of section is devoted to show that Algorithm 3 is well defined. We firstly give the following hypothesis on the choice of the matrix H_k.

Assumption 4. There exist positive constants C_1 and C_2 such that for all k and $d \in R^n$

$$C_1 \|d\|^2 \leq d^T H_k d \leq C_2 \|d\|^2. \quad (26)$$

It is not difficult to see from $c_i(x^k) < 0$, for every k in nonnegative integer set $\mathcal{N}$, the inner iteration (S.1) terminates finitely.

Lemma 5. $\nabla f(x^k) = 0$, *if there exists some k such that the following conditions hold.*

(a) $I(n_{k-1}, k, \varepsilon_k) = \emptyset$, (b) $1/(n_k + l)^{\delta_2} > M\|d^{k_2}\|$ for all $l \geq 0$.

Proof. From condition (b), we have that $d^{k_2} \to 0$. It follows that

$$\left\|\nabla f\left(x^k\right)\right\| \leq \left\|\nabla f\left(x^k\right) - g_{n_k+l}\left(x^k\right)\right\| + \left\|g_{n_k+l}\left(x^k\right)\right\| \leq \frac{C\left[\log\left(n_k + l\right)\right]^{m-1}}{n_k + l} + \left\|H_k d^{k_2}\right\| \longrightarrow 0. \quad (27)$$

From independence of k and l, the result follows. □

Lemma 6. $\nabla f(x^k) = 0$, *if there exists some k such that the following conditions hold.*

(a) $I(n_{k-1}, k, \varepsilon_k) \neq \emptyset$, (b) $1/(n_k+l)^{\delta_2} > M\|d^{k_1}\|$ for all $l \geq 0$.

Proof. From condition (b), $d^{k_1} \to 0$. It follows that, as $l \to \infty$

$$g_{n_k+l}\left(x^k\right)^T d^{k_1} = -d^{k_0} H_k d^{k_0} - \sum_{i\in\Gamma_{k_0}^-} \left(\lambda_i^{k_0}\right)^2 - \sum_{i\in I(n_{k-1},k,\varepsilon_k)\setminus\Gamma_{k_0}^-} \lambda_i^{k_0} \min\left\{-c_i\left(x^k\right), \lambda_i^{k_0}\right\} = 0. \quad (28)$$

Therefore, we get $d^{k_0} \to 0$, and

$$\lambda_i^{k_0} \longrightarrow 0 \quad \text{if } i \in \Gamma_{k_0}^-, \qquad \lambda_i^{k_0} c_i\left(x^k\right) \longrightarrow 0 \quad \text{if } i \in I\left(n_{k-1}, k, \varepsilon_k\right) \setminus \Gamma_{k_0}^-. \tag{29}$$

Since $c_i(x^k) < 0$, $\lambda^{k_0} \to 0$, we have, as $l \to \infty$,

$$\begin{aligned}\left\|\nabla f\left(x^k\right)\right\| &\le \left\|\nabla f\left(x^k\right) - g_{n_k+l}\left(x^k\right)\right\| + \left\|g_{n_k+l}\left(x^k\right)\right\| \\ &= \left\|\nabla f\left(x^k\right) - g_{n_k+l}\left(x^k\right)\right\| \\ &\quad + \left\|\nabla c\left(n_{k-1}, x^k\right)\lambda^{k_0} + H_k d^{k_0}\right\| \longrightarrow 0.\end{aligned} \tag{30}$$

This completes the proof. □

It is easy to see from Lemmas 5 and 6 that x^k is a unconstrained stationary point of f, if we are not able to get the next iteration x^{k+1} from the current iteration x^k; that is, the inner iterations (S2.1)–(S2.8) terminate infinitely. Since we always have $x^k \in \mathcal{F}$, this means that x^k is actually a KKT point of problem (1a), (1b), (1c), and (1d). In the following section, we assume that the inner iterations (S2.1)–(S2.8) terminate finitely for all $k \in \mathcal{N}$; namely, there always exists $l_0 \in \mathcal{N}$ such that, for every $k \in \mathcal{N}$, one of the following conditions holds.

(i) $I(n_{k-1}, k, \varepsilon_k) = \emptyset$, $1/(n_k + l_0)^{\delta_2} \le M\|d^{k_2}\|$.

(ii) $I(n_{k-1}, k, \varepsilon_k) \ne \emptyset$, $1/(n_k + l_0)^{\delta_2} \le M\|d^{k_1}\|$.

Therefore, the algorithm generates an infinite iterative sequence $\{x^k\}$.

Lemma 7. *If there exists k_0 such that $I(n_{k-1}, k, \varepsilon_k) = \emptyset$ for all $k > k_0$, then there exists $\bar{\varepsilon} > 0$ such that $\varepsilon_k \ge \bar{\varepsilon}$ for all $k \in \mathcal{N}$.*

Proof. Since $I(n_{k-1}, k, \varepsilon_k) = \emptyset$, we have that $\|M(n_{k-1}, k)^{-1}\| = \|H_k^{-1}\|$. So the result follows from Assumption 4. □

Lemma 8. *If there exist k_0 and subset $K \subset \mathcal{N}$ such that $I(n_{k-1}, k, \varepsilon_k) \ne \emptyset$ with $k \in K$ and all $k > k_0$, then $\varepsilon_{k+1} = \varepsilon_k$ for sufficiently large k.*

Proof. Assume to contrary that for any $k_0 \in K$ there always exists $k > k_0$ such that $\varepsilon_{k+1} \ne \varepsilon_k$. From construction of the algorithm, we have that $\varepsilon_k \to 0$. By Assumption 1 and the finiteness of set I, without loss of generality, we can assume that

(i) $\{x^k\}_K \to x^*$ with $x^* \in \mathcal{F}$ and $\varepsilon_{k+1} < \varepsilon_k$ for all $k \in K$;

(ii) $I(n_{k-1}, k, \varepsilon_k)$, $k \in K$ keep changeless.

For simplicity, let $I_K := I(n_{k-1}, k, \varepsilon_k)$. Since $\psi_{n_{k-1}}(x^{k-1}, \lambda^{k-1})$ is bounded, it follows from $\varepsilon_k \to 0$ that $I_K \subset I_0(x^*)$ for sufficiently large k. Hence, by Assumption 4 and step (S.1), we get

$$\begin{aligned}\left\|\nabla c_{I_K}\left(x^k\right)^T \nabla c_{I_K}\left(x^k\right)\right\| &\longrightarrow \left\|\nabla c_{I_K}\left(x^*\right)^T \nabla c_{I_K}\left(x^*\right)\right\| \\ &= 0,\end{aligned} \tag{31}$$

which contradicts with Assumption 2, and the proof is complete. □

From Lemmas 7 and 8 we can directly obtain Lemma 9.

Lemma 9. *There exists $\bar{\varepsilon} > 0$ such that $\varepsilon_k > \bar{\varepsilon}$ for all k.*

Since $\mathcal{F}$ is compact, we get Lemma 10.

Lemma 10. *$M(n_{k-1}, k)$ is nonsingular and uniformly bounded with respect to $k \in \mathcal{N}$; that is, there exists $\overline{W} > 0$ and $\overline{M} > 0$ such that, for all $k \in \mathcal{N}$,*

$$\overline{W} \le \det\left(M\left(n_{k-1}, k\right)\right) \le \overline{M}. \tag{32}$$

From Assumption 1 and Lemma 10, the following lemma is then obvious.

Lemma 11. *$\{(d^{k_j}, \lambda^{k_j})\}$ are bounded for $j = 0, 1, 2, 3$.*

Lemma 12. *If $I(n_{k-1}, k, \varepsilon_k) \ne \emptyset$, the following results hold.*

(a) $g_{n_k+l}(x^k)^T d^{k_0} = -d^{k_0} H_k d^{k_0}$ *for* $0 \le l \le l_k$.

(b) $g_{n_k+l}(x^k)^T d^{k_1} \le g_{n_k+l}(x^k)^T d^{k_0}$ *for* $0 \le l \le l_k$.

(c) $g_{n_k+l_k}(x^k)^T d^{k_2} \le \theta g_{n_k+l_k}(x^k)^T d^{k_1}$.

Proof. (a) is a direct consequence of linear system (16). It is easy to see from linear systems (16), (18), and (19) that

$$\begin{aligned}{d^{k_0}}^T H_k d^{k_1} &= -g_{n_k+l}\left(x^k\right)^T d^{k_0}, \\ {d^{k_1}}^T H_k d^{k_0} + {\lambda^{k_0}}^T v^k &= -g_{n_k+l}\left(x^k\right)^T d^{k_1}, \\ {d^{k_0}}^T H_k d^{k_2} &= -g_{n_k+l_k}\left(x^k\right)^T d^{k_0}, \\ {d^{k_2}}^T H_k d^{k_0} + {\lambda^{k_0}}^T \mu^k &= -g_{n_k+l_k}\left(x^k\right)^T d^{k_2}.\end{aligned} \tag{33}$$

Therefore, we have

$$\begin{aligned}g_{n_k+l}\left(x^k\right)^T d^{k_1} &= g_{n_k+l}\left(x^k\right)^T d^{k_0} - {\lambda^{k_0}}^T v^k \\ &= g_{n_k+l}\left(x^k\right)^T d^{k_0} - \sum_{i \in \Gamma_{k_0}^-}\left(\lambda_i^{k_0}\right)^2 \\ &\quad - \sum_{i \in I(n_{k-1}, k, \varepsilon_k) \setminus \Gamma_{k_0}^-} \lambda_i^{k_0} \min\left\{-c_i\left(x^k\right), \lambda_i^{k_0}\right\} \\ &\le g_{n_k+l}\left(x^k\right)^T d^{k_0}, \\ g_{n_k+l_k}\left(x^k\right)^T d^{k_2} &= g_{n_k+l_k}\left(x^k\right)^T d^{k_1} + {\lambda^{k_0}}^T\left(v^k - \mu^k\right) \\ &= g_{n_k+l_k}\left(x^k\right)^T d^{k_1} \\ &\quad + \frac{(\theta - 1)\, g_{n_k+l_k}\left(x^k\right)^T d^{k_1} \sum_{i \in I(n_{k-1}, k, \varepsilon_k)}\left|\lambda_i^{k_0}\right| \left\|d^{k_1}\right\|^\eta}{1 + \sum_{i \in I(n_{k-1}, k, \varepsilon_k)}\left|\lambda_i^{k_0}\right| \left\|d^{k_1}\right\|^\eta} \\ &\le \theta g_{n_k+l_k}\left(x^k\right)^T d^{k_1}.\end{aligned} \tag{34}$$

This completes the proof. □

3. Convergence

Lemma 13. *Suppose the following conditions hold.*

(i) $\{x^k\}_K \to x^*$.

(ii) $I(n_{k-1}, k, \varepsilon_k) = \emptyset$ *for every* $k \in K$.

(iii) *There exists* $K_0 \subset K$ *such that* $\{x^{k-1}, \lambda^{k-1}\}_{K_0} \to (\overline{x}^*, \overline{\lambda}^*)$*, and* $\rho(\overline{x}^*, \overline{\lambda}^*) \neq 0$.

Then x^* *is a stationary point; namely,* $\nabla f(x^*) = 0$.

Proof. We show the conclusion by contradiction. Suppose that $\nabla f(x^*) \neq 0$. Without loss of generality, we assume that $\{d^{k_2}\}_K \to \overline{d} \neq 0$ and $H_k \to H^*$. So there exists $\gamma_1 > 0$ such that for sufficiently large $k \in K$

$$\begin{aligned} & g_{n_k+l_k}\left(x^k\right)^T d^{k_2} \\ & = \left(g_{n_k+l_k}\left(x^k\right) - \nabla f\left(x^k\right)\right)^T d^{k_2} \\ & \quad + \nabla f\left(x^k\right)^T H_k^{-1} g_{n_k+l_k}\left(x^k\right) \\ & \leq -\nabla f\left(x^k\right)^T H_k^{-1} \nabla f\left(x^k\right) \\ & \quad + \frac{C\left(\log\left(n_k+l_k\right)\right)^{m-1}}{n_k+l_k}\left[\left\|d^{k_2}\right\| + \left\|\nabla f\left(x^k\right)^T H_k^{-1}\right\|\right] \\ & \leq -\gamma_1. \end{aligned} \tag{35}$$

Since $\rho(\overline{x}^*, \overline{\lambda}^*) \neq 0$, for sufficiently big $k \in K_0$, there exists $\rho_0 > 0$ such that

$$\psi_{n_{k-1}}\left(x^{k-1}, \lambda^{k-1}\right) > \rho_0. \tag{36}$$

So

$$c_i\left(x^k\right) \leq -\overline{\varepsilon}\psi_{n_{k-1}}\left(x^{k-1}, \lambda^{k-1}\right) \leq -\overline{\varepsilon}\rho_0. \tag{37}$$

Let $\gamma = \min\{\gamma_1, \overline{\varepsilon}\rho_0\}$. Since $\{d^{k_2}\}_K \to \overline{d} \neq 0$, it is obvious that $o(\|td^k\|) = o(t)$. So we have that, for sufficiently large $k \in K$,

$$\begin{aligned} & f_{n_k+l_k}\left(x^k + td^{k_2}\right) - f_{n_k+l_k}\left(x^k\right) \\ & \quad \leq f\left(x^k + td^{k_2}\right) - f\left(x^k\right) \\ & \qquad + \left(f_{n_k+l_k}\left(x^k + td^{k_2}\right) - f\left(x^k + td^{k_2}\right)\right) \\ & \qquad + \left(f\left(x^k\right) - f_{n_k+l_k}\left(x^k\right)\right) \\ & \quad \leq f\left(x^k + td^{k_2}\right) - f\left(x^k\right) + 2 \\ & \qquad \cdot \frac{C\left(\log\left(n_k+l_k\right)\right)^{m-1}}{n_k+l_k} \\ & \quad = t\nabla g_{n_k+l_k}\left(x^k\right)^T d^{k_2} \\ & \qquad + t\left(\nabla f\left(x^k\right) - \nabla g_{n_k+l_k}\left(x^k\right)\right)^T d^{k_2} + o(t) + 2 \\ & \qquad \cdot \frac{C\left(\log\left(n_k+l_k\right)\right)^{m-1}}{n_k+l_k} \\ & \quad \leq t\nabla g_{n_k+l_k}\left(x^k\right)^T d^{k_2} + o(t) + \left(t\left\|d^{k_2}\right\| + 2\right) \\ & \qquad \cdot \frac{C\left(\log\left(n_k+l_k\right)\right)^{m-1}}{n_k+l_k} \\ & \quad \leq ut\nabla g_{n_k+l_k}\left(x^k\right)^T d^{k_2} - (1-u)t\gamma + o(t) \\ & \qquad + \left(t\left\|d^{k_2}\right\| + 2\right) \cdot \frac{C\left(\log\left(n_k+l_k\right)\right)^{m-1}}{n_k+l_k}. \end{aligned} \tag{38}$$

Therefore, we have

$$\begin{aligned} & f_{n_k+l_k}\left(x^k + td^{k_2}\right) - f_{n_k+l_k}\left(x^k\right) \\ & \quad \leq ut\nabla g_{n_k+l_k}\left(x^k\right)^T d^{k_2} - (1-u)t\gamma + o(t) + \alpha_k. \end{aligned} \tag{39}$$

By (37), for all $i \in I$

$$\begin{aligned} c_i\left(x^k + td^{k_2}\right) & = c_i\left(x^k\right) + t\nabla c_i\left(x^k\right)^T d^{k_2} + o(t) \\ & \leq -\gamma + t\nabla c_i\left(x^k\right)^T d^{k_2} + o(t). \end{aligned} \tag{40}$$

It follows that from (39) and (40) that there exists $\overline{t} > 0$ independent of k such that, for any $t \in (0, \overline{t}]$, both (23) and (24) hold. From (39), there exists k_0 such that, for all $k \geq k_0$ with $k \in K, \overline{t} \geq t_k \geq \beta\overline{t}$,

$$f_{n_k+l_k}\left(x^k + t_k d^{k_2}\right) - f_{n_k+l_k}\left(x^k\right) \leq -u\overline{t}\beta\gamma + \alpha_k. \tag{41}$$

It is not difficult to see from (23) and Lemma 12 that, for sufficiently large k,

$$\begin{aligned} & f\left(x^k + t_k d^{k_2}\right) - f\left(x^k\right) \\ & \quad = f_{n_k+l_k}\left(x^k + t_k d^{k_2}\right) - f_{n_k+l_k}\left(x^k\right) \\ & \qquad + \left(f\left(x^k + t_k d^{k_2}\right) - f_{n_k+l_k}\left(x^k + t_k d^{k_2}\right)\right) \\ & \qquad + \left(f_{n_k+l_k}\left(x^k\right) - f\left(x^k\right)\right) \\ & \quad \leq f_{n_k+l_k}\left(x^k + t_k d^{k_2}\right) - f_{n_k+l_k}\left(x^k\right) + \alpha_k \\ & \quad \leq ut_k g_{n_k+l_k}\left(x^k\right)^T d^{k_2} + \alpha_k + \alpha_k \leq 2\alpha_k. \end{aligned} \tag{42}$$

Combining with (41), we get

$$\sum_{k=k_0}^{\infty}\left(f\left(x^{k+1}\right)-f\left(x^{k}\right)\right)\le\sum_{k=k_0,\,k\notin K}^{\infty}2\alpha_k+\sum_{k\ge k_0,\,k\in K}^{\infty}\left(f_{n_k+l_k}\left(x^k+t_kd^{k_2}\right)-f_{n_k+l_k}\left(x^k\right)+\alpha_k\right)\le\sum_{k=k_0}^{\infty}4\alpha_k+\sum_{k\ge k_0,\,k\in K}^{\infty}\left(-u\bar{t}\beta\gamma\right)\longrightarrow-\infty. \quad (43)$$

It follows that $f(x^k)\to-\infty$, which contradicts with the fact that $\{f(x^k)\}$ is bounded, and the proof is complete. □

Lemma 14. *Suppose the following conditions hold.*

(i) $\{x^k\}_K\to x^*$.

(ii) $I(n_{k-1},k,\varepsilon_k)=\emptyset$ *for every* $k\in K$.

(iii) *There exists* $K_0\subset K$ *such that* $\{x^{k-1},\lambda^{k-1}\}_{K_0}\to(\overline{x}^*,\overline{\lambda}^*)$, *and* $\rho(\overline{x}^*,\overline{\lambda}^*)=0$.

If $I(n_{k-2},k-1,\varepsilon_{k-1})=\emptyset$ *for every* $k\in K_0$, *then* x^* *is a stationary point; namely,* $\nabla f(x^*)=0$.

Proof. From the above conditions, we have that, for every $k\in K_0$, $\lambda^k=\lambda^{k-1}=0$, $\nabla f(\overline{x}^*)=0$, and therefore

$$\left\{d^{k_j-1}\right\}_{K_0}\longrightarrow H^{*-1}\nabla f\left(\overline{x}^*\right)=0,\quad j=0,1,2,3. \quad (44)$$

It follows that as $k\to\infty$ and $k\in K_0$

$$x^k=x^{k-1}+t_kd^{k_2-1}+t_k^2\left(d^{k_3-1}-d^{k_2-1}\right)\longrightarrow\overline{x}^*=x^*. \quad (45)$$

This completes the proof. □

Let π^{k_j}, $j=0,1,2,3$, denote the vectors on R^m with components $\pi_i^{k_j}$, respectively, where

$$\pi_i^{k_j}:=\begin{cases}\lambda_i^{k_j}, & i\in I\left(n_{k-1},k,\varepsilon_k\right),\\ 0, & i\in I\setminus I\left(n_{k-1},k,\varepsilon_k\right).\end{cases} \quad (46)$$

Lemma 15. *Suppose conditions (i)–(iii) hold in Lemma 14. If* $I(n_{k-2},k-1,\varepsilon_{k-1})\neq\emptyset$ *for every* $k\in K_0$, *then* x^* *is a stationary point; namely,* $\nabla f(x^*)=0$.

Proof. Without loss of generality, we suppose that $I_K:=I(n_{k-1},k,\varepsilon_k)$ and $I_{K_0}:=I(n_{k-2},k-1,\varepsilon_{k-1})$ keep changeless.

From condition (iii) in Lemma 14

$$\left\{\rho_{n_k}\left(x^{k-1},\lambda^{k-1}\right)\right\}_{K_0}\longrightarrow0,\qquad \left\{\psi_{n_{k-1}}\left(x^{k-1},\lambda^{k-1}\right)\right\}_{K_0}\longrightarrow0. \quad (47)$$

Combining with the first equation of linear system (16), $\{d^{k_0-1}\}_{K_0}\to0$. It is easy to see from (47) that

$$\left\{\min\left\{-c_i\left(x^{k-1}\right),\lambda_i^{k_0-1}\right\}\right\}_{K_0}\longrightarrow0\quad\text{for } i\in I\left(n_{k-2},k-1,\varepsilon_{k-1}\right)\setminus\Gamma_{k_0-1}^-,\qquad \left\{\lambda_i^{k_0-1}\right\}_{K_0}\longrightarrow0\quad\text{for } i\in\Gamma_{k_0-1}^-. \quad (48)$$

Therefore, we have

$$\left\{d^{k_j-1}\right\}_{K_0}\longrightarrow0,\qquad \left\{\pi^{k_j-1}\right\}_{K_0}\longrightarrow\overline{\lambda}^*,\qquad j=1,2,3. \quad (49)$$

So we get

$$x^k=x^{k-1}+t_kd^{k_2-1}+t_k^2\left(d^{k_3-1}-d^{k_2-1}\right)\longrightarrow\overline{x}^*, \quad (50)$$

and for sufficiently large $k\in K_0$

$$\left\|d^{k_2-1}\right\|=O\left(\left\|d^{k_1-1}\right\|\right),\qquad \left\|d^{k_3-1}\right\|=O\left(\left\|d^{k_1-1}\right\|\right). \quad (51)$$

Let $I_0^+(\overline{x}^*)=\{i\in I\mid\overline{\lambda}_i^*>0\}$. Since $(\overline{x}^*,\overline{\lambda}^*)$ is a KKT pair of problem (1a), (1b), (1c), and (1d), we have

$$I_0^+\left(\overline{x}^*\right)\subset I_{K_0}. \quad (52)$$

Therefore, from (50)

$$x^*=\overline{x}^*,\qquad c_i\left(x^*\right)=c_i\left(\overline{x}^*\right)=0,\quad i\in I_0^+\left(\overline{x}^*\right). \quad (53)$$

If there is $M>0$ such that $\|d^{k_0-1}\|\ge M\|d^{k_1-1}\|$, then we have from (51) that for arbitrary $i\in I_0^+(\overline{x}^*)$ and sufficiently large $k\in K_0$

$$c_i\left(x^k\right)+\bar{\varepsilon}\psi_{n_{k-1}}\left(x^{k-1},\lambda^{k-1}\right)=c_i\left(x^k\right)+\bar{\varepsilon}\left\{2\left[\left(\frac{1}{n_{k-1}^{\delta_2}}\right)^2+\rho_{n_{k-1}}\left(x^{k-1},\lambda^{k-1}\right)^4\right]\right\}^{1/4}\ge c_i\left(x^k\right)+2^{1/4}\bar{\varepsilon}\left[\left\|\nabla_xL_{n_{k-1}}\left(x^{k-1},\lambda^{k-1}\right)\right\|^2+\sum_{i=1}^{m}\left|\min\left\{-c_i\left(x^{k-1}\right),\lambda_i^{k-1}\right\}\right|^2\right]^{1/4}=c_i\left(x^k\right)+2^{1/4}\bar{\varepsilon}\left[\left\|H_{k-1}d^{k_0-1}\right\|^2\right.$$

$$+\sum_{i=1}^{m}\left|\min\left\{-c_i\left(x^{k-1}\right),\lambda_i^{k-1}\right\}\right|^2\Bigg]^{1/4}\geq c_i\left(x^k\right)$$
$$+2^{-1/2}\bar{\varepsilon}\Bigg[\left\|H_{k-1}d^{k_0-1}\right\|^{1/2}$$
$$+\left(\sum_{i=1}^{m}\left|\min\left\{-c_i\left(x^{k-1}\right),\lambda_i^{k-1}\right\}\right|^2\right)^{1/4}\Bigg]>0. \tag{54}$$

For $\|d^{k_1-1}\| = 0(\|d^{k_0-1}\|)$, since $M(n_{k-1},k)$ is nonsingular

$$\left\|d^{k_1-1}\right\|=O\left(\sqrt{\sum_{i=1}^{m}\left|\min\left\{-c_i\left(x^{k-1}\right),\lambda_i^{k-1}\right\}\right|^2}\right). \tag{55}$$

So we can also get that for sufficiently large $k\in K_0$

$$c_i\left(x^k\right)+\bar{\varepsilon}\psi_{n_{k-1}}\left(x^{k-1},\lambda^{k-1}\right)>0. \tag{56}$$

So we have

$$I_0^{+}\left(\overline{x}^*\right)\subset I_K\subset I_0\left(x^*\right). \tag{57}$$

Since $I_K=\emptyset$, $I_0^{+}(\overline{x}^*)=\emptyset$. It follows that $\overline{\lambda}^*=0$, and, therefore, $\nabla f(x^*)=0$. This completes the proof. □

Lemma 16. *Suppose that $\{(x^k,\lambda^k)\}_K \to (x^*,\lambda^*)$. If $\{g_{n_k+l_k}(x^k)^T d^{k_2}\}_K\to 0$, then (x^*,λ^*) is a KKT pair of problem (1a), (1b), (1c), and (1d).*

Proof. If, for every $k\in K$, $I(n_{k-1},k,\varepsilon_k)=\emptyset$, the result can be directly obtained from $d^{k_2}=-H_k g_{n_k}(x^k)$ and $\lambda^k=0$. Without loss of generality, we suppose that, for all $k\in K$,

(i) $I(n_{k-1},k,\varepsilon_k)\neq\emptyset$,

(ii) $I_K := I(n_{k-1},k,\varepsilon_k)$ keep changeless.

By Lemma 12 and linear systems (16), (18), and (19), we have

$$\begin{aligned}g_{n_k+l_k}\left(x^k\right)^T d^{k_2} &\leq \theta g_{n_k+l_k}\left(x^k\right)^T d^{k_1}\\ &\leq -d^{k_0\,T}H_k d^{k_0}-\sum_{i\in\Gamma_{k_0}^-}\left(\lambda_i^{k_0}\right)^2\\ &\quad-\sum_{i\in I_K\setminus\Gamma_{k_0}^-}\lambda_i^{k_0}\min\left\{-c_i\left(x^k\right),\lambda_i^{k_0}\right\}\\ &\longrightarrow 0.\end{aligned} \tag{58}$$

So

$$\begin{aligned}\left\{d^{k_0}\right\}_K&\longrightarrow 0,\\ \left\{\lambda^{k_0}v^k\right\}_K&\longrightarrow 0.\end{aligned} \tag{59}$$

Let λ^* be an arbitrary accumulation point of $\{\lambda^{k_0}\}_K$. Since $\nabla f(x)$ and $\nabla c(x)$ are continuously differentiable, we get from (6), (16), and (59) that

$$\begin{aligned}\nabla f\left(x^*\right)+\nabla c\left(x^*\right)\lambda^*&=0,\\ \lambda_i^*&\geq 0,\\ \lambda_i^* c_i\left(\lambda_i^*\right)&=0,\\ c_i\left(x^*\right)&\leq 0,\\ i&\in I.\end{aligned} \tag{60}$$

This completes the proof. □

Lemma 17. *Assume that the following conditions hold:*

(i) $\{(x^k,\lambda^k)\}_K\to(x^*,\lambda^*)$.

(ii) *there exists subset $K_0\subset K$ such that $\{(x^{k-1},\lambda^{k-1})\}_{K_0}\to(\overline{x}^*,\overline{\lambda}^*)$ and $\rho(\overline{x}^*,\overline{\lambda}^*)\neq 0$.*

Then (x^*,λ^*) is a KKT pair of problem (1a), (1b), (1c), and (1d).

Proof. Assume to the contrary that (x^*,λ^*) is not a KKT pair of problem (1a), (1b), (1c), and (1d). Without loss of generality, we suppose that conditions (i) and (ii), which are given in proof for Lemma 16, hold for all $k\in K$. It is not difficult to see from Lemma 16 that there exists $\gamma_1>0$ such that, for sufficiently large $k\in K$,

$$\begin{aligned}g_{n_k+l_k}\left(x^k\right)^T d^{k_1}&<-\gamma_1,\\ g_{n_k+l_k}\left(x^k\right)^T d^{k_2}&<-\gamma_1.\end{aligned} \tag{61}$$

So that, for sufficiently large $k\in K$,

$$\begin{aligned}&f_{n_k+l_k}\left(x^k+td^{k_2}\right)-f_{n_k+l_k}\left(x^k\right)\\ &\quad\leq ut\nabla g_{n_k+l_k}\left(x^k\right)^T d^{k_2}-(1-u)t\gamma_1+o(t)+\alpha_k.\end{aligned} \tag{62}$$

From (61), $\{d^{k_1}\}_K$ does not converge to 0. Therefore, without loss of generality, we also can suppose that $\{d^{k_1}\}_K\to d_1\neq 0$ and $H_k\to H^*$. Since $\rho(\overline{x}^*,\overline{\lambda}^*)\neq 0$, for sufficiently large $k\in K_0$, there exists $\rho_0>0$ such that

$$\psi_{n_{k-1}}\left(x^{k-1},\lambda^{k-1}\right)>\rho_0. \tag{63}$$

It follows that $I_K\supset I_0(x^*)$. So, for every $i\in I_K$,

$$\begin{aligned}&c_i\left(x^k+td^{k_2}+t^2\left(d^{k_3}-d^{k_2}\right)\right)\\ &=c_i\left(x^k\right)+t\nabla c_i\left(x^k\right)^T d^{k_2}+o(t)\\ &=\begin{cases}c_i\left(x^k\right)+t\lambda_i^{k_0}-t\rho_k+o(t), & \lambda_i^{k_0}<0,\\ c_i\left(x^k\right)+t\cdot\min\left\{-c_i\left(x^k\right),\lambda_i^{k_0}\right\}-t\rho_k+o(t), & \lambda_i^{k_0}\geq 0,\end{cases}\end{aligned} \tag{64}$$

and, for $i \notin I_K$,

$$\begin{aligned} c_i\left(x^k + td^{k_2} + t^2\left(d^{k_3} - d^{k_2}\right)\right) &= c_i\left(x^k\right) + t\nabla c_i\left(x^k\right)^T d^{k_2} + o(t) \\ &\le -\frac{1}{2}\bar{\varepsilon}\rho_0 + O(t). \end{aligned} \tag{65}$$

In a way similar to the proof of Lemma 13, we get that $\{f(x^k)\}_{K_0} \to -\infty$, which contradicts with the boundedness of $f(x)$, $x \in \mathcal{F}$. This completes the proof. □

Lemma 18. *Assume that the following conditions hold.*

(i) $\{(x^k, \lambda^k)\}_K \to (x^*, \lambda^*)$.

(ii) *There exists subset* $K_0 \in K$ *such that* $\{(x^{k-1}, \lambda^{k-1})\}_{K_0} \to (\bar{x}^*, \bar{\lambda}^*)$ *and* $\rho(\bar{x}^*, \bar{\lambda}^*) = 0$.

(iii) $I(n_{k-1}, k, \varepsilon_k) \ne \emptyset$ *for every* $k \in K$.

If $I(n_{k-2}, k-1, \varepsilon_{k-1}) = \emptyset$ *for every* $k \in K_0$*, then* (x^*, λ^*) *is a KKT pair of problem (1a), (1b), (1c), and (1d).*

Proof. Without loss of generality, we suppose that $I_K := I(n_{k-1}, k, \varepsilon_k)$ keep changeless. Let

$$M(x^*) := \begin{pmatrix} H^* & \nabla c_{I_K}(x^*) \\ \nabla c_{I_K}(x^*)^T & 0 \end{pmatrix}. \tag{66}$$

By Lemma 10, $M(x^*)$ is nonsingular. Therefore, there exists $\bar{d}$ such that $\{d^{k_0}\}_K \to \bar{d}$ and $(\bar{d}, \lambda^*)$ is the unique solution of the following linear system:

$$M(x^*)\begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} \nabla f(x^*) \\ 0 \end{pmatrix}. \tag{67}$$

From Lemma 13, $\lambda^{k-1} = 0$ for every $k \in K_0$, and

$$\left\{d^{k_j - 1}\right\}_{K_0} \longrightarrow 0, \quad j = 0, 1, 2, 3, \tag{68}$$

where $d^{k_j-1} = -H_{k-1} g_{n_{k-1}+l_{k-1}}(x^{k-1})$.

So we get, as $k \to \infty$ and $k \in K_0$,

$$x^k = x^{k-1} + t_k d^{k_2-1} + t_k^2\left(d^{k_3-1} - d^{k_2-1}\right) \longrightarrow \bar{x}^* = x^*. \tag{69}$$

It follows from Lemma 13 that $\nabla f(x^*) = \nabla f(\overline{x^*}) = 0$. Therefore, we have that $\bar{d} = 0$, $\lambda^* = 0$, and the proof is complete. □

Lemma 19. *Assume that conditions (i)–(iii) in Lemma 18 hold. If* $I(n_{k-2}, k-1, \varepsilon_{k-1}) \ne \emptyset$ *for every* $k \in K_0$*, then* (x^*, λ^*) *is a KKT pair of problem (1a), (1b), (1c), and (1d).*

Proof. Without loss of generality, we suppose that $I_K := I(n_{k-1}, k, \varepsilon_k)$ and $I_{K_0} := I(n_{k-2}, k-1, \varepsilon_{k-1})$ keep changeless. In a way similar to Lemma 15, we have

$$\begin{gathered} x^* = \bar{x}^*, \\ I_0^+(\bar{x}^*) \subset I_K \subset I_0(x^*). \end{gathered} \tag{70}$$

Let $\pi^* = (\lambda_i^* \mid i \in I_K)$ and $\bar{\pi}^*$ denote a vector with the following components:

$$\bar{\pi}_i^* = \begin{cases} \bar{\lambda}_i^*, & i \in I_0^+(x^*), \\ 0, & i \in I_K \setminus I_0^+(x^*). \end{cases} \tag{71}$$

Since $(x^*, \bar{\lambda}^*)$ is a KKT pair of (1a), (1b), (1c), and (1d), we have from (70) that $\left(\begin{smallmatrix} 0 \\ \bar{\pi}^* \end{smallmatrix}\right)$ is the solution of the following linear system:

$$M(x^*)\begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} \nabla f(x^*) \\ 0 \end{pmatrix}. \tag{72}$$

On the other hand, since $\{(x^k, \lambda^k)\}_K \to (x^*, \lambda^*)$, there exists d^* such that $\{d^{k_0}\}_K \to d^*$. From Assumption 2, $M(x^*)$ is nonsingular. Therefore, (d^*, λ^*) is unique solution of the linear system (72). So we have that $\pi^* = \bar{\pi}^*$. So $\lambda^* = \bar{\lambda}^*$, and the proof is complete. □

From Lemmas 13–19, we have the following.

Theorem 20. *If* $\{(x^k, \lambda^k)\}_K \to (x^*, \lambda^*)$*, then* (x^*, λ^*) *is a KKT pair of problem (1a), (1b), (1c), and (1d).*

4. Rate of Convergence

In this section, we will establish the superlinear convergence of Algorithm 3. We suppose that the algorithm generates an infinite iterative sequence $\{x^k\}$ and there exists k_0 such that $I(n_{k-1}, k, \varepsilon_k) \ne \emptyset$ with $k > k_0$. That is, (S2.1)–(S2.4) will never be run when $k > k_0$ and the inner iterations (S2.5)–(S2.8) terminate finitely. Let $\tau^* = (x^*, \lambda^*)$ be an accumulation point of the sequence $\{(x^k, \lambda^k)\}$ generated by Algorithm 3. We assume that $\nabla^2 f$, $\nabla^2 c_i$, $i \in I$ are locally Lipschitz continuous on a neighborhood of x^*. To ensure the whole sequence $\{(x^k, \lambda^k)\}$ converges to (x^*, λ^*), we need the following assumption.

Assumption 21. The second-order sufficient condition holds at τ^*; that is, the Hessian $\nabla_{xx} L(x^*, \lambda^*)$ is positive definite on the space $\{\alpha \mid \langle \nabla c_i(x^*), \alpha \rangle = 0, \forall i \in I_0(x^*)\}$.

We first introduce a useful proposition as follows.

Proposition 22 (see [25, Proposition 4.1]). *Assume that* $\omega^* \in R^t$ *is an isolated accumulation point of a sequence* $\{\omega^k\} \subset R^t$ *such that for every subsequence* $\{\omega^k\}_K$ *converges to* ω^**; there is an infinite subset* $\bar{K} \subset K$ *such that* $\{\|\omega^{k+1} - \omega^k\|\}_{\bar{K}} \to 0$*; then the whole sequence* $\{\omega^k\}$ *converges to* ω^*.

Lemma 23. *If* $\{x^k\}_K \to x^*$*, then* $\{d^{k_0}\}_K \to 0$.

Proof. Assume to the contrary that there exists subset $\bar{K} \subset K$ such that $\{d^{k_0}\}_{\bar{K}} \to \bar{d} \ne 0$. By the finiteness of set I and boundedness of sequence $\{\lambda^k\}$, there exists subset $K_0 \subset \bar{K}$

such that $\{\lambda^k\}_{K_0} \to \lambda^*$ and $I(n_{k-1}, k, \varepsilon_k)$, $k \in K_0$ keep changeless. It is not difficult to see from linear system (18) that

$$\rho(x^*, \lambda^*) = \left\| H^* \overline{d} \right\| \neq 0. \tag{73}$$

On the other hand, from Theorem 20, (x^*, λ^*) is a KKT pair of problem (1a), (1b), (1c), and (1d); it follows that $\rho(x^*, \lambda^*) = 0$, which contradicts with (73). So we have that $\{d^{k_0}\}_K \to 0$. □

Lemma 24. *If $\{x^k\}_K \to x^*$, then $\{\lambda^k\}_K \to \lambda^*$.*

Proof. Since multiplier λ^* is unique with respect to x^* and $\{\lambda^k\}$ is bounded, it follows from Theorem 20 that $\{\lambda^k\}_K \to \lambda^*$. □

Lemma 25. *If $\{(x^k, \lambda^k)\}_K \to (x^*, \lambda^*)$, then $\{(x^{k-1}, \lambda^{k-1})\}_K \to (x^*, \lambda^*)$.*

Proof. Suppose that $(\overline{x}^*, \overline{\lambda}^*)$ is a arbitrary accumulation point of $\{(x^{k-1}, \lambda^{k-1})\}_K$. Then, from Theorem 20

$$\left\{\psi_{n_{k-1}}\left(x^{k-1}, \lambda^{k-1}\right)\right\}_{K_0} \longrightarrow \rho\left(\overline{x}^*, \overline{\lambda}^*\right) = 0, \quad K_0 \subset K. \tag{74}$$

In a way similar to the proof of Lemmas 18 and 19, we get that $x^* = \overline{x}^*$, $\lambda^* = \overline{\lambda}^*$. From the boundedness of $\{(x^{k-1}, \lambda^{k-1})\}_K$, the result follows. □

Lemma 26. *If $\{x^k\}_K \to x^*$, then $\{d^{k_j}\}_K \to 0$, $j = 0, 1, 2, 3$.*

Proof. By Lemma 25, $I(n_{k-1}, k, \varepsilon_k) \subset I_0(x^*)$. From Lemma 23, the result follows. □

Lemma 27. *Under Assumptions 1, 2, 4, and 21, the whole sequence $\{(x^k, \lambda^k)\}$ converges to (x^*, λ^*).*

Proof. Suppose that $\{x^k\}_K \to x^*$. Assumptions 2 and 21 imply that x^* is an isolated accumulation point of $\{x^k\}$ [30]. By (S.3) in Algorithm 3, $\|d^{k_3} - d^{k_2}\| \leq \|d^{k_2}\|$. It follows from Lemma 23 that

$$\left\| x^{k+1} - x^k \right\| \leq \left\| d^{k_2} \right\| + \left\| d^{k_3} - d^{k_2} \right\| \leq 2 \left\| d^{k_2} \right\| \longrightarrow 0. \tag{75}$$

Therefore, we have from Proposition 22 that the whole sequence $\{x^k\}$ converges to x^*. By Lemma 24, we have that λ^k converges to λ^*. This completes the proof. □

Assumption 28. The strict complementarity condition holds at τ^*; that is, $\lambda^* - c(x^*) \neq 0$.

Lemma 29. *Let $I_0^+(x^*) = \{i \in I \mid \lambda_i^* > 0\}$; then for all sufficiently big k*

$$I_0^+\left(x^*\right) = I\left(n_{k-1} + l_k, k, \varepsilon_k\right) = I_0\left(x^*\right). \tag{76}$$

Proof. By Theorem 20 and Lemma 27, it is easy to see that

$$\left\{\psi_{n_{k-1}}\left(x^{k-1}, \lambda^{k_1-1}\right)\right\} \longrightarrow \rho\left(\overline{x}^*, \overline{\lambda}^*\right) = 0. \tag{77}$$

In a way similar to the proof of (70) in Theorem 20, we have the following result:

$$I_0^+\left(x^*\right) \subset I\left(n_{k-1} + l_k, k, \varepsilon_k\right) \subset I_0\left(x^*\right). \tag{78}$$

By Assumption 28, the result follows. □

By Lemmas 23, 27, and 29, we can directly obtain the following corollary.

Corollary 30. *If Assumptions 1, 2, 4, and 21 hold, then for every $i = 0, 1, 2, 3$*

$$\begin{gathered} d^{k_i} \longrightarrow 0, \\ \lambda^{k_i} \longrightarrow \lambda^* \\ \text{as } k \longrightarrow \infty. \end{gathered} \tag{79}$$

By linear systems (18), (19), and (21), we have

$$\begin{aligned} M\left(n_{k-1}, k\right)\begin{pmatrix} d^{k_2} - d^{k_1} \\ \lambda^{k_2} - \lambda^{k_1} \end{pmatrix} &= \begin{pmatrix} 0 \\ \rho^k e \end{pmatrix}, \\ M\left(n_{k-1}, k\right)\begin{pmatrix} d^{k_3} - d^{k_2} \\ \lambda^{k_3} - \lambda^{k_2} \end{pmatrix} &= \begin{pmatrix} 0 \\ \omega^k - \mu^k \end{pmatrix}. \end{aligned} \tag{80}$$

Combining with the fact that $\rho_k = o(\|d^{k_1}\|^2)$ and $\omega^k - \mu^k = O(\|d^{k_2}\|^2)$, we have the following.

Lemma 31. *For sufficiently large k, the following results hold.*

$$\begin{aligned} \left\| d^{k_2} - d^{k_1} \right\| &= o\left(\left\| d^{k_1} \right\|^2\right), \\ \left\| \lambda^{k_2} - \lambda^{k_1} \right\| &= o\left(\left\| d^{k_1} \right\|^2\right), \\ \left\| d^{k_3} - d^{k_2} \right\| &= O\left(\left\| d^{k_2} \right\|^2\right), \\ \left\| \lambda^{k_3} - \lambda^{k_2} \right\| &= O\left(\left\| d^{k_2} \right\|^2\right). \end{aligned} \tag{81}$$

Assumption 32. The sequence of matrices $\{H_k\}$ satisfies

$$\frac{\left\| P_k\left(H_k - \nabla_{xx}^2 L\left(x^*, \lambda^*\right)\right) d^{k_2} \right\|}{\left\| d^{k_2} \right\|} \longrightarrow 0, \tag{82}$$

where $P_k = E - N_k(N_k^T N_k)^{-1} N_k^T$, $N_k = \nabla c_{I_0(x^*)}(x^k)$.

Note. Assumption 32 is an extended Dennis-Moré condition. It is used in Qp-free algorithm for nonlinear optimization problems by Yang et al. [27]. We will show that it is a sufficient condition for our algorithm to be superlinearly convergent. In order to show the superlinear convergence, we first introduce the following proposition.

Proposition 33 (see [27, Lemma 4.3]). *For sufficiently large k, the direction d^{k_2} can be decomposed into*

$$d^{k_2} = P_k d^{k_2} + \widetilde{d}^{k_2}, \tag{83}$$

with $\|\widetilde{d}^{k_2}\| = O(\|c_{I_0(x^*)}(x^k)\|) + o(\|d^{k_1}\|^2)$.

Lemma 34. *For sufficiently large k, if $I(n_{k-1}, k, \varepsilon_k) \neq \emptyset$, then the step $t_k = 1$ is accepted.*

Proof. For $i \notin I_0(x^*)$, due to $c_i(x^*) < 0$, it is not difficult to see from Corollary 30 that when k is sufficiently large, $c_i(x^k + d^{k_3}) < 0$. For $i \in I_0(x^*)$, we have from linear system (21) and Lemma 31 that

$$\begin{aligned} c_i\left(x^k + d^{k_3}\right) &= c_i\left(x^k + d^{k_2}\right) \\ &\quad + \nabla c_i\left(x^k + d^{k_2}\right)^T\left(d^{k_3} - d^{k_2}\right) \\ &\quad + O\left(\left\|d^{k_3} - d^{k_2}\right\|^2\right) \\ &= c_i\left(x^k + d^{k_2}\right) \\ &\quad + \nabla c_i\left(x^k + d^{k_2}\right)^T\left(d^{k_3} - d^{k_2}\right) \\ &\quad + O\left(\left\|d^{k_2}\right\|^3\right) \\ &= -\left\|d^{k_2}\right\|^\eta + O\left(\left\|d^{k_2}\right\|^3\right) = o\left(\left\|d^{k_2}\right\|^2\right). \end{aligned} \tag{84}$$

It follows from $\eta \in (2, 3)$ that, for sufficiently large k, $c_i(x^k + d^{k_3}) < 0$. So when k is sufficiently large, $x^k + d^{k_3}$ is a strictly feasible point of problem (1a), (1b), (1c), and (1d). By (21) and (81), we have

$$\begin{aligned} c_i\left(x^k + d^{k_3}\right) &= c_i\left(x^k\right) + \nabla c_i\left(x^k\right)^T d^{k_3} \\ &\quad + \frac{1}{2}{d^{k_3}}^T\nabla^2 c_i\left(x^k\right)d^{k_3} + o\left(\left\|d^{k_2}\right\|^2\right) \\ &= \nabla c_i\left(x^k\right)^T\left(d^{k_3} - d^{k_2}\right) \\ &\quad + \nabla c_i\left(x^k\right)^T\left(d^{k_2} - d^{k_1}\right) \\ &\quad + \frac{1}{2}{d^{k_3}}^T\nabla^2 c_i\left(x^k\right)d^{k_3} + o\left(\left\|d^{k_2}\right\|^2\right) \\ &= \nabla c_i\left(x^k\right)^T\left(d^{k_3} - d^{k_2}\right) \\ &\quad + \frac{1}{2}{d^{k_2}}^T\nabla^2 c_i\left(x^k\right)d^{k_2} + o\left(\left\|d^{k_2}\right\|^2\right). \end{aligned} \tag{85}$$

Combining with (84), we have

$$\begin{aligned} &\nabla c_i\left(x^k\right)^T\left(d^{k_3} - d^{k_2}\right) + \frac{1}{2}{d^{k_2}}^T\nabla^2 c_i\left(x^k\right)d^{k_2} \\ &\quad = o\left(\left\|d^{k_2}\right\|^2\right). \end{aligned} \tag{86}$$

It follows that for sufficiently large k

$$\begin{aligned} &\nabla f\left(x^k\right)^T\left(d^{k_3} - d^{k_2}\right) \\ &\quad = \left[\nabla f\left(x^k\right) - g_{\bar{n}_k}\left(x^k\right)\right]^T\left(d^{k_3} - d^{k_2}\right) \\ &\quad\quad + g_{\bar{n}_k}\left(x^k\right)^T\left(d^{k_3} - d^{k_2}\right) \\ &\quad = -{d^{k_2}}^T H_k\left(d^{k_3} - d^{k_2}\right) \\ &\quad\quad - \sum_{i\in I_0(x^*)}\lambda_i^{k_2}\nabla c_i\left(x^k\right)^T\left(d^{k_3} - d^{k_2}\right) + o\left(\left\|d^{k_2}\right\|^2\right) \\ &\quad = \frac{1}{2}\sum_{i\in I_0(x^*)}\lambda_i^{k_2}{d^{k_2}}^T\nabla^2 c_i\left(x^k\right)d^{k_2} + o\left(\left\|d^{k_2}\right\|^2\right). \end{aligned} \tag{87}$$

From Proposition 33 and Assumption 32, we have that

$$\begin{aligned} &\frac{1}{2}{d^{k_2}}^T\left[H_k - \left(\nabla^2 f\left(x^k\right) + \sum_{i\in I_0(x^*)}\lambda_i^{k_2}\nabla^2 c_i\left(x^k\right)\right)\right]d^{k_2} = \frac{1}{2} \\ &\quad\cdot {d^{k_2}}^T\left(H_k - \nabla_{xx}^2 L\left(x^k, \lambda^{k_2}\right)\right)d^{k_2} = \frac{1}{2} \\ &\quad\cdot {d^{k_2}}^T P_k\left(H_k - \nabla_{xx}^2 L\left(x^k, \lambda^{k_2}\right)\right)d^{k_2} + \frac{1}{2} \\ &\quad\cdot {d^{k_2}}^T\left(H_k - \nabla_{xx}^2 L\left(x^k, \lambda^{k_2}\right)\right)\tilde{d}^{k_2} = \frac{1}{2} \\ &\quad\cdot {d^{k_2}}^T P_k\left(H_k - \nabla_{xx}^2 L\left(x^*, \lambda^*\right)\right)d^{k_2} \\ &\quad + o\left(\left\|c_{I_0(x^*)}\left(x^k\right)\right\|\right) + o\left(\left\|d^{k_2}\right\|^2\right) \\ &\quad = o\left(\left\|c_{I_0(x^*)}\left(x^k\right)\right\|\right) + o\left(\left\|d^{k_2}\right\|^2\right). \end{aligned} \tag{88}$$

From linear system (19), for sufficiently large k,

$$\begin{aligned} &\nabla f\left(x^k\right)^T d^{k_2} \\ &\quad = \left(f\left(x^k\right)^T d^{k_2} - g_{\bar{n}_k}\left(x^k\right)^T d^{k_2}\right) + g_{\bar{n}_k}\left(x^k\right)^T d^{k_2} \\ &\quad \leq g_{\bar{n}_k}\left(x^k\right)^T d^{k_2} + \frac{2C\left(\log \bar{n}_k\right)^{m-1}}{\bar{n}_k} \\ &\quad \leq \frac{1}{2}g_{\bar{n}_k}\left(x^k\right)^T d^{k_2} \\ &\quad\quad + \frac{1}{2}\left[-{d^{k_2}}^T H_k d^{k_2} + \sum_{i\in I_0(x^*)}\lambda_i^{k_2}c_i\left(x^k\right)\right] + \frac{1}{2}\alpha_k \\ &\quad\quad + o\left(\left\|d^{k_2}\right\|^2\right). \end{aligned} \tag{89}$$

Since $\lambda_i^{k_2} \to \lambda_i^* > 0$ for all $i \in I_0(x^*)$, we have, for sufficiently large k,

$$\frac{1}{2}\sum_{i\in I_0(x^*)}\lambda_i^{k_2}c_i\left(x^k\right) + o\left(\left\|c_{I_0(x^*)}\left(x^k\right)\right\|\right) < 0. \tag{90}$$

By (87), (88), and (89), we get

$$
\begin{aligned}
& f_{\bar{n}_k}\left(x^k+d^{k_3}\right)-f_{\bar{n}_k}\left(x^k\right) \\
& \quad=\left(f_{\bar{n}_k}\left(x^k+d^{k_3}\right)-f\left(x^k+d^{k_3}\right)\right) \\
& \qquad+\left(f\left(x^k\right)-f_{\bar{n}_k}\left(x^k\right)\right) \\
& \qquad+\left(f\left(x^k+d^{k_3}\right)-f\left(x^k\right)\right) \\
& \quad\leq \nabla f\left(x^k\right)^T d^{k_3}+\frac{1}{2} d^{k_2 T} \nabla^2 f\left(x^k\right) d^{k_2}+o\left(\left\|d^{k_2}\right\|^2\right) \\
& \qquad+\frac{2C\left(\log \bar{n}_k\right)^{m-1}}{\bar{n}_k} \\
& \quad\leq \nabla f\left(x^k\right)^T d^{k_2}+\nabla f\left(x^k\right)^T\left(d^{k_3}-d^{k_2}\right) \\
& \qquad+\frac{1}{2} d^{k_2 T} \nabla^2 f\left(x^k\right) d^{k_2}+o\left(\left\|d^{k_2}\right\|^2\right)+\frac{1}{2} \alpha_k \\
& \quad=\nabla f\left(x^k\right)^T d^{k_2}+\frac{1}{2} \sum_{i \in I_0(x^*)} \lambda_i^{k_2} d^{k_2 T} \nabla^2 c_i\left(x^k\right) d^{k_2} \\
& \qquad+\frac{1}{2} d^{k_2 T} \nabla^2 f\left(x^k\right) d^{k_2}+o\left(\left\|d^{k_2}\right\|^2\right)+\frac{1}{2} \alpha_k \\
& \quad\leq \frac{1}{2} g_{\bar{n}_k}\left(x^k\right)^T d^{k_2} \\
& \qquad+\left[\frac{1}{2} \sum_{i \in I_0(x^*)} \lambda_i^{k_2} c_i\left(x^k\right)+o\left(\left\|c_{I_0(x^*)}\left(x^k\right)\right\|\right)\right] \\
& \qquad+o\left(\left\|d^{k_2}\right\|^2\right)+\alpha_k \\
& \quad\leq u g_{\bar{n}_k}\left(x^k\right)^T d^{k_2} \\
& \qquad+\left(u-\frac{1}{2}\right)\left[d^{k_2 T} H_k d^{k_2}-\sum_{i \in I_0(x^*)} \lambda_i^{k_2} c_i\left(x^k\right)\right] \\
& \qquad+o\left(\left\|d^{k_2}\right\|^2\right)+\alpha_k \\
& \quad\leq u g_{\bar{n}_k}\left(x^k\right)^T d^{k_2} \\
& \qquad+\left(u-\frac{1}{2}\right)\left[d^{k_2 T} H_k d^{k_2}+o\left(\left\|d^{k_2}\right\|^2\right)\right]+\alpha_k \\
& \quad\leq u g_{\bar{n}_k}\left(x^k\right)^T d^{k_2}+\alpha_k,
\end{aligned}
\tag{91}
$$

which completes the proof. □

Theorem 35. *Under stated assumptions, we have*

$$\left\|x^{k+1}-x^*\right\|=o\left(\left\|x^k-x^*\right\|\right). \tag{92}$$

Proof. By the definition of P_k, we have

$$
\begin{aligned}
P_k H_k d^{k_3} & =-P_k g_{\bar{n}_k}\left(x^k\right)=-P_k\left(\nabla f\left(x^k\right)-\nabla f\left(x^*\right)\right. \\
& \quad\left.+\lambda^*\left(\nabla c_{I_0(x^*)}\left(x^k\right)-\nabla c_{I_0(x^*)}\left(x^*\right)\right)\right) \\
& \quad-P_k\left(g_{\bar{n}_k}\left(x^k\right)-\nabla f\left(x^k\right)\right)=-P_k \nabla_{xx}^2 L\left(x^*, \lambda^*\right)\left(x^k\right. \\
& \quad\left.-x^*\right)+o\left(\left\|x^k-x^*\right\|\right)+o\left(\left\|d^{k_3}\right\|\right).
\end{aligned}
\tag{93}
$$

It follows from (93) that

$$
\begin{aligned}
& P_k \nabla_{xx}^2 L\left(x^*, \lambda^*\right)\left(x^k+d^{k_3}-x^*\right) \\
& \quad=-P_k\left(H_k-\nabla_{xx}^2 L\left(x^*, \lambda^*\right)\right) d^{k_3}+o\left(\left\|x^k-x^*\right\|\right) \\
& \qquad+o\left(\left\|d^{k_3}\right\|\right).
\end{aligned}
\tag{94}
$$

Since $c_{I_0(x^*)}(x^*)=0$, it is clear from linear system (21) that

$$
\begin{aligned}
\nabla c_{I_0(x^*)}\left(x^k\right)^T d^{k_3} & =-c_{I_0(x^*)}\left(x^k\right)+o\left(\left\|d^{k_3}\right\|\right) \\
& =-\nabla c_{I_0(x^*)}\left(x^k\right)^T\left(x^k-x^*\right) \\
& \quad+o\left(\left\|x^k-x^*\right\|\right)+o\left(\left\|d^{k_3}\right\|\right).
\end{aligned}
\tag{95}
$$

Let $G_k := \begin{pmatrix} P_k \nabla_{xx}^2 L(x^*,\lambda^*) \\ \nabla c_{I_0(x^*)}(x^k) \end{pmatrix}$ and $B_k := \begin{pmatrix} -P_k(H_k-\nabla_{xx}^2 L(x^*,\lambda^*))d^{k_3} \\ 0 \end{pmatrix}$.

From (94) and (95), we have

$$
\begin{aligned}
G_k\left(x^k+d^{k_3}-x^*\right) & =B_k+o\left(\left\|x^k-x^*\right\|\right) \\
& \quad+o\left(\left\|d^{k_3}\right\|\right).
\end{aligned}
\tag{96}
$$

From Assumption 21, it is not difficult to see that when k is sufficiently large, G_k have full column rank. It follows from (96) and Assumption 32 that

$$\left\|x^k+d^{k_3}-x^*\right\|=o\left(\left\|x^k-x^*\right\|\right)+o\left(\left\|d^{k_3}\right\|\right), \tag{97}$$

which implies that

$$\left\|x^k+d^{k_3}-x^*\right\|=o\left(\left\|x^k-x^*\right\|\right). \tag{98}$$

This completes the proof. □

In sequel, we consider the following case: the KKT point x^* of problem (1a), (1b), (1c), and (1d) is an unconstrained stationary with multiplier vector $\lambda^* = 0$. It is clear that $\nabla f(x^*) = 0$ and also $I_0(x^*) = \emptyset$ in this case. Therefore, we have form the construction of Algorithm 3 that, for sufficiently large k, $I(n_{k-1}, k, \varepsilon_k) = \emptyset$. In order to show the superlinear convergence under this case. we firstly give two well-known propositions.

Proposition 36. *Assume that $f(x)$ is twice continuously differentiable and $\nabla^2 f(x)$ is Lipschitz continuous on open convex subset D of $\mathcal{F}$. Then, for arbitrary $x, u, v \in D$, we have*

$$
\begin{aligned}
& \left\|\nabla f(u)-\nabla f(v)-\nabla^2 f(x)(u-v)\right\| \\
& \quad\leq \frac{\gamma}{2}(\|u-x\|+\|v-x\|)\|u-v\|,
\end{aligned}
\tag{99}
$$

where γ is a Lipschitz constant.

Proposition 37. *Assume that $f(x)$ and $\nabla^2 f(x)$ satisfy the conditions in Proposition 36. If $\nabla^2 f(x)$ is symmetric positive definite, then there exist $\varepsilon > 0$, $\beta > \alpha > 0$ such that when $\max\{\|u-x\|, \|v-x\|\} \le \varepsilon$ with $u, v \in D$*

$$\alpha \|u-x\| \le \|\nabla f(u) - \nabla f(v)\| \le \|u-v\|. \tag{100}$$

In order to obtain the superlinear convergence of problem (1a), (1b), (1c), and (1d) under the condition $I_0(x^*) = \emptyset$, we give the following assumption.

Assumption 38. The sequence of matrices $\{H_k\}$ satisfies

$$\frac{\left\|\left(H_k - \nabla^2_{xx} L(x^*, \lambda^*)\right) d^{k_2}\right\|}{\left\|d^{k_2}\right\|} \longrightarrow 0. \tag{101}$$

Lemma 39. *If $I_0(x^*) = \emptyset$, then the step $t_k = 1$ is accepted for sufficiently large k.*

Proof. Since $I_0(x^*) = \emptyset$, $c_i(x^*) \ne 0$, $\forall i \in I$. It follows from $d^{k_2} \to 0$ that, for sufficiently large k, $c_i(x^k + d^{k_2}) \ne 0$, $\forall i \in I$. That is, $x^k + d^{k_2}$ is strictly feasible. To the end, we show that inequality (23) also holds when $t_k = 1$. By (5), (6), and $g_{\bar{n}_k}(x^K) = -H_k d^{k_2}$, we have, for sufficiently large k,

$$\begin{aligned}
& f_{\bar{n}_k}\left(x^k + d^{k_2}\right) - f_{\bar{n}_k}\left(x^k\right) \\
&\quad = \left(f\left(x^k + d^{k_2}\right) - f\left(x^k\right)\right) \\
&\qquad + \left(f_{\bar{n}_k}\left(x^k + d^{k_2}\right) - f\left(x^k + d^{k_2}\right)\right) \\
&\qquad + \left(f\left(x^k\right) - f_{\bar{n}_k}\left(x^k\right)\right) \\
&\quad \le \nabla f\left(x^k\right)^T d^{k_2} + \frac{1}{2} {d^{k_2}}^T \nabla^2 f\left(x^k\right) d^{k_2} + \frac{1}{2}\alpha_k \\
&\qquad + o\left(\left\|d^{k_2}\right\|\right) \\
&\quad = g_{\bar{n}_k}\left(x^k\right)^T d^{k_2} + \left(\nabla f\left(x^k\right) - g_{\bar{n}_k}\left(x^k\right)\right)^T d^{k_2} \\
&\qquad + \frac{1}{2} {d^{k_2}}^T \nabla^2 f\left(x^k\right) d^{k_2} + \frac{1}{2}\alpha_k + o\left(\left\|d^{k_2}\right\|\right) \\
&\quad \le -\frac{1}{2} {d^{k_2}}^T H_k d^{k_2} + \frac{1}{2} {d^{k_2}}^T \left(\nabla^2 f\left(x^k\right) - H_k\right) d^{k_2} \\
&\qquad + \alpha_k + o\left(\left\|d^{k_2}\right\|\right) \\
&\quad \le -\frac{1}{2} {d^{k_2}}^T H_k d^{k_2} + \alpha_k + o\left(\left\|d^{k_2}\right\|\right) \\
&\quad \le -u {d^{k_2}}^T H_k d^{k_2} + \alpha_k = u g_{\bar{n}_k}\left(x^k\right)^T d^{k_2} + \alpha_k,
\end{aligned} \tag{102}$$

which completes the proof. □

Theorem 40. *Assume that $I_0(x^*) = \emptyset$. If $f(x)$ and $\nabla^2 f(x)$ satisfy the conditions in Propositions 36 and 37, then*

$$\left\|x^{k+1} - x^*\right\| = o\left(\left\|x^k - x^*\right\|\right). \tag{103}$$

Proof. Since $g_{\bar{n}_k}(x^k) = -H_k d^{k_2}$, we have

$$\begin{aligned}
& \left[H_k - \nabla^2 f\left(x^*\right)\right]\left(x^{k+1} - x^k\right) \\
&\quad = -g_{\bar{n}_k}\left(x^k\right) - \nabla^2 f\left(x^*\right)\left(x^{k+1} - x^k\right) \\
&\quad = \left[\nabla f\left(x^k\right) - g_{\bar{n}_k}\left(x^k\right)\right] \\
&\qquad + \left[\nabla f\left(x^{k+1}\right) - \nabla f\left(x^k\right) - \nabla^2 f\left(x^*\right)\left(x^{k+1} - x^k\right)\right] \\
&\qquad - \nabla f\left(x^{k+1}\right).
\end{aligned} \tag{104}$$

It follows from Proposition 36 that

$$\begin{aligned}
\frac{\left\|\nabla f\left(x^{k+1}\right)\right\|}{\left\|x^{k+1} - x^k\right\|} & \le \frac{\left\|\left[H_k - \nabla^2 f\left(x^*\right)\right]\left(x^{k+1} - x^k\right)\right\|}{\left\|x^{k+1} - x^k\right\|} \\
&\quad + \frac{\left\|\nabla f\left(x^{k+1}\right) - \nabla f\left(x^k\right) - \nabla^2 f\left(x^*\right)\left(x^{k+1} - x^*\right)\right\|}{\left\|x^{k+1} - x^k\right\|} \\
&\quad + \frac{\left\|\nabla f\left(x^k\right) - g_{\bar{n}_k}\left(x^k\right)\right\|}{\left\|x^{k+1} - x^k\right\|} \\
&\le \frac{\left\|\left[H_k - \nabla^2 f\left(x^*\right)\right]\left(x^{k+1} - x^k\right)\right\|}{\left\|x^{k+1} - x^k\right\|} \\
&\quad + \frac{\left\|\nabla f\left(x^k\right) - g_{\bar{n}_k}\left(x^k\right)\right\|}{\left\|x^{k+1} - x^k\right\|} \\
&\quad + \frac{1}{2}\left(\left\|x^k - x^*\right\| + \left\|x^{k+1} - x^*\right\|\right).
\end{aligned} \tag{105}$$

By inequality (6), we have

$$\begin{aligned}
\frac{\left\|\nabla f\left(x^k\right) - g_{\bar{n}_k}\left(x^k\right)\right\|}{\left\|x^{k+1} - x^k\right\|} & \le \frac{1}{\left\|d^{k_2}\right\|} \cdot \frac{C\left(\log \bar{n}_k\right)^{m-1}}{\bar{n}_k^{\delta_1}} \cdot \frac{1}{\bar{n}_k^{\delta_2}} \\
& \le M \cdot \frac{C\left(\log \bar{n}_k\right)^{m-1}}{\bar{n}_k^{\delta_1}} \longrightarrow 0.
\end{aligned} \tag{106}$$

Hence, from (105) and (106), we have

$$\frac{\left\|\nabla f\left(x^{k+1}\right)\right\|}{\left\|d^{k_2}\right\|} \longrightarrow 0 \quad \text{as } k \longrightarrow \infty. \tag{107}$$

By $\nabla f(x^*) = 0$ and Proposition 37,

$$\begin{aligned}
\left\|\nabla f\left(x^{k+1}\right)\right\| & = \left\|\nabla f\left(x^{k+1}\right) - \nabla f\left(x^*\right)\right\| \\
& \ge \beta \left\|x^{k+1} - x^*\right\| \longrightarrow 0.
\end{aligned} \tag{108}$$

So, we have

$$\frac{\left\|\nabla f\left(x^{k+1}\right)\right\|}{\left\|d^{k_2}\right\|} \ge \frac{\beta\left\|x^{k+1} - x^*\right\|}{\left\|x^{k+1} - x^*\right\| + \left\|x^k - x^*\right\|} \longrightarrow 0, \tag{109}$$

which implies that

$$\frac{\left\|x^{k+1} - x^*\right\|}{\left\|x^k - x^*\right\|} \longrightarrow 0. \tag{110}$$

This completes the proof. □

5. Conclusion

In this paper, by quasi-Monte-Carlo-based approximations of the objective function and its first derivative, we have proposed a feasible sequential system of linear equations method for two-stage stochastic quadratic programming problem with inequality constraint. A new technique to update the "working set" is suggested. The feature of the new technique is that, in order to update the "working set," at each iteration we directly make use of the solution λ^{k_0} of linear system (16), while we do not calculate the inverse of matrix $M^{-1}(x)$ [27]. Moreover, it also does not need to approximate the Hessian by Monte Carlo (or quasi-Monte-Carlo) rule. Therefore, our algorithm saves the computational cost. The other remarkable feature of this technique is that it can accurately identify active constraints of problem (1a), (1b), (1c), and (1d). It should be pointed out that the technique also is useful for deterministic nonlinear programming problem with inequality constraints. We have shown that the sequence generated by the proposed algorithm converges to a KKT point of the problem globally. In particular the convergence rate is locally superlinear under some additional conditions. To get the superlinear convergence of the algorithm, we still need the strict complementarity assumption. However, we believe that, by using quasi-Monte-Carlo-based approximations and the new identification technique, it is possible to find a new algorithm without strict complementarity assumption. Moreover, how to use parallel optimization techniques [31–33] for the large scale stochastic programs with recourse is an important topic for further research.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Bureau of Statistics of the People's Republic of China (2014LZ41).

References

[1] J. R. Birge and F. Louveaux, *Introduction to Stochastic Programming*, Springer, New York, NY, USA, 1997.

[2] S. W. Wallace and W. T. Ziemba, "Applications of stochastic programming," in *Proceedings of the MPS/SIAM Series on Optimization*, vol. 5, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2005.

[3] X. Liu, Y. Li, and W. Zhang, "Stochastic linear quadratic optimal control with constraint for discrete-time systems," *Applied Mathematics and Computation*, vol. 228, pp. 264–270, 2014.

[4] G. Li and W. Zhang, "Study on indefinite stochastic linear quadratic optimal control with inequality constraint," *Journal of Applied Mathematics*, vol. 2013, Article ID 805829, 5004 pages, 2013.

[5] L. Liu and X. Meng, "Optimal harvesting control and dynamics of two-species stochastic model with delays," *Advances in Difference Equations*, vol. 2017, 18 pages, 2017.

[6] G. Li and M. Chen, "Infinite horizon linear quadratic optimal control for stochastic difference time-delay systems," *Advances in Difference Equations*, vol. 2015, article 14, 2017.

[7] H. Yang, "Study on stochastic linear quadratic optimal control with quadratic and mixed terminal state constraints," *Journal of Applied Mathematics*, vol. 2013, Article ID 674327, 11 pages, 2013.

[8] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, "Robust stochastic approximation approach to stochastic programming," *SIAM Journal on Optimization*, vol. 19, no. 4, pp. 1574–1609, 2009.

[9] J. Linderoth, A. Shapiro, and S. Wright, "The empirical behavior of sampling methods for stochastic programming," *Annals of Operations Research*, vol. 142, no. 1, pp. 215–241, 2006.

[10] J. Yu, M. Li, Y. Wang, and G. He, "A decomposition method for large-scale box constrained optimization," *Applied Mathematics and Computation*, vol. 231, no. 12, pp. 9–15, 2014.

[11] G. Bayraksan and D. P. Morton, "A sequential sampling procedure for stochastic programming," *Operations Research*, vol. 59, no. 4, pp. 898–913, 2011.

[12] T. Ekin, N. G. Polson, and R. Soyer, "Augmented Markov chain MONte Carlo simulation for two-stage stochastic programs with recourse," *Decision Analysis*, vol. 11, no. 4, pp. 250–264, 2014.

[13] R. T. Rockafellar and R. J. Wets, "A dual solution procedure for quadratic stochastic programs with simple recourse," in *Numerical methods*, vol. 1005 of *Lecture Notes in Math*, pp. 252–265, Springer, Berlin, Germany, 1983.

[14] R. T. Rockafellar and R. J. Wets, "A Lagrangian finite generation technique for solving linear-quadratic problems in stochastic programming," *Mathematical Programming Study*, no. 28, pp. 63–93, 1986.

[15] L. Q. Qi and R. S. Womersley, "An SQP algorithm for extended linear-quadratic problems in stochastic programming," *Annals of Operations Research*, vol. 56, pp. 251–285, 1995.

[16] X. J. Chen, L. Q. Qi, and R. S. Womersley, "Newton's method for quadratic stochastic programs with recourse," *Journal of Computational and Applied Mathematics*, vol. 60, no. 1-2, pp. 29–46, 1995.

[17] J. Birge, X. Chen, and L. Qi, "A stochastic Newton method for stochastic quadratic programs with recourse," https://www.researchgate.net/publication/2255017.

[18] S. Joe and I. H. Sloan, "Imbedded lattice rules for multidimensional integration," *SIAM Journal on Numerical Analysis*, vol. 29, no. 4, pp. 1119–1135, 1992.

[19] S. Joe and I. H. Sloan, "Implementation of a Lattice Method for Numerical Multiple Integration," *ACM Transactions on Mathematical Software (TOMS)*, vol. 19, no. 4, pp. 523–545, 1993.

[20] H. Niederreiter, "Random number generation and quasi-Monte Carlo methods," *Journal of the American Statistical Association*, vol. 88, no. 89, pp. 147–153, 1992.

[21] J. R. Birge, "Quasi-Monte Carlo approaches to option pricing," *American Anthropologist*, vol. 81, no. 2, pp. 87–388, 1995.

[22] E. R. Panier, A. L. Tits, and J. N. Herskovits, "A QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization," *SIAM Journal on Control and Optimization*, vol. 26, no. 4, pp. 788–811, 2006.

[23] L. Sun, G. He, Y. Wang, and C. Zhou, "An accurate active set Newton algorithm for large scale bound constrained optimization," *Applications of Mathematics*, vol. 56, no. 3, pp. 297–314, 2011.

[24] Y. Li, T. Tan, and X. Li, “A log-exponential smoothing method for mathematical programs with complementarity constraints,” *Applied Mathematics and Computation*, vol. 218, no. 10, pp. 5900–5909, 2012.

[25] H.-D. Qi and L. Qi, “A new QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization,” *SIAM Journal on Optimization*, vol. 11, no. 1, pp. 113–132, 2000.

[26] Z. Y. Gao, G. P. He, and F. Wu, “Sequential systems of linear equations algorithm for nonlinear optimization problems with general constraints,” *Journal of Optimization Theory and Applications*, vol. 147, no. 3, pp. 211–226, 2002.

[27] Y.-F. Yang, D.-H. Li, and L. Qi, “A feasible sequential linear equation method for inequality constrained optimization,” *SIAM Journal on Optimization*, vol. 13, no. 4, pp. 1222–1244, 2003.

[28] F. Facchinei, A. Fischer, and C. Kanzow, “On the accurate identification of active constraints,” *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 14–32, 1996.

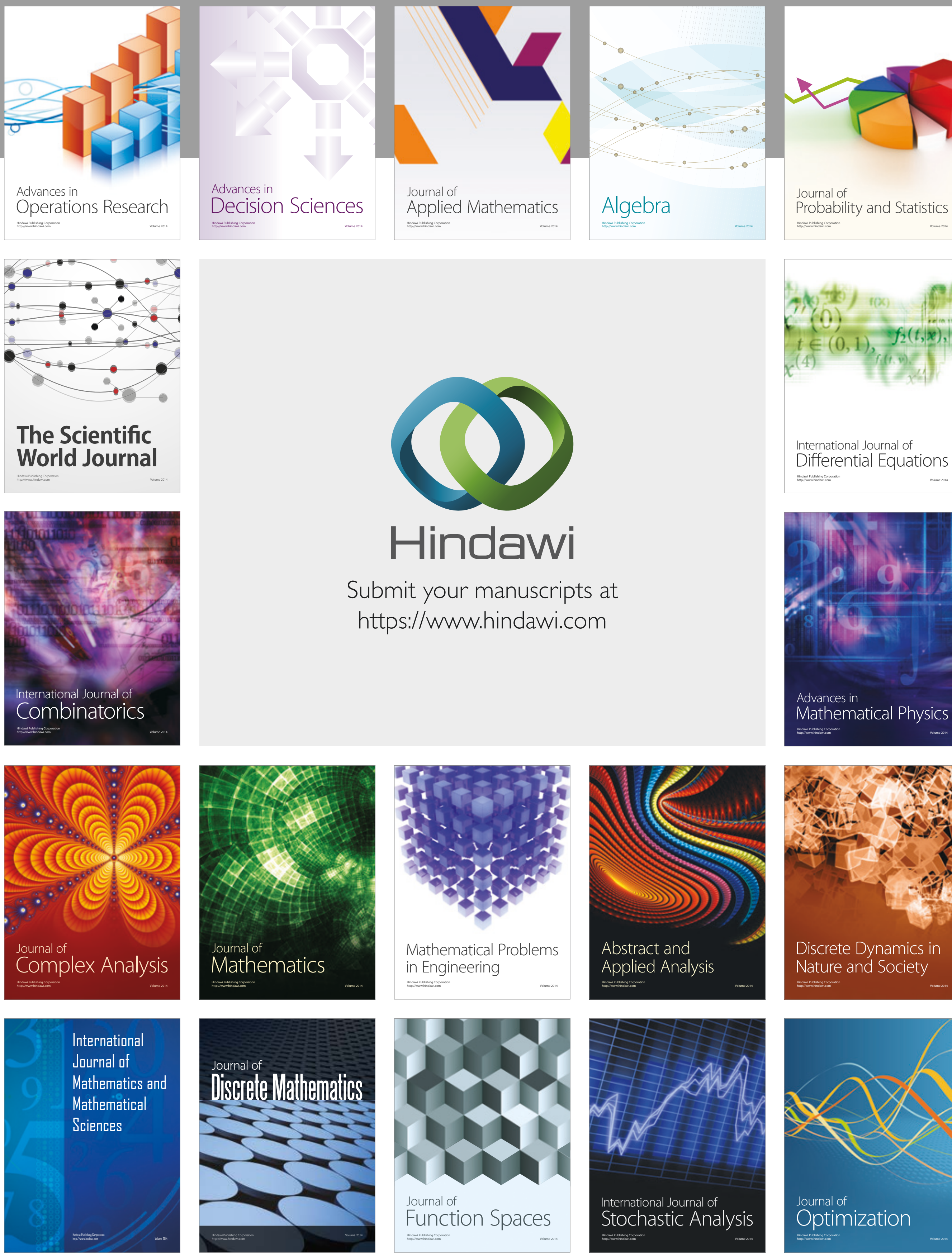
[29] M. J. D. Powell, “A fast algorithm for nonlinearly constrained optimization calculations, in Numerical Analysis,” *Lecture Notes in Mathematics*, vol. 630, pp. 144–157, 1978.

[30] S. M. Robinson, “Strongly regular generalized equations,” *Mathematics of Operations Research*, vol. 5, no. 1, pp. 43–62, 1980.

[31] C. Han, F. Zheng, T. Guo, and G. He, “Parallel algorithms for large-scale linearly constrained minimization problem,” *Acta Mathematicae Applicatae Sinica*, vol. 30, no. 3, pp. 707–720, 2014.

[32] C. Han, T. Feng, G. He, and T. Guo, “Parallel variable distribution algorithm for constrained optimization with nonmonotone technique,” *Journal of Applied Mathematics*, vol. 2013, Article ID 295147, 420 pages, 2013.

[33] F. Zheng, C. Han, and Y. Wang, “Parallel SSLE algorithm for large scale constrained optimization,” *Applied Mathematics and Computation*, vol. 217, no. 12, pp. 5377–5384, 2011.


Advances in Operations Research
Advances in Decision Sciences
Journal of Applied Mathematics
Algebra
Journal of Probability and Statistics
The Scientific World Journal
Hindawi
Submit your manuscripts at
https://www.hindawi.com
International Journal of Differential Equations
International Journal of Combinatorics
Advances in Mathematical Physics
Journal of Complex Analysis
Journal of Mathematics
Mathematical Problems in Engineering
Abstract and Applied Analysis
Discrete Dynamics in Nature and Society
International Journal of Mathematics and Mathematical Sciences
Journal of Discrete Mathematics
Journal of Function Spaces
International Journal of Stochastic Analysis
Journal of Optimization