Broyden-like Methods for Variational Inequality and Nonlinear Complementarity Problems with Global and Superlinear Convergence

Donghui Li [1]
Department of Applied Mathematics
Hunan University
Changsha, China 410082
e-mail: dhli@mail.hunu.edu.cn

Masao Fukushima
Department of Applied Mathematics and Physics
Graduate School of Informatics
Kyoto University
Kyoto 606-8501, Japan
e-mail: fuku@kuamp.kyoto-u.ac.jp

September 17, 1998

Abstract In this paper, we propose a Broyden-like quasi-Newton method for solving finite-dimensional variational inequality problems. By means of a perturbed Fischer-Burmeister function, we successively approximate the KKT system for a variational inequality problem by a sequence of smooth equations. Based on these approximating equations, we present a Broyden-like method. The method is well defined. We show that under suitable conditions, the proposed method converges to a KKT point of the variational inequality problem globally and superlinearly. In particular, we show that when specialized to a nonlinear complementarity problem, the proposed method converges globally whenever the problem involves a P_0 function with Lipschitzian Jacobian. Moreover, if the strict complementarity condition holds, then the convergence rate is superlinear.

Key words: Variational inequality problem, nonlinear complementarity problem, Broyden-like method, global convergence, superlinear convergence

[1] Present address (available until October, 1999): Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan, e-mail: lidh@kuamp.kyoto-u.ac.jp

1. Introduction

Many practical problems such as economic and traffic equilibrium problems can be formulated as the variational inequality problem of finding an $x \in X$ such that

$$F(x)^T(y-x) \geq 0, \qquad \forall y \in X, \tag{1.1}$$

where $F : R^n \to R^n$ is continuously differentiable and $X \subset R^n$ is a closed convex set. In most cases X can be specified as

$$X = \{x \in R^n \,|\, g_i(x) \geq 0, i = 1,2,\ldots,m_1, h_j(x) = 0, j = 1,2,\ldots m_2\},$$

where g_i, $i = 1,2,\ldots,m_1$, are twice continuously differentiable concave functions and h_j, $j = 1,2,\ldots,m_2$, are affine functions. Problem (1.1) is denoted by VI(X,F). When $X = R^n_+ = \{x \,|\, x \geq 0\}$, that is, $m_1 = n$, $m_2 = 0$ and $g_i(x) = x_i$, $i = 1,2,\ldots,n$, (1.1) is called the nonlinear complementarity problem, denoted by NCP(F), which is equivalent to finding an $x \in R^n$ such that

$$x \geq 0, \quad F(x) \geq 0 \quad \text{and} \quad x^T F(x) = 0. \tag{1.2}$$

An important task in the study of variational inequality and nonlinear complementarity problems is to construct practical and efficient iterative methods. There have been developed various kinds of iterative methods. For example, linearization methods successively approximate VI(X,F) by linearized variational inequalities VI(X,F^k), where $F^k(x) = F(x_k) + B_k(x - x_k)$ with some matrix B_k. Generally, linearization methods converge only locally unless line search is used. A comprehensive review of those methods can be found in [12]. In recent years, much effort has been made to enlarge the convergence domain of an iterative method by reformulating VI(X,F) as an optimization problem or a nonsmooth equation. Using appropriate reformulations, various kinds of damped Newton-type methods have been proposed [1, 4, 5, 6, 10, 11, 13, 19, 20, 21, 22, 23, 24]. These methods converge globally and superlinearly/quadratically under suitable conditions. Compared with Newton-type methods, however, little attention has been paid to quasi-Newton methods. One of the reasons is perhaps the lack of effective line search technique for quasi-Newton methods. In iterative methods for solving optimization problems, line searches such as Armijo-type and Wolfe-type searches are widely used. The implementation of these line searches relies on the calculation of gradients of the objective function. These techniques can be applied to Newton-type methods for solving VI(X,F) since Newton's method admits the calculation of ∇F. Indeed, Armijo-type line search has been implemented successfully in the context of damped Newton methods for solving VI(X,F) (see the above-mentioned references). For quasi-Newton methods, however, line searches such as Armijo-type search and Wolfe-type search seem inappropriate because the objective function of

an equivalent optimization problem for VI(X,F) usually depends on F and its gradients rely on the Jacobian ∇F. On the other hand, even for smooth equations, the development of line searches in quasi-Newton methods is relatively scarce. This fact makes it more difficult to study quasi-Newton methods for nonsmooth equations derived from VI(X,F). In is therefore desirable to develop an effective line search technique that does not require the calculation of ∇F. Such a line search is called a derivative-free line search. To the authors' knowledge, the first derivative-free line search in quasi-Newton methods is the one due to Griewank [9] for Broyden's method for solving smooth equations. This line search has been modified and applied quasi-Newton methods for solving the nonlinear complementarity problem in [18] and [25]. In this paper, we propose a new derivative-free line search technique and apply it to Broyden-like methods for finding a vector $z^* = (x^*, \lambda^*, \mu^*) \in R^{n+m_1+m_2}$ that satisfies the KKT system

$$\left\{ \begin{array}{l} F(x^*) - \nabla g(x^*)\lambda^* - \nabla h(x^*)\mu^* = 0, \\ \lambda^* \geq 0, \quad g(x^*) \geq 0, \quad g(x)^T\lambda = 0, \\ \qquad\qquad\qquad\qquad h(x^*) = 0, \end{array} \right. \tag{1.3}$$

where $\nabla g(x)$ and $\nabla h(x)$ denote the Jacobian matrices of $g(x) = (g_1(x), g_2(x), \ldots, g_{m_1}(x))^T$ and $h(x) = (h_1(x), h_2(x), \ldots, h_{m_2}(x))^T$, respectively, whose columns are the gradients $\nabla g_i(x)$, $i = 1, 2, \ldots, m_1$, and $\nabla h_j(x)$, $j = 1, 2, \ldots, m_2$, respectively. The vectors $\lambda^* \in R^{m_1}$ and $\mu^* \in R^{m_2}$ are called Lagrange multipliers at z^*. It is well-known that under certain constraint qualifications, a solution of VI(X,F) together with some multipliers satisfies the KKT system (1.3) (see e.g. [12]).

An effective way of solving VI(X,F) is to develop iterative methods for finding a vector z that satisfies the KKT system (1.3) or its nonsmooth equation reformulation

$$\left(\begin{array}{c} L(z) \\ \min(\lambda, g(x)) \\ h(x) \end{array} \right) = 0, \tag{1.4}$$

where

$$L(z) = F(x) - \nabla g(x)\lambda - \nabla h(x)\mu, \tag{1.5}$$

and

$$\min(\lambda, g(x)) = (\min(\lambda_1, g_1(x)), \min(\lambda_2, g_2(x)), \ldots, \min(\lambda_{m_1}, g_{m_1}(x)))^T.$$

In [19] and [20], a Newton-type method is proposed on the basis of (1.4) and its global and superlinear convergence is proved under suitable conditions. In this paper, we develop a quasi-Newton method of Broyden-type for the KKT system (1.3) by way of another equivalent nonsmooth equation reformulation. In particular, we propose a new derivative-free line search pertaining to quasi-Newton methods. Under suitable conditions, we show global and superlinear convergence of

the proposed method. Moreover, for the nonlinear complementarity problem (1.2), we specialize the method so that the subproblems are linear equations of dimension n. We show that the specialized Broyden-like method converges globally if F is a P_0 function with Lipschitz continuous Jacobian. Moreover, if the strict complementarity condition holds, then the convergence rate is superlinear.

The paper is organized as follows: In the next section, we give some preliminary results concerning smooth approximations of (1.3). In Section 3, we propose a Broyden-like method for solving (1.3) based on its smooth approximations and prove some useful lemmas. In Sections 4 and 5, we establish global and superlinear convergence of the proposed method, respectively. In Section 6, we specialize the method to the nonlinear complementarity problem and establish its global and superlinear convergence under weaker conditions. Section 7 concludes the paper.

2. Preliminaries

Throughout the paper, we consider approximating the KKT system (1.3) by means of the perturbed Fischer-Burmeister function $\phi^\epsilon : R^2 \to R$ defined by

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

where ϵ is a nonnegative parameter. When $\epsilon = 0$, ϕ^0 reduces to the ordinary Fischer-Burmeister function. The Fischer-Burmeister function and its perturbed version have been widely used in the context of the nonlinear complementarity problem and related problems (see e.g. [1, 4, 5, 6, 8, 13, 14, 24]). An important property of ϕ^0 is that $\phi^0(a,b) = 0$ holds if and only if $a \geq 0, b \geq 0$ and $ab = 0$. Hence, (1.3) can be rewritten as

$$\begin{aligned} F(x^*) - \nabla g(x^*)\lambda^* - \nabla h(x^*)\mu^* &= 0, \\ \phi^0(\lambda_i^*, g_i(x^*)) &= 0, \quad i = 1,2,\ldots,m_1, \\ h_j(x^*) &= 0, \quad j = 1,2,\ldots,m_2. \end{aligned}$$

Let $z = (x, \lambda, \mu)$ and define $\Phi^\epsilon(z) = (\phi_1^\epsilon(z), \phi_2^\epsilon(z), \ldots, \phi_{m_1}^\epsilon(z))^T$ with elements

$$\phi_i^\epsilon(z) \stackrel{\triangle}{=} \phi^\epsilon(\lambda_i, g_i(x)) = \sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2} - (\lambda_i + g_i(x)), \quad i = 1,2,\cdots,m_1 \tag{2.1}$$

and

$$H^\epsilon(z) \stackrel{\triangle}{=} \begin{pmatrix} L(z) \\ -\Phi^\epsilon(z) \\ -h(x) \end{pmatrix}. \tag{2.2}$$

The KKT system (1.3) is then equivalent to the nonsmooth equation

$$H^0(z) = 0. \tag{2.3}$$

Note that for all $z = (x, \lambda, \mu) \in R^{n+m_1+m_2}$, any $\epsilon > 0$ and every $i = 1, 2, \ldots, m_1$

$$\begin{aligned} |\phi_i^\epsilon(z) - \phi_i^0(z)| &= \sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2} - \sqrt{\lambda_i^2 + g_i(x)^2} \\ &= \frac{\epsilon^2}{\sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2} + \sqrt{\lambda_i^2 + g_i(x)^2}} \\ &\leq \epsilon. \end{aligned}$$

This implies that $\|\Phi^\epsilon(z) - \Phi^0(z)\| \leq \sqrt{m_1}\epsilon$ holds for all z and any $\epsilon > 0$. Moreover, H^ϵ is differentiable everywhere, provided that $\epsilon > 0$, F is differentiable and g_i, $i = 1, 2, \ldots, m_1$ and h_j, $j = 1, 2, \ldots m_2$, are twice differentiable. Therefore when $\epsilon > 0$ is small, the equation

$$H^\epsilon(z) = 0 \tag{2.4}$$

can be regarded as a smooth approximation of (2.3). The following lemma summarizes some useful properties of ϕ^ϵ.

Lemma 2.1 *For any $\epsilon' > \epsilon \geq 0$, we have*

$$0 < \phi^{\epsilon'}(a, b) - \phi^\epsilon(a, b) \leq \epsilon' - \epsilon, \qquad \forall (a, b) \in R^2. \tag{2.5}$$

Consequently, we have for any $\epsilon' > \epsilon \geq 0$

$$\|H^\epsilon(z) - H^{\epsilon'}(z)\| = \|\Phi^\epsilon(z) - \Phi^{\epsilon'}(z)\| \leq \sqrt{m_1}(\epsilon' - \epsilon), \quad \forall z \in R^{n+m_1+m_2}. \tag{2.6}$$

In particular, we have for any $\epsilon > 0$

$$0 < \phi^\epsilon(a, b) - \phi^0(a, b) \leq \epsilon, \qquad \forall (a, b) \in R^2, \tag{2.7}$$

and

$$\|H^\epsilon(z) - H^0(z)\| \leq \sqrt{m_1}\epsilon, \quad \forall z \in R^{n+m_1+m_2}. \tag{2.8}$$

Proof (i) It is obvious that ϕ^ϵ increases with $|\epsilon|$. So, we get for any $\epsilon' > \epsilon \geq 0$ and every $(a, b) \in R^2$

$$\begin{aligned} 0 &< \phi^{\epsilon'}(a, b) - \phi^\epsilon(a, b) \\ &= \sqrt{a^2 + b^2 + (\epsilon')^2} - \sqrt{a^2 + b^2 + \epsilon^2} \\ &= \frac{(\epsilon')^2 - \epsilon^2}{\sqrt{a^2 + b^2 + (\epsilon')^2} + \sqrt{a^2 + b^2 + \epsilon^2}} \\ &\leq \frac{(\epsilon')^2 - \epsilon^2}{\epsilon' + \epsilon} \\ &= \epsilon' - \epsilon. \end{aligned}$$

This shows inequality (2.5). The inequality (2.6) then follows from (2.5) and the definition of H^ϵ directly. □

Next we consider the differentiability of ϕ^ϵ and related properties. Put

$$a_i^\epsilon(x,\lambda) = \frac{\lambda_i}{\sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2}} - 1, \quad b_i^\epsilon(x,\lambda) = \frac{g_i(x)}{\sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2}} - 1.$$

Then the following lemma is straightforward from the definition of $a_i^\epsilon(x,\lambda)$ and $b_i^\epsilon(x,\lambda)$.

Lemma 2.2 *For every $\epsilon \neq 0$ and any $(x,\lambda) \in R^{n+m_1}$, we have $-2 < a_i^\epsilon(x,\lambda) < 0$ and $-2 < b_i^\epsilon(x,\lambda) < 0$ for all $i = 1,2,\ldots,m_1$.*

It is also easy to show the following lemma by using the chain rule of derivatives for composite functions.

Lemma 2.3 *Let F be continuously differentiable and g_i, $i = 1,2,\ldots,m_1$, and h_j, $j = 1,2,\ldots,m_2$, be twice continuously differentiable. Then Φ^ϵ and H^ϵ are continuously differentiable for every $\epsilon \neq 0$. Moreover, for every $z = (x,\lambda,\mu) \in R^{n+m_1+m_2}$ and any $\epsilon \neq 0$, we have the following expressions:*

$$\nabla_x \Phi^\epsilon(x,\lambda) = \nabla g(x)\ \mathrm{diag}\ (b_i^\epsilon(x,\lambda)), \qquad \nabla_\lambda \Phi^\epsilon(x,\lambda) = \mathrm{diag}\ (a_i^\epsilon(x,\lambda)),$$

and

$$\nabla H^\epsilon(z) = \begin{pmatrix} \nabla_x L(z) & -\nabla g(x)\ \mathrm{diag}\ (b_i^\epsilon(x,\lambda)) & -\nabla h(x) \\ -\nabla g(x)^T & -\ \mathrm{diag}\ (a_i^\epsilon(x,\lambda)) & 0 \\ -\nabla h(x)^T & 0 & 0 \end{pmatrix}. \tag{2.9}$$

In the remainder of this section, we abbreviate $a_i^\epsilon(x,\lambda)$ and $b_i^\epsilon(x,\lambda)$ as a_i^ϵ and b_i^ϵ, respectively. It is clear from Lemma 2.2 that $\mathrm{diag}(a_i^\epsilon)$ and $\mathrm{diag}(b_i^\epsilon)$ are nonsingular. Moreover, a_i^ϵ and b_i^ϵ are negative. This implies that for any $(x,\lambda) \in R^{n+m_1}$, the diagonal matrix $\mathrm{diag}\,((a_i^\epsilon)^{-1}b_i^\epsilon)$ is positive definite, and hence, the matrix $\nabla g(x)\mathrm{diag}((a_i^\epsilon)^{-1}b_i^\epsilon)\nabla g(x)^T$ is symmetric and positive semidefinite for any $(x,\lambda) \in R^{n+m_1}$. The following lemma shows some conditions that ensure regularity of H^ϵ.

Lemma 2.4 *Let $z = (x,\lambda,\mu) \in R^{n+m_1+m_2}$ and $\epsilon \neq 0$. If $\nabla h_j(x), j = 1,2,\ldots,m_2$, are linearly independent and the matrix*

$$\nabla_x L(z) + \nabla g(x)\mathrm{diag}((a_i^\epsilon)^{-1}b_i^\epsilon)\nabla g(x)^T$$

is positive definite on the subspace $\{p \,|\, \nabla h(x)^T p = 0\}$, then $\nabla H^\epsilon(z)$ is nonsingular. In particular, if $\nabla_x L(z)$ is positive definite, then $\nabla H^\epsilon(z)$ is nonsingular.

Proof Let $w = (p, q, r) \in R^{n+m_1+m_2}$ be a solution of the system of linear equations $\nabla H^\epsilon(z)w = 0$, i.e.,

$$\begin{pmatrix} \nabla_x L(z) & -\nabla g(x)\ \mathrm{diag}\ (b_i^\epsilon) & -\nabla h(x) \\ -\nabla g(x)^T & -\ \mathrm{diag}\ (a_i^\epsilon) & 0 \\ -\nabla h(x)^T & 0 & 0. \end{pmatrix} \begin{pmatrix} p \\ q \\ r \end{pmatrix} = 0.$$

We will show that $w = 0$. Writing the above equations blockwise, we get

$$\begin{cases} \nabla_x L(z)p & - & \nabla g(x)\ \mathrm{diag}\ (b_i^\epsilon)q & - & \nabla h(x)r & = & 0, \\ -\nabla g(x)^T p & - & \mathrm{diag}\ (a_i^\epsilon)q & & & = & 0, \\ -\nabla h(x)^T p & & & & & = & 0, \end{cases} \tag{2.10}$$

which implies

$$\begin{cases} \Big(\nabla_x L(z) + \nabla g(x)\mathrm{diag}((a_i^\epsilon)^{-1}b_i^\epsilon)\nabla g(x)^T\Big)p - \nabla h(x)r = 0, \\ \nabla h(x)^T p = 0. \end{cases}$$

Multiplying the first equation by p yields

$$p^T\Big(\nabla_x L(z) + \nabla g(x)\mathrm{diag}((a_i^\epsilon)^{-1}b_i^\epsilon)\nabla g(x)^T\Big)p = 0. \tag{2.11}$$

However, since p lies in the subspace $\{p \,|\, \nabla h(x)^T p = 0\}$, it follows from the positive definiteness of $\nabla_x L(z) + \nabla g(x)\mathrm{diag}((a_i^\epsilon)^{-1}b_i^\epsilon)\nabla g(x)^T$ on the subspace that $p = 0$. Then, the second equation in (2.10) implies that $q = 0$. Finally, since $\nabla h_j(x)$, $j = 1, 2, \ldots, m_2$, are linearly independent, it follows from the first equation of (2.10) that $r = 0$. Consequently, zero is the unique solution of (2.10), or equivalently, $\nabla H^\epsilon(z)$ is nonsingular. The last part of the lemma is a direct consequence of the first part by the positive semidefiniteness of $\nabla g(x)\ \mathrm{diag}\ ((a_i^\epsilon)^{-1}b_i^\epsilon)\nabla g(x)^T$. □

3. Algorithm

This section is devoted to presenting a Broyden-like method for solving the KKT system (1.3). For convenience, we simplify some notations. For a given positive sequence $\{\epsilon_k\}$, we abbreviate H^{ϵ_k}, Φ^{ϵ_k}, ϕ^{ϵ_k}, $a_i^{\epsilon_k}$, $b_i^{\epsilon_k}$, etc. as H^k, Φ^k, ϕ^k, a_i^k, b_i^k, etc., respectively. Note that H^0, Φ^0, ϕ^0, a_i^0, b_i^0, etc. will still denote H^ϵ, Φ^ϵ, ϕ^ϵ, a_i^ϵ, b_i^ϵ, etc. with $\epsilon = 0$, respectively.

The subproblem in the proposed Broyden-like method takes the form

$$B_k p + H^0(z_k) = 0,$$

which is similar to the one in [18] and [25]. If $B_k = \nabla H^k(z_k)$, then it reduces to the subproblem of the successive approximation method proposed by Qi and Chen [23]. However, the function H^0 here is different from those in [18], [23] and [25].

We now propose a new derivative-free line search. Let $\{\eta_k\}$ be a positive sequence satisfying

$$\sum_{k=1}^{\infty} \eta_k \leq \eta < \infty, \tag{3.1}$$

where η is a positive constant. At iteration k, given the current iterate z_k and a direction p_k, we determine a positive steplength $\alpha_k > 0$ so that the following inequality holds for $\alpha = \alpha_k$:

$$\begin{aligned} &\|H^k(z_k + \alpha p_k)\| - \|H^k(z_k)\| \\ &\leq \quad -\sigma_1 \|H^k(z_k + \alpha p_k) - H^k(z_k)\| - \sigma_2 \|\alpha p_k\|^2 + \eta_k, \end{aligned} \tag{3.2}$$

where $\sigma_1, \sigma_2 > 0$ are suitable constants. Note that (3.2) is satisfied for all sufficiently small $\alpha > 0$ since η_k is positive and independent of α. Thus α_k can be obtained by a backtracking process. The inequality (3.2) shows an approximate norm descent property for H^k.

The steps of the algorithm are as follows.

Algorithm 1

Step 0. Choose constants $\rho, \beta \in (0,1)$, $0 < \gamma < \frac{1}{\sqrt{m_1}}$, $\sigma_1 \in (0, \frac{1-\sqrt{m_1}\gamma}{1+\sqrt{m_1}\gamma})$, $\sigma_2 > 0$, $\bar{\theta} \in (0,1)$. Select a positive sequence $\{\eta_k\}$ satisfying (3.1). Choose an initial point $z_1 \in R^{n+m_1+m_2}$ and an initial nonsingular matrix $B_1 \in R^{(n+m_1+m_2)\times(n+m_1+m_2)}$. Choose an $\epsilon_1 > 0$ satisfying $\epsilon_1 \leq \frac{\gamma}{2}\|H^0(z_1)\|$. Let $k := 1$.

Step 1. If $H^0(z_k) = 0$, stop. Otherwise, solve the linear equation

$$B_k p + H^0(z_k) = 0 \tag{3.3}$$

to get p_k.

Step 2. Let α_k be the maximum element in the set $\{1, \rho, \rho^2, \ldots\}$ such that $\alpha = \rho^i$ satisfies the line search condition (3.2).

Step 3. Let $z_{k+1} = z_k + \alpha_k p_k$.

Step 4. Update B_k by the Broyden-like update formula

$$B_{k+1} = B_k + \theta_k \frac{(y_k - B_k s_k) s_k^T}{\|s_k\|^2}, \tag{3.4}$$

where $s_k = z_{k+1} - z_k$ and $y_k = H^k(z_{k+1}) - H^k(z_k)$. The parameter θ_k is chosen so that $|\theta_k - 1| \leq \bar{\theta}$ and B_{k+1} is nonsingular.

Step 5. Update ϵ_{k+1} by the following rule: If $\epsilon_k < \gamma\|H^0(z_{k+1})\|$, set $\epsilon_{k+1} = \epsilon_k$. Otherwise, let ϵ_{k+1} satisfy

$$0 < \epsilon_{k+1} \leq \min\{\frac{\gamma}{2}\|H^0(z_{k+1})\|, \frac{1}{2}\epsilon_k\}.$$

Step 6. Let $k := k + 1$. Go to Step 1.

Remark Step 5 of Algorithm 1 ensures that $\{\epsilon_k\}$ is monotonically nonincreasing and $\epsilon_k \leq$

$\gamma\|H^0(z_k)\|$ holds for every k. Moreover, if the algorithm does not terminate finitely, i.e., $H^0(z_k) \neq 0$ for all k, then $\epsilon_k > 0$ for all k. We also notice that $\{z_k\}$ satisfies the inequality

$$\|H^k(z_{k+1})\| - \|H^k(z_k)\| \leq -\sigma_1\|H^k(z_{k+1}) - H^k(z_k)\| - \sigma_2\|s_k\|^2 + \eta_k.$$

By (2.6), it then follows that for every k

$$\begin{aligned} \|H^{k+1}(z_{k+1})\| &\leq \|H^k(z_{k+1})\| + \|H^{k+1}(z_{k+1}) - H^k(z_{k+1})\| \\ &\leq \|H^k(z_{k+1})\| + \sqrt{m_1}(\epsilon_k - \epsilon_{k+1}) \\ &\leq \|H^k(z_k)\| - \sigma_1\|H^k(z_{k+1}) - H^k(z_k)\| - \sigma_2\|s_k\|^2 \\ &\quad + \eta_k + \sqrt{m_1}(\epsilon_k - \epsilon_{k+1}). \end{aligned} \tag{3.5}$$

Moreover, (3.5) implies

$$\|H^{k+1}(z_{k+1})\| \leq \|H^k(z_k)\| + \eta_k + \sqrt{m_1}(\epsilon_k - \epsilon_{k+1}). \tag{3.6}$$

Since η_k satisfies (3.1), we immediately get the following result.

Lemma 3.1 *Let $\{z_k\}$ be generated by Algorithm 1. Then $\{\|H^k(z_k)\|\}$ converges. Moreover, we have*

$$\sum_{k=1}^{\infty} \|H^k(z_{k+1}) - H^k(z_k)\| < \infty \tag{3.7}$$

and

$$\sum_{k=1}^{\infty} \|s_k\|^2 = \sum_{k=1}^{\infty} \|z_{k+1} - z_k\|^2 < \infty. \tag{3.8}$$

Proof Since (3.1) implies

$$\sum_{k=1}^{\infty} \{\eta_k + \sqrt{m_1}(\epsilon_k - \epsilon_{k+1})\} < \infty,$$

it follows from (3.6) that $\{\|H^k(z_k)\|\}$ converges. The inequality (3.7) then follows from (3.5) directly. Moreover, by (3.5), it is easy to see that

$$\sigma_2\|s_k\|^2 \leq \|H^k(z_k)\| - \|H^{k+1}(z_{k+1})\| + \eta_k + \sqrt{m_1}(\epsilon_k - \epsilon_{k+1}).$$

Summing the above inequalities yields (3.8). □

Define the level set

$$\Omega = \{z \in R^{n+m_1+m_2} \mid \|H^0(z)\| \leq \frac{1 + \sqrt{m_1}\gamma}{1 - \sqrt{m_1}\gamma}\|H^0(z_1)\| + \frac{1}{1 - \sqrt{m_1}\gamma}\eta\}, \tag{3.9}$$

where η is a positive constant such that (3.1) holds.

Lemma 3.2 *Let $\{z_k\}$ be generated by Algorithm 1. Then $\{z_k\} \subset \Omega$.*

Proof Summing (3.6), we get for every k

$$\begin{aligned} \|H^{k+1}(z_{k+1})\| &\leq \|H^1(z_1)\| + \eta + \sqrt{m_1}\epsilon_1 \\ &\leq \|H^0(z_1)\| + \eta + 2\sqrt{m_1}\epsilon_1 \\ &\leq (1+\sqrt{m_1}\gamma)\|H^0(z_1)\| + \eta, \end{aligned} \tag{3.10}$$

where the second inequality follows from (2.8) and the last inequality follows from the choice of ϵ_1. By using (2.8) again, we have

$$\begin{aligned} \|H^0(z_k)\| &\leq \|H^k(z_k)\| + \|H^k(z_k) - H^0(z_k)\| \\ &\leq \|H^k(z_k)\| + \sqrt{m_1}\epsilon_k \\ &\leq \|H^k(z_k)\| + \sqrt{m_1}\gamma\|H^0(z_k)\|, \end{aligned}$$

where the last inequality holds since by Step 5, $\epsilon_k \leq \gamma\|H^0(z_k)\|$ for every k. So, we deduce from (3.10) and the last inequality

$$\|H^0(z_k)\| \leq \frac{1}{1-\sqrt{m_1}\gamma}\|H^k(z_k)\| \leq \frac{1+\sqrt{m_1}\gamma}{1-\sqrt{m_1}\gamma}\|H^0(z_1)\| + \frac{1}{1-\sqrt{m_1}\gamma}\eta.$$

This means that $\{z_k\} \subset \Omega$ as desired. □

In the rest of the paper, we make the following assumption.

Assumption A (i) The level set Ω defined by (3.9) is contained in a bounded convex set D.

(ii) The function F is continuously differentiable on D_x, and g_i, $i = 1, 2, \ldots, m_1$, and h_j, $j = 1, 2, \ldots, m_2$, are twice continuously differentiable on D_x,where $D_x = \{x \in R^n \,|\, (x, \lambda, \mu) \in D\}$.

(iii) ∇F, $\nabla^2 g$ and $\nabla^2 h$ are Lipschitz continuous on D_x, i.e., there is a constant $\tau > 0$ such that

$$\begin{cases} \|\nabla F(x) - \nabla F(y)\| \leq \tau\|x-y\|, & \forall x, y \in D_x \\ \|\nabla^2 g(x) - \nabla^2 g(y)\| \leq \tau\|x-y\|, & \forall x, y \in D_x \\ \|\nabla^2 h(x) - \nabla^2 h(y)\| \leq \tau\|x-y\|, & \forall x, y \in D_x. \end{cases} \tag{3.11}$$

For given $\epsilon > 0$ and $\theta > 0$, define the set

$$\Pi^\epsilon(\theta) = \{z = (x, \lambda, \mu) \,|\, \sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2} \geq \theta, i = 1, 2, \ldots, m\} \cap D. \tag{3.12}$$

Note that $\Pi^\epsilon(\theta)$ is bounded for any $\epsilon > 0$ and $\theta > 0$ since D is bounded. The following lemma shows that the Lipschitz condition (3.11) implies the Lipschitz continuity of ∇H^k for every k with a common Lipschitz constant independent of k.

Lemma 3.3 *Let Assumption A hold. Then for any $\theta > 0$ and $\epsilon > 0$, there is a constant $\tau_1 > 0$ such that*

$$\|\nabla H^\epsilon(z) - \nabla H^\epsilon(z')\| \leq \tau_1\|z - z'\|, \qquad \forall z, z' \in \Pi^\epsilon(\theta). \tag{3.13}$$

Proof We prove the Lipschitz continuity of ∇H^ϵ by showing that every block of ∇H^ϵ in (2.9) is Lipschitz continuous.

In view of the boundedness of $\Pi^\epsilon(\theta)$, the Lipschitz continuity of $\nabla_x L$ follows from (3.11) directly. It is also clear by the twice continuous differentiability of g and h that ∇g^T, ∇h^T and ∇g, ∇h are Lipschitz continuous. So it suffices to show that the second block column of ∇H^ϵ satisfies the Lipschitz condition. By an elementary deduction, we have for any $z = (x, \lambda, \mu), z' = (x', \lambda', \mu') \in R^{n+m_1+m_2}$

$$\begin{aligned}
&\|\nabla g(x)\mathrm{diag}\,(b_i^\epsilon(x,\lambda)) - \nabla g(x')\mathrm{diag}\,(b_i^\epsilon(x',\lambda'))\| \\
&\quad\le \|\nabla g(x) - \nabla g(x')\|\,\|\mathrm{diag}\,(b_i^\epsilon(x,\lambda))\| + \|\nabla g(x')\|\,\|\mathrm{diag}\,(b_i^\epsilon(x,\lambda)) - \mathrm{diag}\,(b_i^\epsilon(x',\lambda'))\| \\
&\quad\le 2\|\nabla g(x) - \nabla g(x')\| + \|\nabla g(x')\|\,\|\mathrm{diag}\,(b_i^\epsilon(x,\lambda)) - \mathrm{diag}\,(b_i^\epsilon(x',\lambda'))\| \\
&\quad\le 2\bar\tau\|x - x'\| + \|\nabla g(x')\|\,\|\mathrm{diag}\,(b_i^\epsilon(x,\lambda)) - \mathrm{diag}\,(b_i^\epsilon(x',\lambda'))\|, \qquad (3.14)
\end{aligned}$$

where $\bar\tau$ denotes an upper bound of $\|\nabla^2 g\|$ on $\Pi^\epsilon(\theta)$, the second inequality follows from Lemma 2.2 and the last inequality follows from the twice continuous differentiability of g. We estimate the last term of (3.14) elementwise. From the definition of b_i^ϵ, we have for every $i = 1, 2, \ldots, m_1$ and any $z, z' \in \Pi^\epsilon(\theta)$

$$\begin{aligned}
&|b_i^\epsilon(x,\lambda) - b_i^\epsilon(x',\lambda')| \\
&\quad= \Big|\Big[\frac{g_i(x)}{\sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2}} - 1\Big] - \Big[\frac{g_i(x')}{\sqrt{(\lambda_i')^2 + g_i(x')^2 + \epsilon^2}} - 1\Big]\Big| \\
&\quad= \Big|\frac{g_i(x)}{\sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2}} - \frac{g_i(x')}{\sqrt{(\lambda_i')^2 + g_i(x')^2 + \epsilon^2}}\Big| \\
&\quad\le \Big|\frac{g_i(x) - g_i(x')}{\sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2}}\Big| \\
&\qquad + \Big|g_i(x')\Big[\frac{1}{\sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2}} - \frac{1}{\sqrt{(\lambda_i')^2 + g_i(x')^2 + \epsilon^2}}\Big]\Big|. \qquad (3.15)
\end{aligned}$$

For the first term of (3.15), we have for every $i = 1, 2, \ldots, m_1$ and any $z, z' \in \Pi^\epsilon(\theta)$

$$\Big|\frac{g_i(x) - g_i(x')}{\sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2}}\Big| \le \frac{\|g(x) - g(x')\|}{\theta} \le \frac{\bar M}{\theta}\|x - x'\|, \qquad (3.16)$$

where $\bar M > 0$ stands for an upper bound of $\|\nabla g\|$ on $\Pi^\epsilon(\theta)$. For the second term of (3.15) we have for every $i = 1, 2, \ldots, m_1$ and any $z, z' \in \Pi^\epsilon(\theta)$

$$\Big|g_i(x')\Big[\frac{1}{\sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2}} - \frac{1}{\sqrt{(\lambda_i')^2 + g_i(x')^2 + \epsilon^2}}\Big]\Big|$$

$$
\begin{aligned}
&= |g_i(x')| \Big| \frac{\sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2} - \sqrt{(\lambda_i')^2 + g_i(x')^2 + \epsilon^2}}{\sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2}\sqrt{(\lambda_i')^2 + g_i(x')^2 + \epsilon^2}} \Big| \\
&\leq \theta^{-2}\|g(x')\| \Big| \sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2} - \sqrt{(\lambda_i')^2 + g_i(x')^2 + \epsilon^2} \Big| \\
&= \theta^{-2}\|g(x')\| \frac{\Big| \Big(\lambda_i^2 + g_i(x)^2\Big) - \Big((\lambda_i')^2 + g_i(x')^2\Big) \Big|}{\sqrt{\lambda_i^2 + g_i(x)^2 + \epsilon^2} + \sqrt{(\lambda_i')^2 + g_i(x')^2 + \epsilon^2}} \\
&\leq \frac{1}{2}\theta^{-3}M'\{|\lambda_i^2 - (\lambda_i')^2| + |g_i(x)^2 - g_i(x')^2|\} \\
&\leq \frac{1}{2}\theta^{-3}M'\{(\|\lambda\| + \|\lambda'\|)\|\lambda - \lambda'\| + (\|g(x)\| + \|g(x')\|)\|g(x) - g(x')\|\},
\end{aligned}
$$

where M' is an upper bound of $\|g\|$ on $\Pi^\epsilon(\theta)$. Notice that z and z' are in the bounded set $\Pi^\epsilon(\theta)$, the above inequality together with (3.16) implies the Lipschitz continuity of diag (b_i^ϵ). By a similar argument, we can show the Lipschitz continuity of diag(a_i^ϵ). The proof is complete. □

4. Global Convergence

In this section, we show that under suitable conditions, Algorithm 1 converges globally.

Let us define the index set $\bar{K} \subset \{1, 2, \ldots\}$ by

$$
\bar{K} = \{k \mid \epsilon_k \geq \gamma\|H^0(z_{k+1})\|\}. \tag{4.1}
$$

Then we have the following result.

Theorem 4.1 *Let $\{z_k\}$ be generated by Algorithm 1. If $\bar{K}$ is infinite, then*

$$
\liminf_{k\to\infty} \|H^0(z_k)\| = 0. \tag{4.2}
$$

Proof Let $\bar{K} = \{k_0 < k_1 < k_2, \cdots\}$. Then by Step 5, for every k such that $k_{i-1} < k \leq k_i$, we have $\epsilon_k = \epsilon_{k_i}$. So, it follows that

$$
\begin{aligned}
\|H^0(z_{k_i+1})\| &\leq \gamma^{-1}\epsilon_{k_i} = \gamma^{-1}\epsilon_{k_{i-1}+1} \leq \frac{1}{2}\|H^0(z_{k_{i-1}+1})\| \\
&\leq \cdots \leq (\frac{1}{2})^i\|H^0(z_{k_0+1})\| \leq \gamma^{-1}(\frac{1}{2})^i\epsilon_{k_0} \leq (\frac{1}{2})^i\|H^0(z_{k_0})\|,
\end{aligned} \tag{4.3}
$$

where the last inequality holds since by Step 5 of Algorithm 1, $\epsilon_k \leq \gamma\|H^0(z_k)\|$ holds for every k. This implies (4.2). □

Theorem 4.1 reveals that to show that Algorithm 1 converges globally, it suffices to verify that $\bar{K}$ defined by (4.1) is infinite. In the following, we prove this assertion by way of contradiction.

First, we show the following useful lemma, which is a slight extension of the related result in [9] and Theorem 3.1 in [15]. Let δ_k be defined by

$$\delta_k = \frac{\|y_k - B_k s_k\|}{\|s_k\|}. \tag{4.4}$$

Lemma 4.1 *Let Assumption A hold. If $\bar{K}$ is finite, then*

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=2}^{k} \delta_i^2 = 0. \tag{4.5}$$

In particular, there is an infinite index set $\bar{K}_1$ such that $\delta_k \to 0$ as $k \to \infty$ with $k \in \bar{K}_1$.

Proof When $\bar{K}$ is finite, we have $\epsilon_k = \epsilon_{\bar{k}}$ for all $k \geq \bar{k}$ with some index $\bar{k}$, which in turn implies that $H^k = H^{\bar{k}}$ for all $k \geq \bar{k}$. Without loss of generality, we assume that $\epsilon_k = \bar{\epsilon}$ and $H^k = \bar{H}$ for all k with some positive constant $\bar{\epsilon}$ and some function $\bar{H} : R^{n+m_1+m_2} \to R^{n+m_1+m_2}$. Let

$$A_k = \int_0^1 \nabla \bar{H}(z_k + ts_k)dt. \tag{4.6}$$

Then we have $y_k = A_k s_k$, and we get from (3.4)

$$\begin{aligned}
\|B_{k+1} - A_k\|_F^2 &= \left\| B_k - A_k + \theta_k (A_k - B_k)\frac{s_k s_k^T}{\|s_k\|^2} \right\|_F^2 \\
&= \|B_k - A_k\|_F^2 + (\theta_k^2 - 2\theta_k)\frac{\|(B_k - A_k)s_k\|^2}{\|s_k\|^2} \\
&= \|B_k - A_k\|_F^2 + (\theta_k^2 - 2\theta_k)\frac{\|y_k - B_k s_k\|^2}{\|s_k\|^2} \\
&= \|B_k - A_k\|_F^2 + (\theta_k^2 - 2\theta_k)\delta_k^2 \\
&\leq \|B_k - A_k\|_F^2 - (1 - \bar{\theta}^2)\delta_k^2,
\end{aligned} \tag{4.7}$$

where the inequality holds since, in Step 4, θ_k is chosen to satisfy $|1 - \theta_k| \leq \bar{\theta}$ for every k, implying that $\theta_k^2 - 2\theta_k \leq -(1 - \bar{\theta}^2)$ holds for every k. However, since $\epsilon_k = \bar{\epsilon} > 0$ for all k, it follows that z_k, z_{k-1}, $z_k + ts_k$ and $z_{k-1} + ts_{k-1}$ are in $\Pi^{\bar{\epsilon}}(\bar{\epsilon})$ for all $t \in (0, 1)$, where $\Pi^{\bar{\epsilon}}(\bar{\epsilon})$ is as defined by (3.12). Therefore, we get from Lemma 3.3 that there is a positive constant τ_2 such that

$$\begin{aligned}
\|A_k - A_{k-1}\| &\leq \int_0^1 \|\nabla \bar{H}(z_k + ts_k) - \nabla \bar{H}(z_{k-1} + ts_{k-1})\|dt \\
&\leq \tau_2 \int_0^1 \|(z_k + ts_k) - (z_{k-1} + ts_{k-1})\|dt \\
&= \tau_2 \int_0^1 \|ts_k + (1-t)s_{k-1}\|dt \\
&\leq \tau_2 \Big(\|s_k\| + \|s_{k-1}\| \Big) \\
&= \tau_2 \nu_k,
\end{aligned}$$

where $\nu_k = \|s_k\| + \|s_{k-1}\|$. So we get from (4.7)

$$\begin{aligned}
\|B_{k+1} - A_k\|_F &\leq \|B_k - A_k\|_F \\
&\leq \|B_k - A_{k-1}\|_F + \|A_k - A_{k-1}\|_F \\
&\leq \|B_k - A_{k-1}\|_F + \tau_2\nu_k \\
&\vdots \\
&\leq \|B_2 - A_1\|_F + \tau_2 \sum_{i=2}^{k} \nu_i.
\end{aligned}$$

It then follows from (4.7) again that

$$\begin{aligned}
(1-\bar{\theta}^2)\delta_k^2 &\leq \|B_k - A_k\|_F^2 - \|B_{k+1} - A_k\|_F^2 \\
&\leq \Big(\|B_k - A_{k-1}\|_F + \|A_k - A_{k-1}\|_F\Big)^2 - \|B_{k+1} - A_k\|_F^2 \\
&\leq \Big(\|B_k - A_{k-1}\|_F + \tau_2\nu_k\Big)^2 - \|B_{k+1} - A_k\|_F^2 \\
&= \|B_k - A_{k-1}\|_F^2 - \|B_{k+1} - A_k\|_F^2 + 2\tau_2\nu_k\|B_k - A_{k-1}\|_F + \tau_2^2\nu_k^2.
\end{aligned}$$

Summing these inequalities yields

$$\begin{aligned}
(1-\bar{\theta}^2)\sum_{i=2}^{k} \delta_i^2 &\leq \|B_2 - A_1\|_F^2 + 2\tau_2 \sum_{i=2}^{k} \nu_i\|B_i - A_{i-1}\|_F + \tau_2^2 \sum_{i=2}^{k} \nu_i^2 \\
&\leq \|B_2 - A_1\|_F^2 + 2\tau_2 \sum_{i=2}^{k} \nu_i\Big(\|B_2 - A_1\|_F + \tau_2 \sum_{j=2}^{i-1} \nu_j\Big) + \tau_2^2 \sum_{i=2}^{k} \nu_i^2 \\
&= \|B_2 - A_1\|_F^2 + 2\tau_2\|B_2 - A_1\|_F \sum_{i=2}^{k} \nu_i + 2\tau_2^2 \sum_{i=2}^{k} \nu_i \sum_{j=1}^{i-1} \nu_j + \tau_2^2 \sum_{i=2}^{k} \nu_i^2 \\
&= \Big(\|B_2 - A_1\|_F + \tau_2 \sum_{i=2}^{k} \nu_i\Big)^2.
\end{aligned}$$

Therefore, we deduce for any integer $t > 2$

$$\begin{aligned}
(1-\bar{\theta}^2)\sum_{i=2}^{k} \delta_i^2 &\leq \Big(\|B_2 - A_1\|_F + \tau_2 \sum_{i=2}^{t} \nu_i + \tau_2 \sum_{i=t+1}^{k} \nu_i\Big)^2 \\
&\leq 2\Big(\|B_2 - A_1\|_F + \tau_2 \sum_{i=2}^{t} \nu_i\Big)^2 + 2\tau_2^2\Big(\sum_{i=t+1}^{k} \nu_i\Big)^2 \\
&\leq 2\Big(\|B_2 - A_1\|_F + \tau_2 \sum_{i=2}^{t} \nu_i\Big)^2 + 2\tau_2^2(k-t)\sum_{i=t+1}^{k} \nu_i^2.
\end{aligned}$$

Dividing the both sides by k and letting k go to infinity with t being fixed arbitrarily yield

$$(1-\bar{\theta}^2)\lim_{k\to\infty} \frac{1}{k}\sum_{i=2}^{k} \delta_i^2 \leq 2\lim_{k\to\infty} \frac{1}{k}\Big(\|B_2 - A_1\|_F + \tau_2 \sum_{i=2}^{t} \nu_i\Big)^2 + 2\tau_2^2 \lim_{k\to\infty} \frac{k-t}{k} \sum_{i=t+1}^{k} \nu_i^2$$

$$= \quad 2\tau_2^2 \sum_{i=t+1}^{\infty} \nu_i^2. \tag{4.8}$$

Since

$$\sum_{k=2}^{\infty} \nu_k^2 \le 2 \sum_{k=2}^{\infty} (\|s_k\|^2 + \|s_{k-1}\|^2) < \infty$$

and t can be arbitrarily large, we get (4.5) from (4.8). □

Now we prove global convergence of Algorithm 1. From Theorem 4.1, it suffices to show that $\bar{K}$ defined by (4.1) is infinite.

Theorem 4.2 *Let Assumption A hold. Assume that, for any accumulation point $\bar{z}$ of the sequence $\{z_k\}$ generated by Algorithm 1, $\nabla H^k(\bar{z})$ is nonsingular for all sufficiently large k. Then the index set $\bar{K}$ defined by (4.1) must be infinite. Moreover, we have*

$$\lim_{k \to \infty} \|H^0(z_k)\| = 0 \tag{4.9}$$

and every accumulation point of $\{z_k\}$ is a solution of (1.4).

Proof We assume that $\bar{K}$ is finite to deduce a contradiction. By Step 5 of Algorithm 1, there is an index $\bar{k}$ such that $\epsilon_k = \bar{\epsilon}$ for all $k \ge \bar{k}$ with some constant $\bar{\epsilon} > 0$. Hence $H^k \equiv H^{\bar{\epsilon}}$ for all $k \ge \bar{k}$. Without loss of generality, we assume that $\epsilon_k = \bar{\epsilon}$ and $H^k = \bar{H}$ for all k with some function $\bar{H} : R^{n+m_1+m_2} \to R^{n+m_1+m_2}$.

From Lemma 4.1, there is an infinite index set $\bar{K}_1$ such that $\lim_{k \in \bar{K}_1, k \to \infty} \delta_k = 0$. Since $\{z_k\}_{k \in \bar{K}_1} \subset \Omega$ is bounded, there exists a subsequence $\{z_k\}_{k \in K} \subset \{z_k\}_{k \in \bar{K}_1}$ converging to some point $\bar{z}$. Let A_k be defined by (4.6). Then $y_k = A_k s_k$ and $A_k \to \nabla \bar{H}(\bar{z})$ as $k \to \infty$ with $k \in K$. Moreover, we get from (3.3)

$$\begin{aligned} \|H^0(z_k)\| &= \|B_k p_k\| \\ &\le \|(B_k - A_k) p_k\| + \|A_k p_k\| \\ &\le \delta_k \|p_k\| + \|A_k\| \, \|p_k\|. \end{aligned}$$

Since $\delta_k \to 0$ as $k \to \infty$ with $k \in \bar{K}_1$, there is a constant $M_1 > 0$ such that for all $k \in K$ large enough

$$\|H^0(z_k)\| \le M_1 \|p_k\|. \tag{4.10}$$

Let $\bar{\alpha} = \limsup_{k \in K, k \to \infty} \alpha_k$. If $\bar{\alpha} > 0$, then we have $\lim_{k \in K, k \to \infty} \|p_k\| = 0$ by (3.8). Hence (4.10) implies (4.9). Now consider the case $\bar{\alpha} = 0$, that is, $\lim_{k \in K, k \to \infty} \alpha_k = 0$. By the assumption that $\nabla \bar{H}(\bar{z})$ is nonsingular, it is clear that when $k \in K$ is sufficiently large, $\nabla \bar{H}(z_k)$ and A_k are uniformly nonsingular. In particular, there exists a positive constant $M_2 > 0$ such that

$\|\nabla \bar{H}(z_k)^{-1}\| \leq M_2$ and $\|A_k^{-1}\| \leq M_2$ hold for all $k \in K$ large enough. This together with (3.3) yields

$$\begin{aligned} \|p_k\| &= \|A_k^{-1}[(A_k - B_k)p_k - H^0(z_k)]\| \\ &\leq \|A_k^{-1}\| \Big(\|(A_k - B_k)p_k\| + \|H^0(z_k)\| \Big) \\ &= \delta_k \|A_k^{-1}\| \, \|p_k\| + \|A_k^{-1}\| \, \|H^0(z_k)\| \\ &\leq M_2 \delta_k \|p_k\| + M_2 \|H^0(z_k)\|. \end{aligned}$$

Since $\delta_k \to 0$ as $k \to \infty$ with $k \in K$, the last inequality implies that there is a constant $M_3 > 0$ such that for all $k \in K$ sufficiently large

$$\|p_k\| \leq M_3 \|H^0(z_k)\|. \tag{4.11}$$

In particular, $\{p_k\}_{k \in K}$ is bounded. Without loss of generality, we assume that $p_k \to \bar{p}$ as $k \to \infty$ with $k \in K$. Again by (3.3) we have

$$\begin{aligned} 0 &= \|B_k p_k + H^0(z_k)\| \\ &\geq \|H^0(z_k) + A_k p_k\| - \|(B_k - A_k)p_k\| \\ &= \|H^0(z_k) + A_k p_k\| - \delta_k \|p_k\|. \end{aligned}$$

Letting $k \in K$ go to infinity yields

$$\nabla \bar{H}(\bar{z})^T \bar{p} + H^0(\bar{z}) = 0. \tag{4.12}$$

On the other hand, the assumption that $\lim_{k \in K, k \to \infty} \alpha_k = 0$ means that when $k \in K$ is sufficiently large, $\alpha_k < 1$. By the line search rule, it means that when k is sufficiently large, $\alpha_k' \equiv \alpha_k / \rho$ does not satisfy (3.2). In other words, we have

$$\begin{aligned} \|\bar{H}(z_k + \alpha_k' p_k)\| - \|\bar{H}(z_k)\| &> -\sigma_1 \|\bar{H}(z_k + \alpha_k' p_k) - \bar{H}(z_k)\| - \sigma_2 \|\alpha' p_k\|^2 + \epsilon_k \\ &\geq -\sigma_1 \|\bar{H}(z_k + \alpha_k' p_k) - \bar{H}(z_k)\| - \sigma_2 \|\alpha' p_k\|^2. \end{aligned}$$

Multiplying the both sides by $\|\bar{H}(z_k + \alpha_k' p_k)\| + \|\bar{H}(z_k)\|$ yields

$$\begin{aligned} &\|\bar{H}(z_k + \alpha_k' p_k)\|^2 - \|\bar{H}(z_k)\|^2 \\ &\geq -\sigma_1 \Big(\|\bar{H}(z_k + \alpha_k' p_k)\| + \|\bar{H}(z_k)\| \Big) \|\bar{H}(z_k + \alpha_k' p_k) - \bar{H}(z_k)\| \\ &\quad - \sigma_2 \|\alpha_k' p_k\|^2 \Big(\|\bar{H}(z_k + \alpha_k' p_k)\| + \|\bar{H}(z_k)\| \Big). \end{aligned}$$

Dividing the both sides by α_k' and then taking the limit as $k \to \infty$ with $k \in K$ yield

$$2 \bar{H}(\bar{z})^T \nabla \bar{H}(\bar{z})^T \bar{p} \geq -2\sigma_1 \|\bar{H}(\bar{z})\| \, \|\nabla \bar{H}(\bar{z}) \bar{p}\|.$$

This together with (4.12) implies

$$\bar{H}(\bar{z})^T H^0(\bar{z}) \le \sigma_1 \|\bar{H}(\bar{z})\| \, \|H^0(\bar{z})\|. \tag{4.13}$$

Since $\bar{\epsilon} \le \gamma \|H^0(z_k)\|$ for all k, we have $\bar{\epsilon} \le \gamma \|H^0(\bar{z})\|$. It then follows from (2.6) that

$$\|\bar{H}(\bar{z}) - H^0(\bar{z})\| = \|H^{\bar{\epsilon}}(\bar{z}) - H^0(\bar{z})\| \le \sqrt{m_1}\bar{\epsilon} \le \sqrt{m_1}\gamma \|H^0(\bar{z})\|, \tag{4.14}$$

and hence

$$\|\bar{H}(\bar{z})\| \le \|H^0(\bar{z})\| + \|\bar{H}(\bar{z}) - H^0(\bar{z})\| \le (1 + \sqrt{m_1}\gamma)\|H^0(\bar{z})\|.$$

Consequently, we get from the last inequality and (4.13)

$$\begin{aligned}
\sigma_1(1 + \sqrt{m_1}\gamma)\|H^0(\bar{z})\|^2 &\ge \sigma_1 \|\bar{H}(\bar{z})\| \, \|H^0(\bar{z})\| \\
&\ge \bar{H}(\bar{z})^T H^0(\bar{z}) \\
&= \|H^0(\bar{z})\|^2 + (\bar{H}(\bar{z}) - H^0(\bar{z}))^T H^0(\bar{z}) \\
&\ge \|H^0(\bar{z})\|^2 - \|\bar{H}(\bar{z}) - H^0(\bar{z})\| \, \|H^0(\bar{z})\| \\
&\ge (1 - \sqrt{m_1}\gamma)\|H^0(\bar{z})\|^2,
\end{aligned}$$

where the last inequality follows from (4.14). However, since $\sigma_1 < \frac{1-\sqrt{m_1}\gamma}{1+\sqrt{m_1}\gamma}$, we get $H^0(\bar{z}) = 0$, which contradicts the fact that $0 < \bar{\epsilon} \le \|H(\bar{z})\|$. The contradiction shows that $\bar{K}$ must be infinite. Therefore by Theorem 4.1, we have $\liminf_{k\to\infty} \|H^0(z_k)\| = 0$.

Now we show that (4.9) holds. Indeed, from (2.8), we get

$$\|H^k(z_k)\| \le \|H^0(z_k)\| + \sqrt{m_1}\epsilon_k \le (1 + \sqrt{m_1}\gamma)\|H^0(z_k)\|.$$

This means that there is a subsequence of $\{H^k(z_k)\}$ converging to zero. However, Lemma 3.1 shows that $\{\|H^k(z_k)\|\}$ itself converges and hence it converges to zero. Again by (2.8) we get

$$\|H^0(z_k)\| \le \|H^k(z_k)\| + \sqrt{m_1}\epsilon_k \le \|H^k(z_k)\| + \sqrt{m_1}\gamma\|H^0(z_k)\|,$$

which implies that

$$\|H^0(z_k)\| \le \frac{1}{1 - \sqrt{m_1}\gamma}\|H^k(z_k)\|.$$

This shows that (4.9) holds. Consequently, every accumulation point of $\{z_k\}$ is a solution of (1.4). □

Remark (i) In Theorem 4.2, the assumption that $\nabla H^k(\bar{z})$ is nonsingular for every accumulation point $\bar{z}$ can be replaced by the weaker assumption that there is an accumulation point $\bar{z}$ of the subsequence $\{z_k\}_{k\in\bar{K}_1}$ such that $\nabla H^k(\bar{z})$ is nonsingular for all k large enough, where $\bar{K}_1$ is the index set as specified in Lemma 4.1.

(ii) Since by Step 5 in Algorithm 1, $\epsilon_k \le \gamma\|H(z_k)\|$ holds for every k and $\{\epsilon_k\}$ is monotonically nonincreasing, Theorem 4.2 shows that

$$\lim_{k\to\infty} \epsilon_k = 0. \tag{4.15}$$

5. Superlinear Convergence

In this section, we prove superlinear convergence of Algorithm 1. In the analysis of superlinear convergence, the concept of strict complementarity plays an important role. We say that the strict complementarity condition holds at a solution $\bar{z} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ of (1.3) if

$$\bar{\lambda}_i + g_i(\bar{x}) > 0, \qquad i = 1, 2, \ldots, m_1. \tag{5.1}$$

If (5.1) holds, then H^0 is differentiable in a neighbourhood of $\bar{z}$ and

$$\nabla H^0(\bar{z}) = \begin{pmatrix} \nabla_x L(\bar{z}) & -\nabla g(\bar{x}) \text{ diag } (b_i^0(\bar{x}, \bar{\lambda})) & -\nabla h(\bar{x}) \\ -\nabla g(\bar{x})^T & -\text{ diag } (a_i^0(\bar{x}, \bar{\lambda})) & 0 \\ -\nabla h(\bar{x})^T & 0 & 0 \end{pmatrix}. \tag{5.2}$$

It is easy to see that $\lim_{\epsilon \to 0} \nabla H^\epsilon(\bar{z}) = \nabla H^0(\bar{z})$.

To establish superlinear convergence of Algorithm 1, we need further assumptions.

Assumption B (i) The sequence $\{z_k\}$ generated by Algorithm 1 converges to a solution of (1.4), say $\bar{z} = (\bar{x}, \bar{\lambda}, \bar{\mu})$, at which the strict complementarity condition holds.

(ii) $\nabla H^0(\bar{z})$ is nonsingular.

Since $\epsilon_k \to 0$ by (4.15), we have $\nabla H^k(z_k) \to \nabla H^0(\bar{z})$ under condition (i) in Assumption B. Therefore, condition (ii) in Assumption B implies that $\nabla H^k(z)$ are uniformly nonsingular in a neighborhood of $\bar{z}$ for all k sufficiently large. This implies that there is a neighborhood $U(\bar{z})$ of $\bar{z}$ and positive constants c_1', C_1', c_2' and C_2', such that

$$c_1' \le \|\nabla H^0(z)\| \le C_1', \quad c_2' \le \|\nabla H^0(z)^{-1}\| \le C_2', \ \ \forall z \in U(\bar{z}) \tag{5.3}$$

and

$$c_1' \le \|\nabla H^k(z)\| \le C_1', \quad c_2' \le \|\nabla H^k(z)^{-1}\| \le C_2', \ \ \forall z \in U(\bar{z}) \tag{5.4}$$

for all k large enough. Moreover, there are positive constants c' and C' such that for k large enough

$$\begin{cases} c'\|p\| \le \|\nabla H^k(z)p\| \le C'\|p\|, & \forall z \in U(\bar{z}),\ \forall p \in R^{n+m+r} \\ c'\|p\| \le \|\nabla H^k(z)^{-1}p\| \le C'\|p\|, & \forall z \in U(\bar{z}),\ \forall p \in R^{n+m+r} \end{cases} \tag{5.5}$$

and

$$\begin{cases} c'\|p\| \le \|\nabla H^0(z)p\| \le C'\|p\|, & \forall z \in U(\bar{z}),\ \forall p \in R^{n+m+r} \\ c'\|p\| \le \|\nabla H^0(z)^{-1}p\| \le C'\|p\|, & \forall z \in U(\bar{z}),\ \forall p \in R^{n+m+r}. \end{cases} \tag{5.6}$$

In particular, we have for all k sufficiently large

$$c'\|z - z'\| \le \|H^k(z) - H^k(z')\| \le C'\|z - z'\|, \qquad \forall z, z' \in U(\bar{z}) \tag{5.7}$$

and

$$c'\|z-z'\| \le \|H^0(z)-H^0(z')\| \le C'\|z-z'\|, \qquad \forall z,z' \in U(\bar{z}). \tag{5.8}$$

In the following, we show some useful lemmas. They will play an important role in the proof of superlinear convergence of Algorithm 1.

Lemma 5.1 *Let $\bar{z}=(\bar{x},\bar{\lambda},\bar{\mu})$ be a point (not necessarily a KKT point of VI(X,F)) such that for every $i=1,2,\ldots,m_1$, $\bar{\lambda}_i$ and $g_i(\bar{x})$ do note vanish at the same time, i.e.,*

$$\bar{\lambda}_i^2 + g_i(\bar{x})^2 > 0, \qquad i=1,2,\ldots,m_1. \tag{5.9}$$

Then there is a neighborhood $U(\bar{z})$ of $\bar{z}$ and a constant $M_4>0$ such that for any $z,z' \in U(\bar{z})$ and every $\epsilon>0$

$$\|(H^0(z)-H^\epsilon(z))-(H^0(z')-H^\epsilon(z'))\| \le M_4\epsilon^2\|z-z'\|. \tag{5.10}$$

Proof We verify (5.10) by showing that there is a neighborhood $U(\bar{z})$ of $\bar{z}$ such that for every $i=1,2,\ldots,n+m_1+m_2$, the inequality

$$\|(H_i^0(z)-H_i^\epsilon(z))-(H_i^0(z')-H_i^\epsilon(z'))\| \le M_4'\epsilon^2\|z-z'\|$$

holds for all $z,z' \in U(\bar{z})$ and every $\epsilon>0$ with some positive constant M_4'. In view of the expression of H^ϵ, it then suffices to show that there is a neighborhood $U(\bar{z})$ of $\bar{z}$ and a positive constant M_5 such that for every $i=1,2,\ldots,m_1$, any $\epsilon>0$, all $z,z' \in U(\bar{z})$, and any $\epsilon>0$

$$\|(\phi_i^0(z)-\phi_i^\epsilon(z))-(\phi_i^0(z')-\phi_i^\epsilon(z'))\| \le M_5\epsilon^2\|z-z'\|. \tag{5.11}$$

By condition (5.9), it is clear that there is a neighborhood $U(\bar{z})$ of $\bar{z}$ such that for all $z=(x,\lambda,\mu) \in U(\bar{z})$ and every $i=1,2,\ldots,m_1$

$$\sqrt{\lambda_i^2+g_i(x)^2} \ge \frac{1}{2}\min_{1\le j\le m_1}\sqrt{\bar{\lambda}_j^2+g_j(\bar{x})^2} \triangleq \bar{\theta} > 0.$$

Therefore, it follows from the definition of ϕ^ϵ that for every $i=1,2,\ldots,m_1$ and all $z,z' \in U(\bar{z})$

$$\begin{aligned}
&\left|(\phi_i^0(z)-\phi_i^\epsilon(z))-(\phi_i^0(z')-\phi_i^\epsilon(z'))\right| \\
&= \left|\left(\sqrt{\lambda_i^2+g_i(x)^2}-\sqrt{\lambda_i^2+g_i(x)^2+\epsilon^2}\right)-\left(\sqrt{(\lambda_i')^2+g_i(x')^2}-\sqrt{(\lambda_i')^2+g_i(x')^2+\epsilon^2}\right)\right| \\
&= \left|\frac{-\epsilon^2}{\sqrt{\lambda_i^2+g_i(x)^2}+\sqrt{\lambda_i^2+g_i(x)^2+\epsilon^2}}-\frac{-\epsilon^2}{\sqrt{(\lambda_i')^2+g_i(x')^2}+\sqrt{(\lambda_i')^2+g_i(x')^2+\epsilon^2}}\right| \\
&= \epsilon^2\frac{\left|\left(\sqrt{(\lambda_i')^2+g_i(x')^2}+\sqrt{(\lambda_i')^2+g_i(x')^2+\epsilon^2}\right)-\left(\sqrt{\lambda_i^2+g_i(x)^2}+\sqrt{\lambda_i^2+g_i(x)^2+\epsilon^2}\right)\right|}{\left(\sqrt{\lambda_i^2+g_i(x)^2}+\sqrt{\lambda_i^2+g_i(x)^2+\epsilon^2}\right)\left(\sqrt{(\lambda_i')^2+g_i(x')^2}+\sqrt{(\lambda_i')^2+g_i(x')^2+\epsilon^2}\right)}
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{\epsilon^2}{4\bar{\theta}^2}\Big|\Big(\sqrt{(\lambda_i')^2+g_i(x')^2}+\sqrt{(\lambda_i')^2+g_i(x')^2+\epsilon^2}\Big)-\Big(\sqrt{\lambda_i^2+g_i(x)^2}+\sqrt{\lambda_i^2+g_i(x)^2+\epsilon^2}\Big)\Big| \\
&\leq \frac{\epsilon^2}{4\bar{\theta}^2}\Big(\Big|\sqrt{(\lambda_i')^2+g_i(x')^2}-\sqrt{\lambda_i^2+g_i(x)^2}\Big|+\Big|\sqrt{(\lambda_i')^2+g_i(x')^2+\epsilon^2}-\sqrt{\lambda_i^2+g_i(x)^2+\epsilon^2}\Big|\Big) \\
&= \frac{\epsilon^2}{4\bar{\theta}^2}\Big\{\frac{|[(\lambda_i')^2-\lambda_i^2]+[g_i(x')^2-g_i(x)^2]|}{\sqrt{(\lambda_i')^2+g_i(x')^2}+\sqrt{\lambda_i^2+g_i(x)^2}}+\frac{|[(\lambda_i')^2-\lambda_i^2]+[g_i(x')^2-g_i(x)^2]|}{\sqrt{(\lambda_i')^2+g_i(x')^2+\epsilon^2}+\sqrt{\lambda_i^2+g_i(x)^2+\epsilon^2}}\Big\} \\
&\leq \frac{\epsilon^2}{4\bar{\theta}^3}\{|(\lambda_i')^2-\lambda_i^2|+|g_i(x')^2-g_i(x)^2|\}.
\end{aligned}
$$

Therefore, by the continuous differentiability of g, we get (5.11) for every $i=1,2,\ldots,m_1$ and any $z,z'\in U(\bar{z})$, which in turn implies (5.10). □

The following lemma shows that ∇H^ϵ as a function of ϵ is Lipschitz continuous.

Lemma 5.2 *Suppose that $\bar{z}=(\bar{x},\bar{\lambda},\bar{\mu})$ is a point (not necessarily a KKT point of VI(X,F)) such that condition (5.9) holds. Then there is a neighborhood $U(\bar{z})$ of $\bar{z}$ and a constant $C>0$ such that for any $z\in U(\bar{z})$ and $\epsilon'>\epsilon\geq 0$*

$$
\|\nabla H^\epsilon(z)-\nabla H^{\epsilon'}(z)\|\leq C(\epsilon'-\epsilon). \tag{5.12}
$$

Proof First, we point out that if condition (5.9) holds, then H^0 is differentiable in a neighborhood of $\bar{z}$. Moreover, we have

$$
\nabla H^0(z)=\begin{pmatrix} \nabla_x L(z) & -\nabla g(x)\ \mathrm{diag}\ (b_i^0) & -\nabla h(x) \\ -\nabla g(x)^T & -\ \mathrm{diag}\ (a_i^0) & 0 \\ -\nabla h(x)^T & 0 & 0 \end{pmatrix}.
$$

This means that ∇H^0 and ∇H^ϵ $(\epsilon>0)$ have the same block structure. To verify (5.12), it thus suffices to show that there is a neighborhood $U(\bar{z})$ of $\bar{z}$ and constants $C_1,C_2>0$ such that for any $z\in U(\bar{z})$

$$
|b_i^\epsilon-b_i^{\epsilon'}|\leq C_1(\epsilon'-\epsilon),\quad i=1,2,\ldots,m_1 \tag{5.13}
$$

and

$$
|a_i^\epsilon-a_i^{\epsilon'}|\leq C_2(\epsilon'-\epsilon),\quad i=1,2,\ldots,m_1. \tag{5.14}
$$

By (5.9), it is clear that there is a neighborhood $U(\bar{z})$ of $\bar{z}$ such that for every $i=1,2,\ldots,m_1$

$$
\sqrt{\lambda_i^2+g_i(x)^2}\geq\frac{1}{2}\min_{1\leq j\leq m_1}\{\sqrt{\lambda_i^2+g_j(\bar{x})^2}\}\triangleq\bar{\theta}>0,\quad\forall z=(x,\lambda,\mu)\in U(\bar{z}).
$$

So, it follows from the definition of b_i^ϵ that for any $z\in U(\bar{z})$ and every $i=1,2,\ldots,m_1$

$$
|b_i^\epsilon-b_i^{\epsilon'}| = \Big|\Big[\frac{g_i(x)}{\sqrt{\lambda_i^2+g_i(x)^2+\epsilon^2}}-1\Big]-\Big[\frac{g_i(x)}{\sqrt{\lambda_i^2+g_i(x)^2+(\epsilon')^2}}-1\Big]\Big|
$$

$$
\begin{aligned}
&= \Big|\frac{1}{\sqrt{\lambda_i^2+g_i(x)^2+\epsilon^2}} - \frac{1}{\sqrt{\lambda_i^2+g_i(x)^2+(\epsilon')^2}}\Big|\,|g_i(x)| \\
&= \frac{\Big|\sqrt{\lambda_i^2+g_i(x)^2+\epsilon^2} - \sqrt{\lambda_i^2+g_i(x)^2+(\epsilon')^2}\Big|}{\sqrt{\lambda_i^2+g_i(x)^2+\epsilon^2}\sqrt{\lambda_i^2+g_i(x)^2+(\epsilon')^2}}\,|g_i(x)| \\
&\le \frac{1}{\bar\theta^2}\Big|\sqrt{\lambda_i^2+g_i(x)^2+\epsilon^2} - \sqrt{\lambda_i^2+g_i(x)^2+(\epsilon')^2}\Big|\,|g_i(x)| \\
&= \frac{1}{\bar\theta^2}\frac{|\epsilon^2-(\epsilon')^2|}{\Big(\sqrt{\lambda_i^2+g_i(x)^2+\epsilon^2}+\sqrt{\lambda_i^2+g_i(x)^2+(\epsilon')^2}\Big)}|g_i(x)| \\
&\le \frac{1}{\bar\theta^2}\frac{((\epsilon')^2-\epsilon^2)}{\epsilon'+\epsilon}|g_i(x)| \\
&= \frac{1}{\bar\theta^2}|g_i(x)|(\epsilon'-\epsilon).
\end{aligned}
$$

This implies (5.13). The inequality (5.14) can be proved in a similar way. Therefore we get (5.12). □

The following lemma shows that $H^k(\bar z)$ converges to $H^0(\bar z)$ rapidly.

Lemma 5.3 *Let $\bar z$ be a solution of the KKT system (1.3). Then, under condition (i) in Assumption B, there is a constant $C_3>0$ such that for every k*

$$
\|H^k(\bar z) - H^0(\bar z)\| \le C_3\|z_k-\bar z\|^2. \tag{5.15}
$$

Proof By the definition of ϕ^ϵ, we have

$$
\begin{aligned}
|\phi_i^k(\bar z) - \phi_i^0(\bar z)| &= \Big|\sqrt{\bar\lambda_i^2+g_i(\bar x)^2+\epsilon_k^2} - \sqrt{\bar\lambda_i^2+g_i(\bar x)^2}\Big| \\
&= \frac{\epsilon_k^2}{\sqrt{\bar\lambda_i^2+g_i(\bar x)^2+\epsilon_k^2}+\sqrt{\bar\lambda_i^2+g_i(\bar x)^2}} \\
&\le \frac{\epsilon_k^2}{2\sqrt{\bar\lambda_i^2+g_i(\bar x)^2}}.
\end{aligned}
$$

This implies that

$$
\|\Phi^k(\bar z) - \Phi^0(\bar z)\| \le \frac{\sqrt{m_1}\epsilon_k^2}{2\sqrt{\bar\lambda_i^2+g_i(\bar x)^2}}.
$$

Letting $C_4 = \frac{\sqrt{m_1}}{2\sqrt{\bar\lambda_i^2+g_i(\bar x)^2}}$, we get

$$
\|H^k(\bar z) - H^0(\bar z)\| = \|\Phi^k(\bar z) - \Phi^0(\bar z)\| \le C_4\epsilon_k^2. \tag{5.16}
$$

On the other hand, by Step 5 in Algorithm 1, we have

$$
\epsilon_k \le \gamma\|H^0(z_k)\| = \gamma\|H^0(z_k) - H^0(\bar z)\| \le C'\gamma\|z_k-\bar z\|,
$$

where the inequality follows from (5.7). Putting $C_3 = C_4(C')^2\gamma^2$ yields (5.15). □

In a way similar to the proof of Lemma 3.3, we can show that under Assumptions A and B, ∇H^0 and ∇H^k for all k sufficiently large are Lipschitz continuous in a neighbourhood of $\bar{z}$. Moreover, the Lipschitz constant is independent of k. So, we have the following useful lemma.

Lemma 5.4 *Let Assumptions A and B hold and δ_k be defined by (4.4). Then we have*

$$\lim_{k\to\infty} \delta_k = 0. \tag{5.17}$$

Proof By Lemma 3.1, we have

$$\sum_{k=1}^{\infty} \|s_k\| = \sum_{k=1}^{\infty} \|z_{k+1} - z_k\| < \infty. \tag{5.18}$$

Define

$$\bar{A}_k = \int_0^1 \nabla H^k(z_k + ts_k)dt. \tag{5.19}$$

Then we have $y_k = \bar{A}_k s_k$. Moreover, it follows from (3.4) that

$$\begin{aligned}
\|B_{k+1} - \bar{A}_k\|_F^2 &= \Big\|B_k - \bar{A}_k + \theta_k(\bar{A}_k - B_k)\frac{s_k s_k^T}{\|s_k\|^2}\Big\|_F^2 \\
&= \|B_k - \bar{A}_k\|_F^2 + (\theta_k^2 - 2\theta_k)\frac{\|(B_k - \bar{A}_k)s_k\|^2}{\|s_k\|^2} \\
&= \|B_k - \bar{A}_k\|_F^2 + (\theta_k^2 - 2\theta_k)\frac{\|y_k - B_k s_k\|^2}{\|s_k\|^2} \\
&= \|B_k - \bar{A}_k\|_F^2 + (\theta_k^2 - 2\theta_k)\delta_k^2 \\
&\le \|B_k - \bar{A}_k\|_F^2 - (1 - \bar{\theta}^2)\delta_k^2,
\end{aligned} \tag{5.20}$$

where the last inequality holds since θ_k is chosen to satisfy $|1 - \theta_k| \le \bar{\theta}$ in Step 4. However, from Lemma 5.2, there is a constant $C_5 > 0$ such that for all k sufficiently large

$$\|\nabla H^k(z_{k-1} + ts_{k-1}) - \nabla H^{k-1}(z_{k-1} + ts_{k-1})\| \le C_5(\epsilon_{k-1} - \epsilon_k).$$

Let τ_3 be the Lipschitz constant of ∇H^k. Note that τ_3 is independent of k. Then we deduce

$$\begin{aligned}
\|\bar{A}_k - \bar{A}_{k-1}\| &\le \int_0^1 \|\nabla H^k(z_k + ts_k) - \nabla H^{k-1}(z_{k-1} + ts_{k-1})\|dt \\
&\le \int_0^1 \|\nabla H^k(z_k + ts_k) - \nabla H^k(z_{k-1} + ts_{k-1})\|dt \\
&\quad + \int_0^1 \|\nabla H^k(z_{k-1} + ts_{k-1}) - \nabla H^{k-1}(z_{k-1} + ts_{k-1})\|dt \\
&\le \tau_3 \int_0^1 \|(z_k + ts_k) - (z_{k-1} + ts_{k-1})\|dt + C_5(\epsilon_{k-1} - \epsilon_k) \\
&= \tau_3 \int_0^1 \|ts_k + (1-t)s_{k-1}\|dt + C_5(\epsilon_{k-1} - \epsilon_k) \\
&\le \tau_3\Big(\|s_k\| + \|s_{k-1}\|\Big) + C_5(\epsilon_{k-1} - \epsilon_k).
\end{aligned}$$

This together with (5.18) implies

$$\sum_{k=1}^{\infty} \|\bar{A}_k - \bar{A}_{k-1}\| < \infty. \tag{5.21}$$

Therefore, we get from (5.20)

$$\|B_{k+1} - \bar{A}_k\|_F \leq \|B_k - \bar{A}_k\|_F \leq \|B_k - \bar{A}_{k-1}\|_F + \|\bar{A}_k - \bar{A}_{k-1}\|_F.$$

This along with (5.21) implies that $\{\|B_{k+1} - \bar{A}_k\|_F\}$ converges. Thus we deduce from (5.20) again that (5.17) holds. □

Lemma 5.5 *Let $\bar{z}$ be a solution of the KKT system (1.3). Then under Assumptions A and B, we have*

$$\lim_{k\to\infty} \frac{\|z_k + p_k - \bar{z}\|}{\|z_k - \bar{z}\|} = 0. \tag{5.22}$$

Proof Let $\bar{A}_k$ be defined by (5.19). Then it is obvious that $\bar{A}_k - \nabla H^k(z_k) \to 0$. Moreover, since (5.17) holds, using the arguments for showing (4.11) in the proof of Theorem 4.2, we can deduce that there is a positive constant M_6 such that

$$\|p_k\| \leq M_6 \|H^0(z_k)\| \leq C' M_6 \|z_k - \bar{z}\|, \tag{5.23}$$

where the last inequality follows from (5.7). We then get from (3.3)

$$\begin{aligned}
&\|\nabla H^k(z_k)(z_k + p_k - \bar{z})\| \\
&= \|\nabla H^k(z_k)(z_k - \bar{z}) + \nabla H^k(z_k)p_k - B_k p_k - H^0(z_k)\| \\
&\leq \|(\nabla H^k(z_k) - B_k)p_k\| + \|\nabla H^k(z_k)(z_k - \bar{z}) + H^0(\bar{z}) - H^0(z_k)\| \\
&\leq \|(\bar{A}_k - B_k)p_k\| + \|(\nabla H^k(z_k) - \bar{A}_k)p_k\| + \|\nabla H^k(z_k)(z_k - \bar{z}) + H^0(\bar{z}) - H^0(z_k)\| \\
&\leq \delta_k \|p_k\| + \|\nabla H^k(z_k) - \bar{A}_k\| \, \|p_k\| + \|\nabla H^k(z_k)(z_k - \bar{z}) + H^k(\bar{z}) - H^k(z_k)\| \\
&\quad + \|[H^0(\bar{z}) - H^0(z_k)] - [H^k(\bar{z}) - H^k(z_k)]\| \\
&\leq C' M_6 \Big(\delta_k + \|\nabla H^k(z_k) - \bar{A}_k\|\Big) \|z_k - \bar{z}\| + M_4 \epsilon_k^2 \|z_k - \bar{z}\| + o(\|z_k - \bar{z}\|) \\
&= o(\|z_k - \bar{z}\|),
\end{aligned}$$

where the third inequality follows from (4.4) and $y_k = \bar{A}_k s_k$, the last inequality follows from (5.10) and (5.23), and the last equality follows from (5.17). This together with (5.5) implies (5.22). □

Now, we establish superlinear convergence of Algorithm 1. By Lemma 5.5, it suffices to show that when k is sufficiently large, the unit steplength is always accepted.

Theorem 5.1 *Under Assumptions A and B, $\{z_k\}$ converges superlinearly.*

Proof By Lemma 5.5, it suffices to show that when k is sufficiently large, $\alpha = 1$ satisfies (3.2). Indeed, it follows from (5.7), (5.15) and (5.22) that for all k large enough

$$\begin{aligned}\|H^k(z_k+p_k)\| &= \|H^k(z_k+p_k)-H^0(\bar{z})\| \\ &\le \|H^k(z_k+p_k)-H^k(\bar{z})\|+\|H^k(\bar{z})-H^0(\bar{z})\| \\ &\le C'\|z_k+p_k-\bar{z}\|+C_3\|z_k-\bar{z}\|^2 \\ &= o(\|z_k-\bar{z}\|). \end{aligned} \tag{5.24}$$

On the other hand, we have

$$\begin{aligned}&\|H^k(z_k)\|-\sigma_1\|H^k(z_k+p_k)-H^k(z_k)\|-\sigma_2\|p_k\|^2 \\ &= \|H^k(z_k)-H^0(\bar{z})\|-\sigma_1\|H^k(z_k+p_k)-H^k(z_k)\|+o(\|z_k-\bar{z}\|) \\ &\ge \Big(\|H^k(z_k)-H^k(\bar{z})\|-\|H^k(\bar{z})-H^0(\bar{z})\|\Big) \\ &\quad -\sigma_1\Big(\|H^k(z_k+p_k)-H^k(\bar{z})\|+\|H^k(\bar{z})-H^k(z_k)\|\Big)+o(\|z_k-\bar{z}\|) \\ &= (1-\sigma_1)\|H^k(z_k)-H^k(\bar{z})\|+o(\|z_k-\bar{z}\|) \\ &\ge (1-\sigma_1)c'\|z_k-\bar{z}\|+o(\|z_k-\bar{z}\|), \end{aligned} \tag{5.25}$$

where the first equality follows from (5.23), the last equality follows from (5.22) and (5.15), and the last inequality follows from (5.7). Comparing (5.24) with (5.25), it is clear that when k is sufficiently large

$$\begin{aligned}\|H^k(z_k+p_k)\| &\le \|H^k(z_k)\|-\sigma_1\|H^k(z_k+p_k)-H^k(z_k)\|-\sigma_2\|p_k\|^2 \\ &\le \|H^k(z_k)\|-\sigma_1\|H^k(z_k+p_k)-H^k(z_k)\|-\sigma_2\|p_k\|^2+\eta_k. \end{aligned}$$

This means that when k is sufficiently large, $\alpha_k = 1$ is always accepted. □

6. Broyden-like method for Nonlinear Complementarity Problems

In this section, we focus on the NCP (1.2). For problem (1.2), the KKT system takes the form

$$\begin{cases} F(x)-\lambda = 0, \\ \min\{x,\lambda\} = 0, \end{cases} \tag{6.1}$$

which coincides with (1.4) if we put $g_i(x) = x_i$, $i = 1,2,\ldots,n$ with $m_1 = n$ and $h_j(x) \equiv 0$, $j = 1,2,\ldots,m_2$. Of course, Algorithm 1 can be used to solve the KKT system (6.1). However, by doing so, the dimension of the subproblem (3.3) becomes $2n$, which is twice as many as in (1.2). In this section, we propose a Broyden-like method for solving (1.2) that retains global and

superlinear convergence properties without increasing the size of the subproblem. Moreover, as will be shown later in this section, conditions to ensure global convergence are weaker. Indeed, we will prove that the proposed Broyden-like method converges if F is a P_0 function with Lipschitz continuous Jacobian.

We notice that the first equation in (6.1) implies that $\lambda = F(x)$. So, it is natural to consider an algorithm similar to Algorithm 1 by using $F(x)$ instead of λ. In the following development, we put this idea into practice. For $\epsilon \geq 0$, define $\Psi^\epsilon(x) = (\psi_1^\epsilon(x), \psi_2^\epsilon(x), \ldots, \psi_n^\epsilon(x))^T$ with elements

$$\psi_i^\epsilon(x) = \sqrt{F_i(x)^2 + x_i^2 + \epsilon^2} - (F_i(x) + x_i), \quad i = 1, 2, \ldots, n. \tag{6.2}$$

Then the NCP (1.2) is equivalent to the nonsmooth equation

$$\Psi^0(x) = 0. \tag{6.3}$$

For $\epsilon \neq 0$, the smooth equation

$$\Psi^\epsilon(x) = 0 \tag{6.4}$$

can be regarded as an approximation to (6.3). The relationship between Ψ^ϵ and Φ^ϵ is obvious: If we let $g_i(x) = x_i$ and $\lambda_i = F_i(x)$, $i = 1, 2, \ldots, n = m_1$ in (2.1), then Φ^ϵ coincides with Ψ^ϵ. Therefore, Ψ^ϵ possesses many properties similar to those of Φ^ϵ as described in Section 2. In particular, Ψ^ϵ is continuously differentiable for every $\epsilon \neq 0$ provided that F is continuously differentiable. Moreover, by direct calculation, we have

$$\nabla \psi_i^\epsilon(x) = \bar{a}_i^\epsilon(x) \nabla F_i(x) + \bar{b}_i^\epsilon(x) e_i,$$

where e_i denotes the i-th unit vector in R^n and

$$\bar{a}_i^\epsilon(x) = \frac{F_i(x)}{\sqrt{F_i(x)^2 + x_i^2 + \epsilon^2}} - 1, \qquad \bar{b}_i^\epsilon(x) = \frac{x_i}{\sqrt{F_i(x)^2 + x_i^2 + \epsilon^2}} - 1.$$

It then follows that for every $\epsilon \neq 0$ and every $x \in R^n$

$$\nabla \Psi^\epsilon(x) = \nabla F(x) \text{ diag } (\bar{a}_i^\epsilon(x)) + \text{ diag } (\bar{b}_i^\epsilon(x)).$$

Moreover, it is not difficult to prove the following set of lemmas corresponding to some lemmas in Section 2. We omit the proof of the first two lemmas.

Lemma 6.1 *For any $\epsilon' > \epsilon \geq 0$, we have*

$$0 < \psi^{\epsilon'}(x) - \psi^\epsilon(x) \leq \epsilon' - \epsilon, \qquad \forall x \in R^n.$$

Consequently, we have for any $\epsilon' > \epsilon \geq 0$

$$\|\Psi^\epsilon(x) - \Psi^{\epsilon'}(x)\| \leq \sqrt{n}(\epsilon' - \epsilon), \quad \forall x \in R^n.$$

Lemma 6.2 *We have for every $\epsilon \neq 0$ and any $x \in R^n$, $-2 < \bar{a}_i^\epsilon(x) < 0$ and $-2 < \bar{b}_i^\epsilon(x) < 0$ for all $i = 1, 2, \ldots, n$.*

Lemma 6.3 *Let $\epsilon \neq 0$. Suppose that F is continuously differentiable. If the matrix*

$$G^\epsilon(x) \stackrel{\triangle}{=} \nabla F(x) + \mathrm{diag}(\bar{b}_i^\epsilon(x)\bar{a}_i^\epsilon(x)^{-1})$$

is nonsingular, then $\nabla \Psi^\epsilon(x)$ is nonsingular. In particular, if F is a P_0 function, then $\nabla \Psi^\epsilon(x)$ is nonsingular for every $\epsilon \neq 0$ and any $x \in R^n$.

Proof It is easy to see that for every $\epsilon \neq 0$ and any $x \in R^n$

$$\nabla \Psi^\epsilon(x) = \ \mathrm{diag}\ (\bar{a}_i^k(x))\Big(\ \mathrm{diag}\ (\bar{b}_i^k(x)\bar{a}_i^k(x)^{-1}) + \nabla F(x)\Big) = \ \mathrm{diag}\ (\bar{a}_i^k(x))G^\epsilon(x).$$

This shows that $\nabla \Psi^\epsilon(x)$ is nonsingular if and only if $G^\epsilon(x)$ is nonsingular, which proves the first part of the lemma. If F is a P_0 function, then $\nabla F(x)$ is a P_0 matrix for any $x \in R^n$. Since $\mathrm{diag}(\bar{b}_i^\epsilon(x))$ and $\mathrm{diag}(\bar{a}_i^\epsilon(x))$ are negative diagonal matrices, by Proposition 3.1 in [24], $G^\epsilon(x)$ is nonsingular, and hence $\nabla \Psi_i^\epsilon(x)$ is nonsingular. □

In Lemma 2.4, it is shown that a sufficient condition to ensure the nonsingularity of $\nabla H^\epsilon(z)$ given by (2.9) is that the matrix

$$\nabla_x L(z) + \nabla g(x)\mathrm{diag}(a_i^\epsilon(x)^{-1}b_i^\epsilon(x))\nabla g(x)^T$$

is positive definite on the subspace $\{p \,|\, \nabla h(x)^T p = 0\}$. For the NCP (1.2), this is equivalent to saying that the matrix $G^\epsilon(x)$ is positive definite, since in this case $g(x) \equiv x$, $h(x) \equiv 0$ and hence $\nabla g(x)$ is the unit matrix and $\{p \,|\, \nabla h(x)^T p = 0\} = R^n$. This condition is obviously stronger than that $G^\epsilon(x)$ is nonsingular.

The next two lemmas correspond to Lemmas 5.1 and 5.2, respectively. We omit the proofs.

Lemma 6.4 *Let $\bar{x}$ be a point (not necessarily a solution of the NCP (1.2)) satisfying*

$$\bar{x}_i^2 + F_i(\bar{x})^2 > 0, \qquad i = 1, 2, \ldots, n. \tag{6.5}$$

Then there is a neighborhood $U(\bar{x})$ of $\bar{x}$ and a constant $M > 0$ such that for any $x, x' \in U(\bar{x})$

$$\|(\Psi^0(x) - \Psi^\epsilon(x)) - (\Psi^0(x') - \Psi^\epsilon(x'))\| \leq M\epsilon^2\|x - x'\|.$$

Lemma 6.5 *Suppose that $\bar{x} \in R^n$ is a point (not necessarily a solution of the NCP (1.2)) such that (6.5) holds. Then there is a neighborhood $U(\bar{x})$ of $\bar{x}$ and a constant $C > 0$ such that for any $x \in U(\bar{x})$ and $\epsilon' > \epsilon \geq 0$*

$$\|\nabla \Psi^\epsilon(x) - \nabla \Psi^{\epsilon'}(x)\| \leq C(\epsilon' - \epsilon).$$

Corresponding to Assumption A, we make the following assumption.

Assumption A′ (i) The level set

$$\Omega = \{x \in R^n \mid \|\Psi^0(x)\| \le \frac{1+\sqrt{n}\gamma}{1-\sqrt{n}\gamma}\|\Psi^0(x_1)\| + \frac{1}{1-\sqrt{n}\gamma}\eta\}$$

is contained in a bounded convex set $\bar{D}$, where η is a positive constant such that (3.1) holds for a given positive sequence $\{\eta_k\}$.

(ii) The function F is a continuously differentiable P_0 function on $\bar{D}$.

(ii) ∇F is Lipschitz continuous on $\bar{D}$.

Under Assumption A′, we can establish the following lemma in a way similar to Lemma 3.3.

Lemma 6.6 *Let Assumption A′ hold. For $\theta > 0$ and $\epsilon > 0$, define the set*

$$\bar{\Pi}^\epsilon(\theta) = \{x \in R^n \mid \sqrt{F_i(x)^2 + x_i^2 + \epsilon^2} \ge \theta, i = 1, 2, \ldots, n\} \cap \bar{D}.$$

Then there is a constant $\tau_2 > 0$ such that

$$\|\nabla\Psi^\epsilon(x) - \nabla\Psi^\epsilon(x')\| \le \tau_2\|x - x'\|, \qquad \forall x, x' \in \bar{\Pi}^\epsilon(\theta).$$

As in the previous section, for a sequence $\{\epsilon_k\}$, we will abbreviate Ψ^{ϵ_k} as Ψ^k. Now we state an algorithm for solving the NCP (1.2).

Algorithm 2

Step 0. Choose constants $\rho, \beta \in (0,1)$, $0 < \gamma < \frac{1}{\sqrt{n}}$, $\sigma_1 \in (0, \frac{1-\sqrt{n}\gamma}{1+\sqrt{n}\gamma})$, $\sigma_2 > 0$, $\bar{\theta} \in (0,1)$. Select a positive sequence $\{\eta_k\}$ satisfying (3.1). Choose an initial point $x_1 \in R^n$ and an initial nonsingular matrix $B_1 \in R^{n\times n}$. Choose a positive constant ϵ_1 satisfying $\epsilon_1 \le \frac{\gamma}{2}\|H^0(x_1)\|$. Let $k := 1$.

Step 1. If $H^0(x_k) = 0$, stop. Otherwise, solve the linear equation

$$B_k p + \Psi^0(x_k) = 0$$

to get p_k.

Step 2. Let α_k be the maximum element in the set $\{1, \rho, \rho^2, \ldots\}$ such that $\alpha = \rho^i$ satisfies

$$\begin{aligned}&\|\Psi^k(x_k + \alpha p_k)\| - \|\Psi^k(x_k)\| \\ &\quad\le \quad -\sigma_1\|\Psi^k(x_k + \alpha p_k) - \Psi^k(x_k)\| - \sigma_2\|\alpha p_k\|^2 + \eta_k.\end{aligned}$$

Step 3. Let $x_{k+1} = x_k + \alpha_k p_k$.

Step 4. Update B_k by the Broyden-like update formula

$$B_{k+1} = B_k + \theta_k\frac{(y_k - B_k s_k)s_k^T}{\|s_k\|^2},$$

where $s_k = x_{k+1} - x_k$ and $y_k = \Psi^k(x_{k+1}) - \Psi^k(x_k)$. The parameter θ_k is chosen so that $|\theta_k - 1| \le \bar{\theta}$ and B_{k+1} is nonsingular.

Step 5. Update ϵ_{k+1} by the following rule: If $\epsilon_k < \gamma\|\Psi^0(x_{k+1})\|$, set $\epsilon_{k+1} = \epsilon_k$. Otherwise, let ϵ_{k+1} satisfy

$$0 < \epsilon_{k+1} \le \min\{\frac{\gamma}{2}\|\Psi^0(x_{k+1})\|, \frac{1}{2}\epsilon_k\}.$$

Step 6. Let $k := k + 1$. Go to Step 1.

Using the properties of Ψ^ϵ shown in the first part of this section, we can establish global convergence of Algorithm 2 in the next theorem. The proof is omitted, but it is noted that, by Lemma 6.3, condition (ii) in Assumption A$'$ ensures the nonsingularity of $\nabla\Psi^k(x)$ for all k and any x.

Theorem 6.1 *Let Assumption A$'$ hold and $\{x_k\}$ be generated by Algorithm 2. Then*

$$\lim_{k\to\infty} \|\Psi^0(x_k)\| = 0.$$

Moreover, every accumulation point of $\{x_k\}$ is a solution of the NCP (1.2).

To obtain superlinear convergence of Algorithm 2, we need an assumption corresponding to Assumption B. First, we introduce some notations. Let E denote the $n \times n$ unit matrix. For a matrix $A = (a_{ij}) \in R^{n\times n}$ and index sets $I, J \subset \{1, 2, \ldots, n\}$, A_{IJ} denotes the submatrix of A with elements $a_{ij}, i \in I$ and $j \in J$, and A_I denotes the submatrix of A with elements $a_{ij}, i \in I$ and $j \in \{1, 2, \ldots, n\}$. For a point $x \in R^n$, we define the index sets

$$I(x) = \{i \,|\, x_i > F_i(x)\}, \qquad J(x) = \{i \,|\, x_i < F_i(x)\}.$$

Now, we make the following assumption corresponding to Assumption B.

Assumption B$'$ (i) The sequence $\{x_k\}$ generated by Algorithm 2 converges to a solution $\bar{x}$ of the NCP (1.2) satisfying the strict complementarity condition

$$\bar{x}_i + F_i(\bar{x}) > 0, \qquad i = 1, 2, \ldots, n.$$

(ii) $\nabla F_{\bar{I}\bar{I}}(\bar{x})$ is nonsingular, where $\bar{I} = I(\bar{x})$.

Remark Condition (i) in Assumption B$'$ implies that there is a neighborhood $U(\bar{x})$ of $\bar{x}$ such that Ψ^0 is differentiable on $U(\bar{x})$. Moreover, by direct calculation, we have

$$\nabla\Psi^0(\bar{x}) = (-\nabla F_{\bar{I}}(\bar{x}), E_{\bar{J}}) = \begin{pmatrix} -\nabla F_{\bar{I}\bar{I}}(\bar{x}) & 0 \\ -\nabla F_{\bar{I}\bar{J}}(\bar{x}) & E_{|\bar{J}|} \end{pmatrix},$$

where $\bar{J} = J(\bar{x})$, $E_{|\bar{J}|}$ denotes the unit matrix of order $|\bar{J}|$, and $|\bar{J}|$ denotes the number of elements in $\bar{J}$. Therefore, under condition (ii) in Assumption B$'$, the Jacobian matrix $\nabla\Psi^0(\bar{x})$ is nonsingular.

In view of the above remark, we can prove superlinear convergence of Algorithm 2 in a similar manner to Theorem 5.1.

Theorem 6.2 *Let Assumptions A' and B' hold. Then $\{x_k\}$ converges superlinearly.*

7. Discussion

In this paper, we have presented Broyden-like methods for solving variational inequality and nonlinear complementarity problems. We have established global and superlinear convergence of the algorithms under suitable conditions. In particular, for nonlinear complementarity problems, global convergence only requires that F is a P_0 function with Lipschitz continuous Jacobian. On the other hand, for variational inequality problems, global convergence needs the assumption that $\nabla H^k(\bar{z})$ is nonsingular for all k sufficiently large, as in the global convergence theorems established in [18], [23] and [25]. Although this condition might appear somewhat artificial, it is implied, as shown in Lemma 2.4, by a condition that is reminiscent of the second order sufficient condition in nonlinear programming.

References

[1] T. De Luca, F. Facchinei and C. Kanzow (1996). A semismooth equation approach to the solution of nonlinear complementarity problems. *Math. Programming* **75** 407-440.

[2] J.E. Dennis, Jr. and J.J. Moré (1974). A characterization of superlinear convergence and its application to quasi-Newton methods. *Math. Comp.* **28** 549-560.

[3] F. Facchinei and S. Lucidi (1995). Quadratically and superlinearly convergent algorithms for the solution of the inequality constrained minimization problems. *J. Optim. Theory Appl.* **85** 265-289.

[4] F. Facchinei and J. Soares (1997). A new merit function for nonlinear complementarity problems and related algorithm. *SIAM J. Optim.* **7** 225-247.

[5] A. Fischer (1992). A special Newton-type optimization method. *Optimization* **24** 269-284.

[6] A. Fischer (1995). A Newton-type method for positive semidefinite linear complementarity problems. *J. Optim. Theory Appl.* **86** 585-608.

[7] R. Fletcher (1987). *Practical Methods of Optimization,* Second Edition, John Wiley & Sons, Chichester.

[8] C. Geiger and C. Kanzow (1996). On the solution of monotone complementarity problems. *Comput. Optim. Appl.* **5** 155-173.

[9] A. Griewank (1986). The 'global' convergence of Broyden-like methods with a suitable line search. *J. Austral. Math. Soc. Ser.B* **28** 75-92.

[10] S.P. Han, J.S. Pang and N. Rangaraj (1992). Globally convergent Newton method for nonsmooth equations. *Math. Oper. Res.* **17** 586-607.

[11] P.T. Harker and D. Xiao (1990). Newton's methods for nonlinear complementarity problems: A B-differentiable equation approach. *Math. Programming* **48** 339-358.

[12] P.T. Harker and J.S. Pang (1990). Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Math. Programming* **48** 161-220.

[13] H. Jiang and L. Qi (1997). A new nonsmooth equations approach to nonlinear complementarity problems. *SIAM J. Control Optim.* **35** 178-193.

[14] C. Kanzow and M. Fukushima (1996). Equivalence of the generalized complementarity problem to differentiable unconstrained minimization. *J. Optim. Theory Appl.* **90** 581-603.

[15] D. Li (1995). Global convergence of nonsingular Broyden's method for solving unconstrained optimization. *Chinese J. Numer. Math. Appl.* **17** No.4 85-94.

[16] D. Li and M. Fukushima (1998). *A modified BFGS method with global convergence in nonconvex minimization*, Technical Report 98003, Department of Applied Mathematics and Physics, Kyoto University.

[17] D. Li and M. Fukushima (1998). *A globally and superlinearly convergent Gauss-Newton based BFGS method for symmetric equations*, Technical Report 98006, Department of Applied Mathematics and Physics, Kyoto University.

[18] D. Li and Y. Yang (1998). *A globally convergent quasi-Newton method for nonlinear complementarity problems*, (submitted to *Math. Programming*).

[19] J.S. Pang (1990). Newton's method for B-differentiable equations. *Math. Oper. Res.* **15** 311-341.

[20] J.S. Pang (1991). A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems. *Math. Programming* **51** 101-131.

[21] J.S. Pang (1994). Serial and parallel computation of Karush-Kuhn-Tucker points via nonsmooth equations. *SIAM J. Optim.* **4** 872-893.

[22] L. Qi (1993). Convergence analysis of some algorithms for solving nonsmooth equations. *Math. Oper. Res.* **18** 227-244.

[23] L. Qi and X. Chen (1995). A globally convergent successive approximation method for severely nonsmooth equations. *SIAM J. Control Optim.* **33** 402-418.

[24] N. Yamashita and M. Fukushima (1997). Modified Newton methods for solving a semismooth reformulation of monotone complementarity problems. *Math. Programming* **76** 469-491.

[25] S. Zhou, D. Li and J. Zeng (1995). A successive approximation quasi-Newton process for nonlinear complementarity problems. D-Z Du, L. Qi and R. S. Womersley eds. *Recent Advanced in Nonsmooth Optimization*, World Scientific Publishing Co Pte Ltd, 459-472.