A Nonsmooth Equation Based BFGS Method for Solving KKT Systems in Mathematical Programming

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

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

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

July 15, 1998

Abstract In this paper, we present a BFGS method for solving a KKT system in mathematical programming, based on a nonsmooth equation reformulation of the KKT system. We successively split the nonsmooth equation into equivalent equations with particular structure. Based on the splitting, we develop a BFGS method in which subproblems are systems of linear equations with symmetric and positive definite coefficient matrices. A suitable line search is introduced under which the generated iterates exhibit an approximately norm decent property. The method is well defined and, under suitable conditions, converges to a KKT point globally and superlinearly without convexity assumption on the problem.

Key Words KKT system, splitting function, BFGS 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

Let f, g_i, $i = 1,2,\ldots,m$, and h_j $j = 1,2,\ldots,r$, be twice continuously differentiable functions from R^n to R. Consider the following general constrained mathematical programming problem:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) \geq 0, \quad i = 1,2,\ldots,m \\ & h_j(x) = 0, \quad j = 1,2,\ldots,r. \end{array} \tag{1.1}$$

Associated with problem (1.1) is the Lagrangian function $l : R^{n+m+r} \rightarrow R$ defined by

$$l(z) = f(x) - \lambda^T g(x) - \mu^T h(x) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x) - \sum_{j=1}^{r} \mu_j h_j(x), \tag{1.2}$$

where $z = (x, \lambda, \mu) \in R^{n+m+r}$. Iterative methods for solving (1.1) typically generate a sequence of points $\{z_k\}$ estimating $z^* = (x^*, \lambda^*, \mu^*)$, a solution of the following KKT system:

$$\begin{array}{rl} \nabla_x l(x,\lambda,\mu) = & \nabla f(x) - \sum\limits_{i=1}^{m} \lambda_i \nabla g_i(x) - \sum\limits_{j=1}^{r} \mu_j \nabla h_j(x) = 0, \\ & \lambda_i \geq 0, \;\; g_i(x) \geq 0, \;\; \lambda_i g_i(x) = 0, \quad i = 1,2,\ldots,m, \\ & h_j(x) = 0, \quad j = 1,2,\ldots,r. \end{array} \tag{1.3}$$

There have been developed many iterative methods for solving (1.3). We refer to [14] for a comprehensive treatment of these methods. Among the iterative methods, the so-called successive quadratic programming (SQP) method is considered one of the most important methods and has received much attention (see e.g. [1, 2, 3, 4, 6, 10, 15, 17, 18, 19, 24, 27, 28, 29, 30, 33]). In conventional SQP methods (see e.g. [17, 28, 29, 30]), problem (1.1) is approximated by a sequence of quadratic programming problems

$$\begin{array}{ll} \min & \frac{1}{2} p^T B_k p + \nabla f(x_k)^T p \\ \text{s.t.} & g_i(x_k) + \nabla g_i(x_k)^T p \geq 0, \quad i = 1,2,\ldots,m \\ & h_j(x_k) + \nabla h_j(x_k)^T p = 0, \quad j = 1,2,\ldots,r, \end{array} \tag{1.4}$$

where B_k is the Hessian $\nabla_x^2 l(x_k, \lambda_k, \mu_k)$ of l at $z_k = (x_k, \lambda_k, \mu_k)$ or its approximation, corresponding to Newton's method or quasi-Newton methods, respectively. There are other different versions of SQP methods where the constraints of the corresponding quadratic programming subproblems differ from those of (1.4) (see e.g. [3, 4, 15]). Under suitable conditions, SQP methods converge globally and superlinearly. Unfortunately, for most SQP methods, the subproblems sometimes turn out to be inconsistent, i.e., problem (1.4) may have no feasible solution. In recent years, some remedies have been proposed (see e.g. [3, 4, 33]). However, those methods require

additional computational efforts at each step. Another deficiency of SQP methods is that to establish global convergence of a SQP method, it is often assumed that the matrix sequence $\{B_k\}$ is uniformly bounded and positive definite, i.e., there are constants $0 < m \leq M$ such that the following inequalities hold for all k:

$$m\|p\|^2 \leq p^T B_k p \leq M\|p\|^2, \qquad \forall p \in R^n. \tag{1.5}$$

However, so far no answer has been given as to whether any particular quasi-Newton method satisfies this condition.

In this paper, we present a BFGS method for (1.3) based on an equivalent system of nonsmooth equations. In each step, the proposed method solves a system of linear equations whose coefficient matrix is symmetric and positive definite. Therefore the subproblems are always consistent and can be solved by any efficient method for linear equations. Moreover, the proposed method exhibits an approximately norm decent property. Under suitable conditions, we prove that the proposed BFGS method converges globally and superlinearly to a KKT point of (1.1) without assuming that the condition (1.5) holds for all k. In particular, the convergence analysis does not require the convexity of the problem.

The paper is organized as follows: In the next section, we first transform (1.3) into an equivalent system of nonsmooth equations and then split it into another form with a special structure. We also discuss several properties of the splitting in Section 2. In Section 3, we present BFGS method and prove some useful lemmas. In Sections 4 and 5, we prove global convergence and superlinear convergence, respectively, of the proposed BFGS method. We conclude the paper by giving some remarks in Section 6.

2. Splitting Function and Its Properties

In this section, we first reformulate the KKT system (1.3) as an equivalent nonsmooth equation and then describe a splitting method for the latter equation. The splitting function has a particular structure, which is the basis for designing a BFGS method. Let $\phi : R^2 \to R$ be the Fischer-Burmeister function [12] defined by

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

The Fischer-Burmeister function has been extensively used in the context of the nonlinear complementarity problem and related problems (see e.g. [8, 11, 12, 13, 16, 20, 21, 34]). The function ϕ is differentiable everywhere except at the origin. It is easy to see that $\phi(a, b) = 0$ if and only if

$a \geq 0, b \geq 0$ and $ab = 0$. A function with this property is called an NCP function. By the use of Fischer-Burmeister function, the KKT system (1.3) can be rewritten in the following form:

$$\begin{cases} L(z) = 0, \\ \phi(g_i(x), \lambda_i) = 0, \qquad i = 1, 2, \cdots, m \\ h_j(x) = 0, \qquad j = 1, 2, \cdots, r, \end{cases} \tag{2.1}$$

where

$$\begin{aligned} L(z) &= \nabla_x l(z) = \nabla f(x) - \sum_{i=1}^{m} \lambda_i \nabla g_i(x) - \sum_{j=1}^{r} \mu_j \nabla h_j(x) \\ &= \nabla f(x) - \nabla g(x)\lambda - \nabla h(x)\mu. \end{aligned}$$

Define $\Phi(x, \lambda) = (\phi_1(x, \lambda), \phi_2(x, \lambda), \ldots, \phi_m(x, \lambda))^T$ with elements

$$\phi_i(x, \lambda) = \phi(g_i(x), \lambda_i), \quad i = 1, 2, \cdots, m.$$

Then (2.1) is represented in a compact form:

$$F(z) \stackrel{\triangle}{=} \begin{pmatrix} L(z) \\ -\Phi(x, \lambda) \\ -h(x) \end{pmatrix} = 0. \tag{2.2}$$

The function F is generally not differentiable because of the nondifferentiability of ϕ and hence cannot be solved by ordinary iterative methods for smooth equations. Recently, there have been developed various Newton-type methods for solving nonsmooth equations with global and superlinear/quadratic convergence properties (see e.g. [18, 25, 26, 27, 31, 32]). However, the study on quasi-Newton methods for nonsmooth equations is relatively scarce. In this paper, we propose a BFGS method for solving the KKT system (1.3) or equivalently the nonsmooth equation (2.2). We construct an iterative process based on the splitting

$$F(z) = F_k(z) + G_k(z) = 0, \tag{2.3}$$

where F_k is differentiable everywhere and G_k is not necessary differentiable but relatively small compared with F_k.

The idea of approximating a nonsmooth equation using a splitting function first appeared in [7]. Qi and Chen [32] proposed a splitting function based successive Newton method for a class of nonsmooth equations and showed that under suitable conditions, the method converges globally and quadratically. The results obtained in [32] were extended in [35] to Broyden-like methods

for solving nonlinear complementarity problems. In this paper, we propose a BFGS method that is somewhat similar to the one in [35] but utilizes a different line search rule. Moreover, the conditions to ensure global convergence of the method are weaker than those in [35].

We now split F into an equivalent form so that the corresponding nonsmooth equation (2.2) takes the form (2.3). For a given constant $\epsilon > 0$, define

$$\begin{aligned}\phi^\epsilon(a,b) &= \begin{cases} \phi(a,b), & \text{if } \sqrt{a^2+b^2} \geq \epsilon, \\ \phi(a,b) + \dfrac{(\sqrt{a^2+b^2}-\epsilon)^2}{2\epsilon}, & \text{if } \sqrt{a^2+b^2} < \epsilon \end{cases} \\ &= \begin{cases} \sqrt{a^2+b^2} - (a+b), & \text{if } \sqrt{a^2+b^2} \geq \epsilon, \\ \frac{1}{2\epsilon}[(a^2+b^2) - 2\epsilon(a+b) + \epsilon^2], & \text{if } \sqrt{a^2+b^2} < \epsilon. \end{cases}\end{aligned}$$

Then ϕ^ϵ is everywhere continuously differentiable and

$$\nabla\phi^\epsilon(a,b) = \begin{cases} \left(\dfrac{a}{\sqrt{a^2+b^2}} - 1, \dfrac{b}{\sqrt{a^2+b^2}} - 1\right)^T, & \text{if } \sqrt{a^2+b^2} \geq \epsilon, \\ (\frac{a}{\epsilon} - 1, \frac{b}{\epsilon} - 1)^T, & \text{if } \sqrt{a^2+b^2} < \epsilon. \end{cases}$$

Moreover, we have

$$|\phi(a,b) - \phi^\epsilon(a,b)| \leq \frac{1}{2}\epsilon, \qquad \forall (a,b) \in R^2.$$

Further define

$$\phi_i^\epsilon(x,\lambda) = \phi^\epsilon(g_i(x), \lambda_i), \quad i = 1,2,\ldots,m,$$

$$\Phi^\epsilon(x,\lambda) = (\phi_1^\epsilon(x,\lambda), \phi_2^\epsilon(x,\lambda), \ldots, \phi_m^\epsilon(x,\lambda))^T,$$

$$F^\epsilon(z) = \begin{pmatrix} L(z) \\ -\Phi^\epsilon(x,\lambda) \\ -h(x) \end{pmatrix} = \begin{pmatrix} \nabla f(x) - \nabla g(x)\lambda - \nabla h(x)\mu \\ -\Phi^\epsilon(x,\lambda) \\ -h(x) \end{pmatrix}$$

and

$$G^\epsilon(z) = F(z) - F^\epsilon(z) = \begin{pmatrix} 0 \\ -(\Phi(x,\lambda) - \Phi^\epsilon(x,\lambda)) \\ 0 \end{pmatrix}.$$

Then we get a splitting form of F:

$$F(z) = F^\epsilon(z) + G^\epsilon(z). \tag{2.4}$$

It is not difficult to see that for every $\epsilon > 0$, F^ϵ is continuously differentiable since Φ^ϵ is continuously differentiable. We will give a direct expression of the Jacobian $\nabla F^\epsilon(z)$ later. Note that

G^ϵ is not differentiable, but the norm of it is uniformly restricted by ϵ. This means that if ϵ is small, then so is $\|G^\epsilon(x,\lambda)\|$ for every $(x,\lambda) \in R^{n+m}$. To be precise, we have

$$\|G^\epsilon(z)\| = \|\Phi(x,\lambda) - \Phi^\epsilon(x,\lambda)\| \le \frac{\sqrt{m}}{2}\epsilon, \qquad \forall z \in R^{n+m+r}.$$

We now prove some useful lemmas for the splitting function (2.4). First, by the previous arguments, it is easy to show the following lemma.

Lemma 2.1 *(i) For every* $z = (x,\lambda,\mu) \in R^{n+m+r}$,

$$\|G^\epsilon(z)\| = \|\Phi(x,\lambda) - \Phi^\epsilon(x,\lambda)\| \le \frac{\sqrt{m}}{2}\epsilon. \tag{2.5}$$

(ii) $\Phi(x,\lambda) = 0$ *if and only if* $\min(\lambda, g(x)) = 0$.

(iii) $F(z) = 0$ *if and only of* z *is a KKT point of (1.1).*

Let $I^\epsilon(x,\lambda)$ and $J^\epsilon(x,\lambda)$ be the index sets given by

$$I^\epsilon(x,\lambda) = \{i \mid \sqrt{\lambda_i^2 + g_i(x)^2} \ge \epsilon\} \quad \text{and} \quad J^\epsilon(x,\lambda) = \{i \mid \sqrt{\lambda_i^2 + g_i(x)^2} < \epsilon\}, \tag{2.6}$$

respectively. Put

$$a_i^\epsilon(x,\lambda) = \begin{cases} \dfrac{\lambda_i}{\sqrt{\lambda_i^2 + g_i(x)^2}} - 1, & \text{if} \quad i \in I^\epsilon(x,\lambda), \\ \dfrac{\lambda_i}{\epsilon} - 1, & \text{if} \quad i \in J^\epsilon(x,\lambda), \end{cases}$$

$$b_i^\epsilon(x,\lambda) = \begin{cases} \dfrac{g_i(x)}{\sqrt{\lambda_i^2 + g_i(x)^2}} - 1, & \text{if} \quad i \in I^\epsilon(x,\lambda), \\ \dfrac{g_i(x)}{\epsilon} - 1, & \text{if} \quad i \in J^\epsilon(x,\lambda). \end{cases}$$

The next lemma immediately follows from the definition of $a_i^\epsilon(x,\lambda)$ and $b_i^\epsilon(x,\lambda)$.

Lemma 2.2 *For any* $\epsilon > 0$ *and every* $(x,\lambda) \in R^{n+m}$, *we have* $-2 \le a_i^\epsilon(x,\lambda) \le 0$ *and* $-2 \le b_i^\epsilon(x,\lambda) \le 0$ *for all* $i = 1,2,\ldots,m$. *Moreover,* $a_i^\epsilon(x,\lambda) = 0$ *if and only if* $\lambda_i \ge \epsilon$ *and* $g_i(x) = 0$ *or equivalently* $b_i^\epsilon(x,\lambda) = -1$; $b_i^\epsilon(x,\lambda) = 0$ *if and only if* $g_i(x) \ge \epsilon$ *and* $\lambda_i = 0$ *or equivalently* $a_i^\epsilon(x,\lambda) = -1$.

The following lemma gives a direct expression of $\nabla F^\epsilon(z)$. It can be derived from the expression of $\nabla \phi^\epsilon$ and the chain rule for the derivatives of composite functions.

Lemma 2.3 *Let f, g_i, $i = 1,2,\ldots,m$, and h_j, $j = 1,2,\ldots,r$, be twice continuously differentiable. Then Φ^ϵ and F^ϵ are continuously differentiable. Moreover, for every $z = (x,\lambda,\mu) \in R^{n+m+r}$, we have the following formulas:*

$$\nabla_x \phi_i^\epsilon(x,\lambda) = b_i^\epsilon(x,\lambda)\nabla g_i(x,\lambda), \qquad \nabla_\lambda \phi_i^\epsilon(x,\lambda) = a_i^\epsilon(x,\lambda)e_i,$$

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

and

$$\begin{aligned} \nabla F^\epsilon(z) &= \begin{pmatrix} \nabla_x L(z) & -\nabla_x \Phi^\epsilon(x,\lambda) & -\nabla h(x) \\ \nabla_\lambda L(z) & -\nabla_\lambda \Phi^\epsilon(x,\lambda) & 0 \\ \nabla_\mu L(z) & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \nabla_x L(z) & -\nabla g(x)\ \mathrm{diag}\ (b_i^\epsilon(x,\lambda)) & -\nabla h(x) \\ -\nabla g(x)^T & -\ \mathrm{diag}\ (a_i^\epsilon(x,\lambda)) & 0 \\ -\nabla h(x)^T & 0 & 0 \end{pmatrix}, \end{aligned} \tag{2.7}$$

where e_i denotes the i-th column of the identity matrix.

The next lemma gives a regularity condition on function F^ϵ. For a given $z = (x,\lambda,\mu)$ and $\epsilon > 0$, we define

$$S^\epsilon(z) = \{i \,|\, g_i(x) = 0, \lambda_i \geq \epsilon\}. \tag{2.8}$$

Lemma 2.4 *Let $z = (x,\lambda,\mu) \in R^{n+m+r}$ and $\epsilon > 0$ be given. Suppose that the vectors $\nabla g_i(x), i \in S^\epsilon(z)$, and $\nabla h_j(x), j = 1,2,\ldots,r$, are linearly independent. Suppose also that the matrix*

$$\nabla_x L(z) = \nabla_x^2 l(z) = \nabla^2 f(x) - \sum_{i=1}^{m} \lambda_i \nabla^2 g_i(x) - \sum_{j=1}^{r} \mu_j \nabla^2 h_j(x)$$

is positive definite on the subspace $\{p \,|\, \nabla g_i(x)^T p = 0, \forall i \in S^\epsilon(z), \nabla h(x)^T p = 0\}$. Then $\nabla F^\epsilon(z)$ is nonsingular.

Proof We verify the nonsingularity of $\nabla F^\epsilon(z)$ by showing that the system of linear equations $\nabla F^\epsilon(z)w = 0$ has the unique solution $w = 0$. Let $w = (p,q,s) \in R^{n+m+r}$. Then the linear system $\nabla F^\epsilon(z)w = 0$ is expressed as

$$\left\{ \begin{array}{rcrcrcl} \nabla_x L(z)p & - & \nabla g(x)\ \mathrm{diag}\ (b_i^\epsilon(x,\lambda))q & - & \nabla h(x)s & = & 0, \\ -\nabla g(x)^T p & - & \mathrm{diag}\ (a_i^\epsilon(x,\lambda))q & & & = & 0, \\ -\nabla h(x)^T p & & & & & = & 0. \end{array} \right. \tag{2.9}$$

It then follows that

$$p^T \nabla_x L(z)p + q^T \text{ diag } (a_i^\epsilon(x,\lambda))b_i^\epsilon(x,\lambda))q = 0. \tag{2.10}$$

Since $a_i^\epsilon(x,\lambda) \leq 0$ and $b_i^\epsilon(x,\lambda) \leq 0$, by Lemma 2.2, the second term on the left-hand side of (2.10) is nonnegative. On the other hand, since p satisfies the third equality of (2.9), and $\nabla g_i(x)^T p = 0, \forall i \in S^\epsilon(z)$ by Lemma 2.2, the positive definiteness of $\nabla_x L(z)$ on the subspace $\{p \,|\, \nabla g_i(x)^T p = 0, \forall i \in S^\epsilon(z), \nabla h(x)^T p = 0\}$ implies that the first term on the left-hand side of (2.10) is also nonnegative. Consequently, p must be zero and (2.9) reduces to

$$\begin{cases} \nabla g(x) \text{ diag } (b_i^\epsilon(x,\lambda))q + \nabla h(x)s &= 0, \\ - \text{ diag } (a_i^\epsilon(x,\lambda))q &= 0. \end{cases} \tag{2.11}$$

From the second equality of (2.11), it is obvious that $q_i = 0$ for every i with $a_i^\epsilon(x,\lambda) \neq 0$. In other words, the only possible nonzero elements q_i correspond to the indices i satisfying $a_i^\epsilon(x,\lambda) = 0$ which, by Lemma 2.2, coincide with the indices $i \in S^\epsilon(z)$. Moreover, for these i, $b_i^\epsilon(x,\lambda) = -1$. Therefore, from the first equality of (2.11) we get

$$- \sum_{i \in S^\epsilon(z)} q_i \nabla g_i(x) + \sum_{j=1}^{r} s_j \nabla h_j(x) = 0.$$

By the assumption that $\nabla g_i(x), i \in S^\epsilon(z)$, and $\nabla h_j(x), j = 1,2,\ldots,r$, are linearly independent, it follows that $q_i = 0, \forall i \in S^\epsilon(z)$ and $s_j = 0$, $j = 1,2,\ldots,r$. In other words, zero is the unique solution of (2.9). Therefore, $\nabla F^\epsilon(z)$ is nonsingular. □

3. Algorithm

This section is devoted to presenting a BFGS method for the nonsmooth equation (2.2) based on the splitting function (2.4). For convenience, we simplify some notations. For a given positive sequence $\{\epsilon_k\}$, we abbreviate F^{ϵ_k}, G^{ϵ_k}, $a_i^{\epsilon_k}$, $b_i^{\epsilon_k}$ etc. as F_k, G_k, a_i^k, b_i^k etc., respectively.

To describe the method, it would be helpful to recall the Newton-type method proposed by Qi and Chen [32]. In their method, the subproblem to be solved at each iteration is the following system of linear equations in p:

$$\nabla F_k(z_k)^T p + F(z_k) = 0.$$

If $\nabla F_k(z_k)$ in nonsingular, then the subproblem is equivalent to the linear equation

$$\nabla F_k(z_k) \nabla F_k(z_k)^T p + \nabla F(z_k) F(z_k) = 0. \tag{3.1}$$

Now we consider avoiding the calculation of the second derivatives $\nabla^2 f$, $\nabla^2 g_i$, $i = 1, 2, \ldots m$, and $\nabla^2 h_j$, $j = 1, 2, \ldots, r$, in (3.1). Specifically, we consider the subproblem

$$B_k p + q_k = 0,$$

where matrix $B_k \in R^{(n+m+r)\times(n+m+r)}$ approximates $\nabla F_k(z_k)\nabla F_k(z_k)^T$ and vector $q_k \in R^{n+m+r}$ approximates $\nabla F_k(z_k)F(z_k)$. When ∇F_k is symmetric, the authors [23] proposed a quasi-Newton method that generates such an approximating matrix B_k and an approximating vector q_k. In the present case, however, ∇F_k is generally not symmetric, and hence the method of [23] cannot be applied. To obtain suitable approximation B_k and q_k, let

$$Q_k = \begin{pmatrix} 0 & -\nabla g(x_k)\mathrm{diag}(b_i^k(x_k,\lambda_k)) & -\nabla h(x_k) \\ -\nabla g(x_k)^T & -\mathrm{diag}(a_i^k(x_k,\lambda_k)) & 0 \\ -\nabla h(x_k)^T & 0 & 0 \end{pmatrix}. \tag{3.2}$$

We notice that Q_k does not contain the second derivatives $\nabla^2 f$, $\nabla^2 g_i$, $i = 1, 2, \ldots m$, and $\nabla^2 h_j$, $j = 1, 2, \ldots, r$. By the definition of Q_k, we immediately have the following relationship

$$\nabla F_k(z_k) = Q_k + \begin{pmatrix} \nabla_x L(z_k) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{3.3}$$

Notice that the matrix in the second term on the right-hand side is a symmetric matrix. Let

$$q_k = \alpha_{k-1}^{-1} \begin{pmatrix} L(x_k + \alpha_{k-1}L(z_k), \lambda_k, \mu_k) - L(z_k) \\ 0 \\ 0 \end{pmatrix} + Q_k F_k(z_k), \tag{3.4}$$

where α_{k-1} is the steplength determined at the $(k-1)$-th iteration. Then we have

$$\begin{aligned} q_k &= \begin{pmatrix} \int_0^1 \nabla_x L(x_k + \tau\alpha_{k-1}L(z_k), \lambda_k, \mu_k)d\tau L(z_k) \\ 0 \\ 0 \end{pmatrix} + Q_k F_k(z_k) \\ &= \begin{pmatrix} \int_0^1 \nabla_x L(x_k + \tau\alpha_{k-1}L(z_k), \lambda_k, \mu_k)d\tau & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} F_k(z_k) + Q_k F_k(z_k) \\ &= \nabla F_k(z_k)F_k(z_k) + \begin{pmatrix} \int_0^1 [\nabla_x L(x_k + \tau\alpha_{k-1}L(z_k), \lambda_k, \mu_k) - \nabla_x L(z_k)]d\tau & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} F_k(z_k) \\ &= \nabla F_k(z_k)F_k(z_k) + u_k, \end{aligned} \tag{3.5}$$

where

$$u_k = \begin{pmatrix} \int_0^1 [\nabla_x L(x_k + \tau\alpha_{k-1}L(z_k), \lambda_k, \mu_k) - \nabla_x L(z_k)]d\tau & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} F_k(z_k). \tag{3.6}$$

Note that u_k satisfies

$$\|u_k\| \le \int_0^1 \|\nabla_x L(x_k + \tau\alpha_{k-1}L(z_k), \lambda_k, \mu_k) - \nabla_x L(z_k)\|d\tau\|F_k(z_k)\|.$$

If $\|\alpha_{k-1}L(z_k)\|$ is small, then so is $\|u_k\|$. By (2.4) and (2.5), when ϵ_k is small enough, we have $F(z_k) \approx F_k(z_k)$. Therefore, (3.5) implies that q_k defined by (3.4) is an approximation of $\nabla F(z_k)F(z_k)$. Moreover, the calculation of q_k does not require the second derivatives $\nabla^2 f$, $\nabla^2 g_i, i = 1, 2, \ldots, m$, and $\nabla^2 h_j, j = 1, 2, \ldots, r$.

We now construct B_k such that $B_k p_k \approx \nabla F_k(z_k)\nabla F_k(z_k)^T p_k$ by the BFGS update formula:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k},$$

where $s_k = z_{k+1} - z_k$ and y_k is determined as specified shortly. First, we notice that B_{k+1} satisfies the secant equation $B_{k+1}s_k = y_k$. So, we only need to select y_k appropriately so that $y_k \approx \nabla F_{k+1}(z_{k+1})\nabla F_{k+1}(z_{k+1})^T s_k$. Denote

$$\delta_k = F_k(z_{k+1}) - F_k(z_k) = \int_0^1 \nabla F_k(z_k + \tau s_k)^T d\tau s_k \tag{3.7}$$

and

$$r_k = \begin{pmatrix} L(x_k + L(z_{k+1}) - L(z_k), \lambda_k, \mu_k) - L(z_k) \\ 0 \\ 0 \end{pmatrix} + Q_k\delta_k. \tag{3.8}$$

Then by (3.3), we get

$$\begin{aligned} r_k &= \begin{pmatrix} \int_0^1 \nabla_x L(x_k + \tau(L(z_{k+1}) - L(z_k)), \lambda_k, \mu_k)d\tau(L(z_{k+1}) - L(z_k)) \\ 0 \\ 0 \end{pmatrix} + Q_k\delta_k \\ &= \begin{pmatrix} \int_0^1 \nabla_x L(x_k + \tau(L(z_{k+1}) - L(z_k)), \lambda_k, \mu_k)d\tau & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \delta_k + Q_k\delta_k \end{aligned}$$

$$
\begin{aligned}
&= \nabla F_k(z_k)\delta_k + \begin{pmatrix} \int_0^1 \{\nabla_x L(x_k + \tau(L(z_{k+1}) - L(z_k)), \lambda_k, \mu_k) - \nabla_x L(z_k)\} d\tau & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \delta_k \\
&= \nabla F_k(z_k) \int_0^1 \nabla F_k(z_k + \tau s_k)^T d\tau s_k + v_k \\
&= \nabla F_k(z_k)\nabla F_k(z_k)^T s_k + \nabla F_k(z_k) \int_0^1 [\nabla F_k(z_k + \tau s_k)^T - \nabla F_k(z_k)^T] d\tau s_k + v_k \\
&= \nabla F_k(z_k)\nabla F_k(z_k)^T s_k + \bar{v}_k,
\end{aligned}
$$

where

$$
v_k = \begin{pmatrix} \int_0^1 \{\nabla_x L(x_k + \tau(L(z_{k+1}) - L(z_k)), \lambda_k, \mu_k) - \nabla_x L(z_k)\} d\tau & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \delta_k
$$

and

$$
\bar{v}_k = \nabla F_k(z_k) \int_0^1 [\nabla F_k(z_k + \tau s_k)^T - \nabla F_k(z_k)^T] d\tau s_k + v_k. \tag{3.9}
$$

We can estimate v_k as

$$
\begin{aligned}
\|v_k\| &\le \Big\| \int_0^1 [\nabla_x L(x_k + \tau(L(z_{k+1}) - L(z_k)), \lambda_k, \mu_k) - \nabla_x L(z_k)] d\tau \Big\| \, \|\delta_k\| \\
&\le \int_0^1 \|\nabla_x L(x_k + \tau(L(z_{k+1}) - L(z_k)), \lambda_k, \mu_k) - \nabla_x L(z_k)\| d\tau \, \|\delta_k\| \\
&\le \int_0^1 \|\nabla_x L(x_k + \tau(L(z_{k+1}) - L(z_k)), \lambda_k, \mu_k) - \nabla_x L(z_k)\| d\tau \\
&\quad \cdot \int_0^1 \|\nabla F_k(z_k + \tau s_k)\| d\tau \|s_k\|,
\end{aligned}
$$

where the last inequality follows from the definition (3.7) of δ_k. Similarly we have

$$
\begin{aligned}
\|\bar{v}_k\| &\le \|\nabla F_k(z_k)\| \Big\| \int_0^1 [\nabla F_k(z_k + \tau s_k)^T - \nabla F_k(z_k)^T] d\tau \Big\| \, \|s_k\| + \|v_k\| \\
&\le \|\nabla F_k(z_k)\| \int_0^1 \|\nabla F_k(z_k + \tau s_k) - \nabla F_k(z_k)\| d\tau \, \|s_k\| \\
&\quad + \int_0^1 \|\nabla_x L(x_k + \tau(L(z_{k+1}) - L(z_k)), \lambda_k, \mu_k) - \nabla_x L(z_k)\| d\tau \\
&\quad \cdot \int_0^1 \|\nabla F_k(z_k + \tau s_k)\| d\tau \|s_k\|.
\end{aligned}
$$

Assume that $\|\nabla F_k\|$ is bounded with an upper bound $M > 0$. Then we have

$$
\begin{aligned}
\|\bar{v}_k\| &\le M \Big(\int_0^1 \|\nabla F_k(z_k + \tau s_k) - \nabla F_k(z_k)\| d\tau \\
&\quad + \int_0^1 \|\nabla_x L(x_k + \tau(L(z_{k+1}) - L(z_k)), \lambda_k, \mu_k) - \nabla_x L(z_k)\| d\tau \Big) \|s_k\|.
\end{aligned}
$$

Since $s_k = z_{k+1} - z_k$, we see that $\|\bar{v}_k\| = o(\|s_k\|)$, which implies $r_k = \nabla F_k(z_{k+1})\nabla F_k(z_{k+1})^T s_k + o(\|s_k\|)$. Thus r_k seems to be a reasonable choice for y_k satisfying $B_{k+1}s_k = y_k$. Nevertheless, since $\nabla F_k(z_k)$ can be singular for some k, taking $y_k = r_k$ does not ensure that $y_k^T s_k > 0$ holds for every k which is a sufficient condition for B_{k+1} to inherit the positive definiteness from B_k. To cope with this problem, we adopt a technique used in [22] to take

$$y_k = r_k + \tau_k \|F_k(z_k)\| s_k, \tag{3.10}$$

where τ_k is defined by

$$\tau_k = 1 - \frac{1}{\|F_k(z_k)\|} \min\Big\{\frac{r_k^T s_k}{\|s_k\|^2}, 0\Big\}. \tag{3.11}$$

Note that $\tau_k \geq 1$. If $\|F_k(z_k)\|$ is small, then $y_k \approx r_k$ is also expected to be a suitable choice of y_k. Moreover, we have for every k

$$\begin{aligned} y_k^T s_k &= r_k^T s_k + \tau_k \|F_k(z_k)\| \|s_k\|^2 \\ &= \|F_k(z_k)\| \, \|s_k\|^2 + r_k^T s_k - \min\{r_k^T s_k, 0\} \\ &\geq \|F_k(z_k)\| \, \|s_k\|^2. \end{aligned} \tag{3.12}$$

This means that B_{k+1} will be positive definite provided that B_k is positive definite.

In the algorithm, we use a line search similar to the one proposed in [23]. Let $\{\eta_k\}$ be a positive sequence satisfying

$$\sum_{k=0}^{\infty} \eta_k \leq \eta < \infty, \tag{3.13}$$

where η is a positive constant. We determine a steplength $\alpha_k > 0$ so that the following inequality holds for $\alpha = \alpha_k$:

$$\|F_k(z_k + \alpha p_k)\|^2 - \|F_k(z_k)\|^2 \leq -\sigma_1 \|\alpha F_k(z_k)\|^2 - \sigma_2 \|\alpha p_k\|^2 + \eta_k \|F_k(z_k)\|^2, \tag{3.14}$$

where σ_1 and σ_2 are given positive constants. It is not difficult to see that (3.14) is satisfied for all sufficiently small $\alpha > 0$, because the last term on the right-hand side is positive and independent of α.

We state a BFGS method for solving (2.2).

Algorithm 1

Step 0. Choose constants $\rho \in (0,1), \beta \in (0,1), 0 < \gamma < \frac{2}{\sqrt{m}}, \sigma_1 > 0, \sigma_2 > 0, \lambda_{-1} > 0$. Select a positive sequence $\{\eta_k\}$ satisfying (3.13). Choose an initial point $z_0 \in R^{n+m+r}$, a symmetric positive definite matrix $B_0 \in R^{(n+m+r)\times(n+m+r)}$ and $\epsilon_0 \leq \frac{\gamma}{2}\|F(z_0)\|$. Let $k := 0$.

Step 1. Let q_k be given by (3.4). Solve the linear equation

$$B_k p + q_k = 0 \tag{3.15}$$

to get p_k.

Step 2. If

$$\|F(z_k + p_k)\| \leq \beta \|F(z_k)\|, \tag{3.16}$$

then let $\alpha_k = 1$ and go to Step 4. Otherwise go to Step 3.

Step 3. Let i_k be the smallest nonnegative integer i such that $\alpha = \rho^i$ satisfies (3.14) and let $\alpha_k = \rho^{i_k}$.

Step 4. Let $z_{k+1} = z_k + \alpha_k p_k$.

Step 5. Update B_k by the BFGS formula

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}, \tag{3.17}$$

where $s_k = z_{k+1} - z_k$ and y_k is determined by (3.10) with r_k and τ_k given by (3.8) and (3.11), respectively.

Step 6 If $\epsilon_k < \gamma \|F(z_{k+1})\|$, take $\epsilon_{k+1} = \epsilon_k$. Otherwise, determine ϵ_{k+1} by

$$\epsilon_{k+1} \leq \min\{\frac{\gamma}{2}\|F(z_{k+1})\|, \frac{1}{2}\epsilon_k\}. \tag{3.18}$$

Step 7. Let $k := k + 1$. Go to Step 1.

We now prove some useful properties of the algorithm. In the rest of the paper, we need the following basic assumption.

Assumption A (i) The level set

$$\Omega = \{z \in R^{n+m+r} \mid \|F(z)\| \leq 6e^{\eta}\|F(z_0)\|\} \tag{3.19}$$

is bounded, where η is a positive constant such that (3.13) holds.

(ii) The functions f, $g_i, i = 1, 2, \ldots, m$, and $h_j, j = 1, 2, \ldots, r$, are twice continuously differentiable on a bounded convex set D containing Ω.

We shall show in Theorem 4.1 that under Assumption A, the sequence $\{z_k\}$ generated by Algorithm 1 is contained in Ω.

Let $M > 0$ be an upper bound of $\|\nabla F_k(z)\|$ over D, which can be chosen independently of k by Lemmas 2.2 and 2.3. Then the following proposition follows from the previous discussion.

Proposition 3.1 *Let q_k, δ_k, r_k and y_k be given by (3.4), (3.7), (3.8) and (3.10), respectively. Then we have for every k*

$$q_k = \nabla F_k(z_k)F_k(z_k) + u_k \tag{3.20}$$

with u_k satisfying

$$\|u_k\| \le \int_0^1 \|\nabla_x L(x_k + \tau\alpha_{k-1}L(z_k), \lambda_k, \mu_k) - \nabla_x L(z_k)\|d\tau\|F_k(z_k)\| \tag{3.21}$$

and

$$r_k = \nabla F_k(z_k)\nabla F_k(z_k)^T s_k + \bar{v}_k \tag{3.22}$$

with $\bar{v}_k$ satisfying

$$\begin{aligned} \|\bar{v}_k\| \;\le\; & M\Big(\int_0^1 \|\nabla F_k(z_k + \tau s_k) - \nabla F_k(z_k)\|d\tau \\ & + \int_0^1 \|\nabla_x L(x_k + \tau(L(z_{k+1}) - L(z_k)), \lambda_k, \mu_k) - \nabla_x L(z_k)\|d\tau\Big)\|s_k\| \end{aligned} \tag{3.23}$$

and

$$\bar{v}_k \le 4M^2\|s_k\|. \tag{3.24}$$

Proof Inequalities (3.21) and (3.23) follow from the previous discussion directly. To prove (3.24), notice that $\|\nabla_x L(z)\| \le \|\nabla_x F_k(z)\| \le \|\nabla F_k(z)\| \le M$ holds for any $z \in \Omega$ and every k. Thus from (3.23) we have

$$\begin{aligned} \|\bar{v}_k\| \;\le\; & M\Big(\int_0^1 \Big[\|\nabla F_k(z_k + \tau s_k)\| + \|\nabla F_k(z_k)\|\Big]d\tau \\ & + \int_0^1 \Big[\|\nabla_x L(x_k + \tau(L(z_{k+1}) - L(z_k)), \lambda_k, \mu_k)\| + \|\nabla_x L(z_k)\|\Big]d\tau\Big)\|s_k\| \\ \le\; & 4M^2\|s_k\|. \end{aligned}$$

□

Proposition 3.2 *Let $\bar{M} > 0$ be an upper bound of $\|F(z)\|$ over D. Then for every k, the vector $y_k = r_k + \tau_k\|F_k(z_k)\|s_k$ satisfies*

$$\|y_k\| \le (10M^2 + \bar{M})\|s_k\|. \tag{3.25}$$

Proof By (3.22) and (3.24) we get

$$\|r_k\| \le \|\nabla F_k(z_k)\|^2\|s_k\| + \|\bar{v}_k\| \le 5M^2\|s_k\|.$$

This together with the definition of y_k yields

$$\begin{aligned}
\|y_k\| &\leq 5M^2\|s_k\| + \tau_k\|F_k(z_k)\|\,\|s_k\| \\
&\leq 5M^2\|s_k\| + \Big(\|F_k(z_k)\| - \min\Big\{\frac{r_k^T s_k}{\|s_k\|^2}, 0\Big\}\Big)\|s_k\| \\
&\leq 5M^2\|s_k\| + \Big(\bar{M} + \frac{|r_k^T s_k|}{\|s_k\|^2}\Big)\|s_k\| \\
&\leq \Big(5M^2 + \bar{M} + \frac{\|r_k\|\,\|s_k\|}{\|s_k\|^2}\Big)\|s_k\| \\
&\leq (10M^2 + \bar{M})\|s_k\|.
\end{aligned}$$

□

Denote

$$\bar{K}_1 = \{k \mid \epsilon_k \geq \gamma\|F(z_{k+1})\|\} \quad \text{and} \quad \bar{K}_2 = \{k \mid (3.16) \text{ holds}\}. \tag{3.26}$$

Then we have the following proposition from Algorithm 1 immediately.

Proposition 3.3 *Let $\{z_k\}$ be generated by Algorithm 1. Then for every k,*

$$\gamma\|F(z_k)\| \geq \epsilon_k, \qquad \|G_k(z)\| \leq \frac{\sqrt{m}}{2}\epsilon_k \leq \bar{\gamma}\|F(z_k)\|, \quad \forall z \in R^{n+m+r}, \tag{3.27}$$

where $\bar{\gamma} = \frac{\sqrt{m}}{2}\gamma \in (0,1)$. Moreover, for every k, the following inequality holds:

$$\|F_k(z_{k+1})\|^2 \leq (1+\eta_k)\|F_k(z_k)\|^2. \tag{3.28}$$

Proof The first inequality in (3.27) follows form the choice of ϵ_k, the second inequality follows from (2.5), and the last inequality then follows from the first inequality. Now we prove (3.28). If $k \in \bar{K}_2$, then $\|F_k(z_{k+1})\| \leq \rho\|F_k(z_k)\| \leq (1+\eta_k)\|F_k(z_k)\|$. Otherwise, by the definition of $\bar{K}_2$, $k \notin \bar{K}_2$ means that z_{k+1} is generated by the line search in Step 3. In other words, we have $z_{k+1} = z_k + \alpha_k p_k$ with α_k satisfying (3.14), which also implies (3.28). □

Proposition 3.4 *Let $\{z_k\}$ be generated by Algorithm 1. Then*

$$\liminf_{k\to\infty} \|F(z_k)\| = 0 \tag{3.29}$$

if and only if

$$\liminf_{k\to\infty} \|F_k(z_k)\| = 0. \tag{3.30}$$

Proof By (3.27), we have for every k

$$\|F(z_k)\| \leq \|F_k(z_k)\| + \|G_k(z_k)\| \leq \|F_k(z_k)\| + \bar{\gamma}\|F(z_k)\|,$$

which implies

$$\|F(z_k)\| \leq \frac{1}{1-\bar{\gamma}}\|F_k(z_k)\|, \quad \forall k. \tag{3.31}$$

On the other hand, we have for every k

$$\|F_k(z_k)\| \leq \|F(z_k)\| + \|G_k(z_k)\| \leq (1+\bar{\gamma})\|F(z_k)\|. \tag{3.32}$$

The inequalities (3.31) and (3.32) show the equivalence between (3.29) and (3.30). □

Proposition 3.4 reveals that to prove the global convergence of Algorithm 1, it suffices to show that there is a subsequence of $\{F_k(z_k)\}$ converging to zero. In the next section, we will accomplish this under suitable conditions.

4. Global Convergence

In this section, we prove global convergence of Algorithm 1. First, we show a convergence result similar to that in [32] and [35].

Theorem 4.1 *Let $\{z_k\}$ be generated by Algorithm 1. Then $\{z_k\} \subset \Omega$. Moreover, if $\bar{K}_1$ is infinite, then*

$$\liminf_{k\to\infty} \|F(z_k)\| = 0. \tag{4.1}$$

Proof. Let $\bar{K}_1 = \{k_0 < k_1 < k_2 < \cdots\}$. Then by Step 6 of Algorithm 1, $\epsilon_k = \epsilon_{k_i}$ for every k such that $k_{i-1} < k \leq k_i$. Moreover, for every $k_i \in \bar{K}_1$, $\epsilon_{k_i+1} \leq \frac{\gamma}{2}\|F(z_{k_i+1})\|$. It then follows that

$$\begin{aligned} \|F(z_{k_i+1})\| &\leq \gamma^{-1}\epsilon_{k_i} = \gamma^{-1}\epsilon_{k_{i-1}+1} \leq \frac{1}{2}\|F(z_{k_{i-1}+1})\| \\ &\leq \cdots \leq (\frac{1}{2})^i\|F(z_{k_0+1})\| \leq \gamma^{-1}(\frac{1}{2})^i\epsilon_{k_0} \leq (\frac{1}{2})^i\|F(z_{k_0})\|. \end{aligned} \tag{4.2}$$

This implies (4.1). Moreover, if $\epsilon_0 \geq \gamma\|F(z_1)\|$, i.e. $k_0 = 0$, then $\|F(z_{k_i+1})\| \leq (\frac{1}{2})^i\|F(z_0)\| \leq e^{\eta}\|F(z_0)\|$. Otherwise, for all $k \leq k_0$, $\epsilon_k = \epsilon_0$ and $F_k = F_0$, $G_k = G_0$. Thus it follows from (3.28) that

$$\begin{aligned} \|F(z_{k_0})\| &\leq \|F_{k_0}(z_{k_0})\| + \|G_{k_0}(z_{k_0})\| \\ &= \|F_{k_0-1}(z_{k_0})\| + \|G_{k_0}(z_{k_0})\| \\ &\leq (1+\eta_{k_0-1})\|F_{k_0-1}(z_{k_0-1})\| + \|G_{k_0}(z_{k_0})\| \end{aligned}$$

$$
\begin{aligned}
&= (1+\eta_{k_0-1})\|F_{k_0-2}(z_{k_0-1})\| + \|G_0(z_{k_0})\| \\
&\le (1+\eta_{k_0-1})(1+\eta_{k_0-2})\|F_{k_0-2}(z_{k_0-2})\| + \|G_0(z_{k_0})\| \\
&= (1+\eta_{k_0-1})(1+\eta_{k_0-2})\|F_{k_0-3}(z_{k_0-2})\| + \|G_0(z_{k_0})\| \\
&\le \cdots \le [\prod_{j=0}^{k_0-1}(1+\eta_j)]\|F_0(z_0)\| + \|G_0(z_{k_0})\| \\
&\le [\prod_{j=0}^{k_0-1}(1+\eta_j)](\|F_0(z_0)\| + \|G_0(z_{k_0})\|).
\end{aligned}
$$

Since, by (2.5), $\|G_0(z_{k_0})\| \le \frac{\sqrt{m}}{2}\epsilon_0 \le \frac{\sqrt{m}}{4}\gamma\|F(z_0)\|$ for all z, it follows that

$$
\|F_0(z_0)\| = \|F(z_0) - G_0(z_0)\| \le \|F(z_0)\| + \|G_0(z_0)\| \le (1 + \frac{\sqrt{m}}{4}\gamma)\|F(z_0)\|
$$

and

$$
\|F_0(z_0)\| + \|G_0(z_{k_0})\| \le (1 + \frac{\sqrt{m}}{4}\gamma)\|F(z_0)\| + \frac{\sqrt{m}}{4}\gamma\|F(z_0)\| = (1 + \frac{\sqrt{m}}{2}\gamma)\|F(z_0)\|.
$$

Therefore, we have

$$
\|F(z_{k_0})\| \le [\prod_{j=0}^{k_0-1}(1+\eta_j)](1 + \frac{\sqrt{m}}{2}\gamma)\|F(z_0)\|, \tag{4.3}
$$

which implies

$$
\|F(z_{k_0})\| \le e^{\eta}(1 + \frac{\sqrt{m}}{2}\gamma)\|F(z_0)\| \le 2e^{\eta}\|F(z_0)\|, \tag{4.4}
$$

where the last inequality holds because $\gamma < \frac{2}{\sqrt{m}}$. The inequalities (4.2) and (4.4) reveal that $z_{k_0} \in \Omega$ and $z_{k_j+1} \in \Omega$ for all $k_j \in \bar{K}_1$. It then remains to show that $z_k \in \Omega$ for all other k. For any of these k, it is clear that there exist $k_j, k_{j+1} \in \bar{K}_1$ satisfying $k_j + 1 < k \le k_{j+1}$. In this case, by Step 6, $\epsilon_k = \epsilon_{k_j+1}$ and $F_k = F_{k_j+1}$, $G_k = G_{k_j+1}$. We deduce again from (3.28) that

$$
\begin{aligned}
\|F(z_k)\| &= \|F_{k_j+1}(z_k) + G_{k_j+1}(z_k)\| \\
&\le \|F_{k_j+1}(z_k)\| + \|G_{k_j+1}(z_k)\| \\
&\le (1+\eta_{k-1})\|F_{k_j+1}(z_{k-1})\| + \|G_{k_j+1}(z_k)\| \\
&\le \cdots \le [\prod_{t=k_j+1}^{k-1}(1+\eta_t)]\|F_{k_j+1}(z_{k_j+1})\| + \|G_{k_j+1}(z_k)\| \\
&\le [\prod_{t=k_j+1}^{k}(1+\eta_t)][\|F(z_{k_j+1})\| + \|G_{k_j+1}(z_{k_j+1})\| + \|G_{k_j+1}(z_k)\|]. \qquad (4.5)
\end{aligned}
$$

However, by (3.27), we have

$$
\|G_{k_j+1}(z_k)\| \le \frac{\sqrt{m}}{2}\gamma\|F(z_{k_j+1})\|, \qquad \|G_{k_j+1}(z_{k_j+1})\| \le \frac{\sqrt{m}}{2}\epsilon_{k_j+1} \le \frac{\sqrt{m}}{2}\gamma\|F(z_{k_j+1})\|.
$$

So, we get from (4.5)

$$\begin{aligned}
\|F(z_k)\| &\leq [\prod_{t=k_j+1}^{k}(1+\eta_t)](1+\sqrt{m}\gamma)\|F(z_{k_j+1})\| \\
&\leq [\prod_{t=k_j+1}^{k}(1+\eta_t)](1+\sqrt{m}\gamma)\|F(z_{k_0})\| \\
&\leq [\prod_{t=0}^{k}(1+\eta_t)](1+\sqrt{m}\gamma)(1+\frac{\sqrt{m}}{2}\gamma)\|F(z_0)\| \\
&\leq 6e^{\eta}\|F(z_0)\|,
\end{aligned}$$

where the second and the third inequalities follow from (4.2) and (4.3), respectively. This means that $z_k \in \Omega$ for all k other than k_0 and $k_j \in \bar{K}_1$.

Summarizing the above discussion, we conclude the proof. □

The following lemma is simple but turns out to be very useful in proving global convergence for Algorithm 1.

Lemma 4.1 *Let $\bar{K}_1$ be finite. Then*

$$\sum_{k=0}^{\infty}\|\alpha_k F_k(z_k)\|^2 < \infty, \tag{4.6}$$

and

$$\sum_{k=0}^{\infty}\|s_k\|^2 = \sum_{k=0}^{\infty}\|x_{k+1}-x_k\|^2 < \infty. \tag{4.7}$$

Proof The assumption that $\bar{K}_1$ is finite implies that there is an index $\bar{k}$ such that $\epsilon_k = \epsilon_{\bar{k}}$ holds for all $k \geq \bar{k}$. It means that for all $k \geq \bar{k}$, F_k is independent of k. We denote $F_k \equiv \bar{F}$ for all $k \geq \bar{k}$. Then by the line search condition (3.14), we have

$$\|\bar{F}(z_{k+1})\|^2 \leq \|\bar{F}(z_k)\|^2 - \sigma_1\|\alpha_k\bar{F}(z_k)\|^2 - \sigma_2\|s_k\|^2 + \eta_k\|\bar{F}(z_k)\|^2,$$

which can be rewritten as

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

Since $\{\|\bar{F}(z_k)\|\}$ is bounded and η_k satisfies (3.13), summing these inequalities from $k = \bar{k}$ to infinity yields (4.6) and (4.7). □

The following lemma comes from [5].

Lemma 4.2 *Let $\{z_k\}$ be generated by Algorithm 1. If*

$$\liminf_{k\to\infty} \|F(z_k)\| \neq 0, \tag{4.8}$$

then there are positive constants β_j, $j = 1,2,3$, such that, for any k sufficiently large, the inequalities

$$\|B_i s_i\| \leq \beta_1 \|s_i\| \quad \textit{and} \quad \beta_2 \|s_i\|^2 \leq s_i^T B_i s_i \leq \beta_3 \|s_i\|^2 \tag{4.9}$$

hold for at least $\lceil \frac{k}{2} \rceil$ values of $i \in \{1,2,\ldots,k\}$.

Proof By Proposition 3.4, (4.8) implies that there exists a positive constant ν such that $\|F_k(z_k)\| \geq \nu$ holds for all $k \geq \bar{k}$, where $\bar{k}$ is a positive integer. It then follows from (3.12) that $y_k^T s_k \geq \|F_k(z_k)\| \, \|s_k\|^2 \geq \nu \|s_k\|^2$ holds for all $k \geq \bar{k}$. Combining this with (3.25), we get (4.9) from Theorem 2.1 in [5]. □

Lemma 4.3 *If $\alpha_k \neq 1$, then*

$$\alpha_k \geq \frac{2\rho(p_k^T B_k p_k - t_k \|p_k\| \|F_k(z_k)\|)}{\sigma_1 \|F_k(z_k)\|^2 + (\sigma_2 + M^2)\|p_k\|^2}, \tag{4.10}$$

where M is an upper bound of $\|\nabla F_k(z)\|$ on D and

$$t_k = \int_0^1 \|\nabla_x L(x_k + \tau\alpha_{k-1} L(z_k), \lambda_k, \mu_k) - \nabla_x L(z_k)\| d\tau + \int_0^1 \|\nabla F_k(z_k + \tau\alpha_k' p_k) - \nabla F_k(z_k)\| d\tau,$$

where $\alpha_k' = \alpha_k/\rho$.

Proof If $\alpha_k \neq 1$, then the steplength α_k is determined in Step 3, and by the line search rule, $\alpha_k' = \alpha_k/\rho$ does not satisfy (3.14), i.e.,

$$\begin{aligned} \|F_k(z_k + \alpha_k' p_k)\|^2 - \|F_k(z_k)\|^2 &> -\sigma_1 \|\alpha_k' F_k(z_k)\|^2 - \sigma_2 \|\alpha_k' p_k\|^2 + \eta_k \|F_k(z_k)\|^2 \\ &\geq -\sigma_1 \|\alpha_k' F_k(z_k)\|^2 - \sigma_2 \|\alpha_k' p_k\|^2 \end{aligned}$$

or equivalently

$$(\alpha_k')^2 (\sigma_1 \|F_k(z_k)\|^2 + \sigma_2 \|p_k\|^2) \geq \|F_k(z_k)\|^2 - \|F_k(x_k + \alpha_k' p_k)\|^2. \tag{4.11}$$

By an elementary deduction we get

$$\begin{aligned} &\|F_k(z_k + \alpha_k' p_k)\|^2 - \|F_k(z_k)\|^2 \\ &= (F_k(z_k + \alpha_k' p_k) + F_k(z_k))^T (F_k(z_k + \alpha_k' p_k) - F_k(z_k)) \\ &= 2F_k(z_k)^T (F_k(z_k + \alpha_k' p_k) - F_k(z_k)) + \|F_k(z_k + \alpha_k' p_k) - F_k(z_k)\|^2 \\ &\leq 2F_k(z_k)^T (F_k(z_k + \alpha_k' p_k) - F_k(z_k)) + M^2 \|\alpha_k' p_k\|^2. \end{aligned} \tag{4.12}$$

For the first term of (4.12), we have

$$
\begin{aligned}
& F_k(z_k)^T(F_k(z_k+\alpha'_k p_k)-F_k(z_k)) \\
&= \alpha'_k F_k(z_k)^T \int_0^1 \nabla F_k(z_k+\tau\alpha'_k p_k)^T p_k d\tau \\
&= \alpha'_k F_k(z_k)^T \nabla F_k(z_k)^T p_k + \alpha'_k F_k(z_k)^T \int_0^1 [\nabla F_k(z_k+\tau\alpha'_k p_k)^T - \nabla F_k(z_k)^T] d\tau p_k \\
&= \alpha'_k q_k^T p_k - \alpha'_k u_k^T p_k + \alpha'_k F_k(z_k)^T \int_0^1 [\nabla F_k(z_k+\tau\alpha'_k p_k)^T - \nabla F_k(z_k)^T] d\tau p_k \\
&= -\alpha'_k p_k^T B_k p_k - \alpha'_k u_k^T p_k + \alpha'_k F_k(z_k)^T \int_0^1 [\nabla F_k(z_k+\tau\alpha'_k p_k)^T - \nabla F_k(z_k)^T] d\tau p_k \\
&\le -\alpha'_k p_k^T B_k p_k + \alpha'_k \|u_k\|\,\|p_k\| + \alpha'_k \|F_k(z_k)\| \int_0^1 \|\nabla F_k(z_k+\tau\alpha'_k p_k) - \nabla F_k(z_k)\| d\tau \|p_k\| \\
&\le -\alpha'_k p_k^T B_k p_k + \alpha'_k \|p_k\|\,\|F_k(z_k)\| \Big(\int_0^1 \|\nabla_x L(x_k+\tau\alpha_{k-1} L(z_k),\lambda_k,\mu_k) - \nabla_x L(z_k)\| d\tau \\
&\quad + \int_0^1 \|\nabla F_k(z_k+\tau\alpha'_k p_k) - \nabla F_k(z_k)\| d\tau \Big) \\
&= -\alpha'_k p_k^T B_k p_k + \alpha'_k t_k \|p_k\|\,\|F_k(z_k)\|, \qquad (4.13)
\end{aligned}
$$

where the third equality follows from (3.20), the fourth equality follows from (3.15), and the last inequality follows from (3.21). Applying (4.13) to (4.12), we get from (4.11)

$$
(\alpha'_k)^2(\sigma_1\|F_k(z_k)\|^2 + \sigma_2\|p_k\|^2) \ge 2\alpha'_k p_k^T B_k p_k - 2\alpha'_k t_k \|p_k\|\|F_k(z_k)\| - M^2\|\alpha'_k p_k\|^2.
$$

Dividing the both sides by α'_k yields

$$
\alpha'_k \ge \frac{2(p_k^T B_k p_k - t_k\|p_k\|\|F_k(z_k)\|)}{\sigma_1\|F_k(z_k)\|^2 + (\sigma_2 + M^2)\|p_k\|^2}.
$$

Since $\alpha_k = \rho\alpha'_k$, the above inequality is equivalent to (4.10). □

Now we prove global convergence of Algorithm 1. Denote

$$
\bar{K} = \{i \mid \text{(4.9) holds}\}. \qquad (4.14)
$$

Theorem 4.2 *Let $\bar{K}$ be defined by (4.14) and $\{z_k\}$ be generated by Algorithm 1. Assume that there exists an accumulation point $\bar{z}$ of $\{z_k\}_{k\in\bar{K}}$ such that $\nabla F_k(\bar{z})$ is nonsingular for all sufficiently large $k \in \bar{K}$. Then*

$$
\liminf_{k\to\infty} \|F(z_k)\| = 0. \qquad (4.15)
$$

Proof From Theorem 4.1 it suffices to consider the case where $\bar{K}_1$ is finite. We may denote $F_k \equiv \bar{F}$ for all k sufficiently large. If $\bar{K}_2 = \{k \mid (3.16) \text{ holds}\}$ is infinite, then (4.15) is trivial. So we assume that both $\bar{K}_1$ and $\bar{K}_2$ are finite.

By (4.6), it suffices to show that there is a subsequence of $\{\alpha_k\}$ with a positive lower bound, i.e., $\limsup_{k\to\infty} \alpha_k > 0$. We assume $\lim_{k\to\infty} \alpha_k = 0$ to deduce a contradiction.

Let $\{z_k\}_{k\in K}$ be a subsequence of $\{z_k\}_{k\in\bar{K}}$ converging to $\bar{z}$. The assumption that $\nabla\bar{F}(\bar{z})$ is nonsingular implies that when $k \in K$ is sufficiently large, $\nabla\bar{F}(z_k)$ is uniformly nonsingular. In particular, there exists a constant $M_1 > 0$ such that $\|\bar{F}(z_k)^{-1}\| \le M_1$ holds for all $k \in K$ large enough. By Lemma 4.2, this together with (3.15) and (3.20) implies that, for at least $\lceil \frac{k}{2} \rceil$ values of $i \in \{1, 2, \ldots, k\}$

$$\begin{aligned}
\|\bar{F}(z_i)\| &= \| - \nabla\bar{F}(z_i)^{-1}(B_i p_i + u_i)\| \\
&\le \|\nabla\bar{F}(z_i)^{-1}\| \left(\|B_i p_i\| + \|u_i\|\right) \\
&\le M_1 \left(\beta_1\|p_i\| + \int_0^1 \|\nabla_x L(x_i + \tau\alpha_{i-1}L(z_i), \lambda_i, \mu_i) - \nabla_x L(z_i)\| d\tau \|\bar{F}(z_i)\|\right) \\
&= M_1(\beta_1\|p_i\| + \nu_i\|\bar{F}(z_i)\|), \qquad (4.16)
\end{aligned}$$

where $\nu_i = \int_0^1 \|\nabla_x L(x_i + \tau\alpha_{i-1}L(z_i), \lambda_i, \mu_i) - \nabla_x L(z_i)\| d\tau$ and the last inequality follows from (3.21) and (4.9). It is obvious that $\nu_k \to 0$. Therefore, (4.16) implies that there is a constant $M_2 > 0$ such that for all $k \in K$ sufficiently large, $\|\bar{F}(z_i)\| \le M_2\|p_i\|$ holds for at least $\lceil \frac{k}{2} \rceil$ values of $i \in \{1, 2, \ldots, k\}$. It follows from (4.10) that

$$\begin{aligned}
\alpha_i &\ge \frac{2\rho(p_i^T B_i p_i - t_i\|p_i\|\|\bar{F}(z_i)\|)}{\sigma_1\|\bar{F}(z_i)\|^2 + (\sigma_2 + M^2)\|p_i\|^2} \\
&\ge \frac{2\rho(\beta_2\|p_i\|^2 - M_2 t_i\|p_i\|^2)}{\sigma_1 M_2^2\|p_i\|^2 + (\sigma_2 + M^2)\|p_i\|^2} \\
&= \frac{2\rho(\beta_2 - M_2 t_i)}{\sigma_1 M_2^2 + (\sigma_2 + M^2)}.
\end{aligned}$$

It is not difficult to see that $t_k \to 0$ as $k \to \infty$ with $k \in K$, since $z_k \to \bar{z}$ as $k \to \infty$ with $k \in K$. Consequently, the above inequality implies that $\{\alpha_k\}_{k\in K}$ contains a subsequence bounded away from zero, which contradicts the assumption that $\lim_{k\to\infty} \alpha_k = 0$. The proof is complete. □

We notice that by Step 6 of Algorithm 1, $\epsilon_k \le \gamma\|F(z_k)\|$ for every k (see Proposition 3.3). Since $\{\epsilon_k\}$ is monotonically nonincreasing, we have

$$\lim_{k\to\infty} \epsilon_k = 0 \qquad (4.17)$$

under the conditions of Theorem 4.2. Now we establish global convergence of Algorithm 1 under slightly different conditions than Theorem 4.2.

Theorem 4.3 *Let $\bar{K}$ be defined by (4.14) and $\{z_k\}$ be generated by Algorithm 1. Assume that there exist accumulation points $\bar{z} = (\bar{x}, \bar{\lambda}, \bar{\mu})$ of $\{z_k\}_{k\in\bar{K}}$ and $\bar{\epsilon}$ of $\{\epsilon_k\}_{k\in\bar{K}}$ such that $\nabla g_i(\bar{x}), i \in$*

$S^{\bar{\epsilon}}(\bar{z})$, and $\nabla h_j(\bar{x}), j = 1,2,\ldots,r$, are linearly independent, where $S^{\epsilon}(z)$ is defined by (2.8). Suppose that $\nabla_x L(\bar{z})$ is positive definite on $\{p \mid \nabla g_i(\bar{x})^T p = 0, \forall i \in S^{\bar{\epsilon}}(\bar{z}), \nabla h(\bar{x})^T p = 0\}$. Then (4.15) holds.

Proof If $\bar{K}_1$ is infinite, then Theorem 4.1 has shown the conclusion. If $\bar{K}_1$ is finite, then when k is sufficiently large, ϵ_k is independent of k. We assume that $\epsilon_k = \bar{\epsilon}$ holds for all $k \geq \bar{k}$ with some positive integer $\bar{k}$. It then follows that $F_k(z) = F_{\bar{k}}(z)$ holds for all $k \geq \bar{k}$. By Lemma 2.4, the conditions in the theorem imply that $\nabla F_k(\bar{z})$ is nonsingular for all $k \geq \bar{k}$. This means that the conditions in Theorem 4.2 hold. □

5. Superlinear Convergence

In this section, we prove superlinear convergence of Algorithm 1. To do this, we need the following further assumptions.

Assumption B (i) The sequence $\{z_k\}$ generated by Algorithm 1 converges to a solution of (1.3), say $\bar{z} = (\bar{x}, \bar{\lambda}, \bar{\mu})$, at which the strict complementarity condition holds, i.e.,

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

(ii) $\nabla F_k(\bar{z})$ is nonsingular for all k sufficiently large.

(iii) $\nabla^2 f$, $\nabla^2 g_i, i = 1,2,\ldots,m$, and $\nabla^2 h_j$, $j = 1,2,\ldots,r$, are Lipschitz continuous at $\bar{x}$, i.e., there is a constant $H > 0$ and a neighborhood $U(\bar{x})$ of $\bar{x}$ such that for all $x \in U(\bar{x})$,

$$\begin{cases} \|\nabla^2 f(x) - \nabla^2 f(\bar{x})\| \leq H\|x - \bar{x}\|, \\ \|\nabla^2 g_i(x) - \nabla^2 g_i(\bar{x})\| \leq H\|x - \bar{x}\|, \quad i = 1,2,\ldots,m, \\ \|\nabla^2 h_j(x) - \nabla^2 h_j(\bar{x})\| \leq H\|x - \bar{x}\|, \quad j = 1,2,\ldots,r. \end{cases} \tag{5.2}$$

Since ∇L is linear in λ and μ, it is clear that condition (iii) in Assumption B implies the Lipschitz continuity of ∇L at $\bar{z} = (\bar{x}, \bar{\lambda}, \bar{\nu})$. It is also obvious that ∇g and ∇h are Lipschitz continuous at $\bar{x}$ since $\nabla^2 g$ and $\nabla^2 h$ are continuous. So, there is a constant $H' > 0$ such that for all z close to $\bar{z}$,

$$\begin{cases} \|\nabla L(z) - \nabla L(\bar{z})\| \leq H'\|z - \bar{z}\|, \\ \|\nabla g(x) - \nabla g(\bar{x})\| = \|\nabla g(x)^T - \nabla g(\bar{x})^T\| \leq H'\|x - \bar{x}\| \leq H'\|z - \bar{z}\|, \\ \|\nabla h(x) - \nabla h(\bar{x})\| = \|\nabla h(x)^T - \nabla h(\bar{x})^T\| \leq H'\|x - \bar{x}\| \leq H'\|z - \bar{z}\|. \end{cases} \tag{5.3}$$

As mentioned just after the proof of Theorem 4.2, we have $\epsilon_k \to 0$. So it is not difficult to see that under condition (i) in Assumption B, $\sqrt{[\lambda_k]_i + g_i(z_k)} \geq \epsilon_k$ holds for all k sufficiently large,

where $[\lambda_k]_i$ denotes the i-th element of λ_k. This means that $J^{\epsilon_k}(x_k, \lambda_k) = \emptyset$ for all sufficiently large k, where $J^{\epsilon}(x, \lambda)$ is defined by (2.6). In other words, for all k sufficiently large, $\Phi_k(z) = \Phi(z)$ and hence $F_k(z) = F(z)$ and $G_k(z) = 0$ for all z close to $\bar{z}$. As a result, conditions (i) and (ii) in Assumption B imply that there is a neighborhood $U(\bar{z})$ of $\bar{z}$ such that $\nabla F(z)$ exists and is continuously differentiable and uniformly nonsingular on $U(\bar{z})$. That is, we have the following lemma.

Lemma 5.1 *Let conditions (i) and (ii) in Assumption B hold. Then there is an index $\bar{k}$ such that for any $k \geq \bar{k}$ and every z in a neighborhood $U(\bar{z})$ of $\bar{z}$, we have $F_k(z) = F(z)$ and $G_k(z) = 0$. Moreover, F is continuously differentiable and $\nabla F(z)$ is uniformly nonsingular on $U(\bar{z})$. In particular, there is a constant $c > 0$ such that for all $k \geq \bar{k}$,*

$$c\|z - \bar{z}\| \leq \|F(z) - F(\bar{z})\| \leq M\|z - \bar{z}\|, \quad \forall z \in U(\bar{z}), \tag{5.4}$$

where $M > 0$ is an upper bound of $\|\nabla F(z)\|$ over D.

Lemma 5.2 *Let Assumptions A and B hold. Then there is a neighborhood $U(\bar{z})$ of $\bar{z}$ and a positive constant $\bar{H}$ such that*

$$\|\nabla F(z) - \nabla F(\bar{z})\| \leq \bar{H}\|z - \bar{z}\|, \quad \forall z \in U(\bar{z}). \tag{5.5}$$

Proof From (2.7), by the strict complementarity condition (5.1), there is a neighborhood of $\bar{z}$ in which $\nabla F(z)$ is represented as

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

where

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

We verify (5.5) by showing that every block in $\nabla F(z)$ is Lipschitz continuous at $\bar{z}$. First, the Lipschitz continuity of $\nabla_x L$, ∇g^T, ∇h^T, ∇g, and ∇h follows from (5.3) directly. It then suffices to show that the second column block of $\nabla F(z)$ satisfies the Lipschitz condition. By an elementary deduction we have

$$\begin{aligned} & \|\nabla g(x)\ \mathrm{diag}\ (b_i(x,\lambda)) - \nabla g(\bar{x})\ \mathrm{diag}\ (b_i(\bar{x},\bar{\lambda}))\| \\ \leq\ & \|\nabla g(x) - \nabla g(\bar{x})\|\ \|\ \mathrm{diag}\ (b_i(x,\lambda))\| + \|\nabla g(\bar{x})\|\ \|\ \mathrm{diag}\ (b_i(x,\lambda)) - \ \mathrm{diag}\ (b_i(\bar{x},\bar{\lambda}))\| \\ \leq\ & 2H'\|x - \bar{x}\| + \|\nabla g(\bar{x})\|\ \|\ \mathrm{diag}\ (b_i(x,\lambda)) - \ \mathrm{diag}\ (b_i(\bar{x},\bar{\lambda}))\|, \end{aligned} \tag{5.6}$$

where the last inequality follows from Lemma 2.2 and (5.3). For the second term of (5.6), we have

$$\begin{aligned} & |b_i(x,\lambda) - b_i(\bar{x},\bar{\lambda})| \\ = & \Big| \frac{g_i(x)}{\sqrt{\lambda_i^2 + g_i(x)^2}} - \frac{g_i(\bar{x})}{\sqrt{\bar{\lambda}_i^2 + g_i(\bar{x})^2}} \Big| \\ = & \frac{|\, g_i(x)\sqrt{\bar{\lambda}_i^2 + g_i(\bar{x})^2} - g_i(\bar{x})\sqrt{\lambda_i^2 + g_i(x)^2}\,|}{\sqrt{\lambda_i^2 + g_i(x)^2}\sqrt{\bar{\lambda}_i^2 + g_i(\bar{x})^2}} \\ \le & \frac{|\, g_i(x) - g(\bar{x})|\sqrt{\bar{\lambda}_i^2 + g_i(\bar{x})^2} + |\, g_i(\bar{x})\,|\,|\sqrt{\bar{\lambda}_i^2 + g_i(\bar{x})^2} - \sqrt{\lambda_i^2 + g_i(x)^2}\,|}{\sqrt{\lambda_i^2 + g_i(x)^2}\sqrt{\bar{\lambda}_i^2 + g_i(\bar{x})^2}}. \end{aligned} \tag{5.7}$$

Since g_i is continuously differentiable, there is a constant $H_1 > 0$ such that for all z near $\bar{z}$,

$$|\, g_i(x) - g_i(\bar{x})| \le H_1\|x - \bar{x}\| \le H_1\|z - \bar{z}\|.$$

We also have

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

By the strict complementarity condition (5.1), there is a neighborhood of $\bar{z}$ in which $\sqrt{\lambda_i^2 + g_i(x)^2}$ is bounded away from zero. Therefore, the last inequality shows that there is a positive constant H_2 such that for all z near $\bar{z}$,

$$\Big|\, \sqrt{\bar{\lambda}_i^2 + g_i(\bar{x})^2} - \sqrt{\lambda_i^2 + g_i(x)^2}\,\Big| \le H_2\Big(\|x - \bar{x}\| + \|\lambda - \bar{\lambda}\|\Big) \le H_2\|z - \bar{z}\|.$$

Combining the above discussion with (5.7), we claim that there is a constant $H_3 > 0$ such that for all z near $\bar{z}$,

$$|b_i(x,\lambda) - b_i(\bar{x},\bar{\lambda})| \le H_3\|z - \bar{z}\|, \qquad i = 1, 2, \ldots, m.$$

This together with (5.6) reveals that there is a positive constant H_4 such that for all z near $\bar{z}$,

$$\|\nabla g(x)\ \mathrm{diag}\ (b_i(x,\lambda)) - \nabla g(\bar{x})\ \mathrm{diag}\ (b_i(\bar{x},\bar{\lambda}))\| \le H_4\|z - \bar{z}\|.$$

Similarly, there is a positive constant H_5 such that for all z near $\bar{z}$,

$$\|\ \mathrm{diag}\ (a_i(x,\lambda)) -\ \mathrm{diag}\ (a_i(\bar{x},\bar{\lambda}))\| \le H_5\|z - \bar{z}\|.$$

Hence the second column block of $\nabla F(z)$ satisfies the Lipschitz condition. This completes the proof. □

Lemma 5.3 *Let Assumptions A and B hold. Then $\tau_k = 1$ holds for all k large enough, where τ_k is defined by (3.11). Moreover, there exist positive constants c_1, c_2, C_1, C_2 and M_3 such that for all k sufficiently large,*

$$c_1\|s_k\|^2 \le y_k^T s_k \le C_1\|s_k\|^2, \tag{5.8}$$

$$c_2\|s_k\| \le \|y_k\| \le C_2\|s_k\|, \tag{5.9}$$

and

$$\|y_k - \nabla F(z_k)\nabla F(z_k)^T s_k\| \le M_3\Big(\|F(z_k)\| + \nu_k\Big)\|s_k\|, \tag{5.10}$$

where

$$\nu_k = \max\{\|z_{k+1} - \bar{z}\|, \|z_k - \bar{z}\|\}. \tag{5.11}$$

Proof We first show that there is a positive constant C_3 such that when k is sufficiently large,

$$\|\bar{v}_k\| \le C_3\nu_k\|s_k\|, \tag{5.12}$$

where $\bar{v}_k$ is defined by (3.9). Indeed, from (3.23), we deduce that for all k sufficiently large,

$$\begin{aligned}
\|\bar{v}_k\| &\le M\Big(\int_0^1 \|\nabla F(z_k + \tau s_k) - \nabla F(\bar{z})\|d\tau + \|\nabla F(z_k) - \nabla F(\bar{z})\| \\
&\quad + \int_0^1 \|\nabla_x L(x_k + \tau(L(z_{k+1}) - L(z_k)), \lambda_k, \mu_k) - \nabla_x L(\bar{z})\|d\tau + \|\nabla_x L(z_k) - \nabla_x L(\bar{z})\|\Big)\|s_k\| \\
&\le M\Big[\bar{H}\Big(\|z_k - \bar{z}\| + \|s_k\| + \|z_k - \bar{z}\|\Big) + H'\Big(\|z_k - \bar{z}\| + \|L(z_{k+1}) - L(z_k)\| + \|z_k - \bar{z}\|\Big)\Big]\|s_k\| \\
&\le M\Big[\bar{H}\Big(2\|z_k - \bar{z}\| + \|z_{k+1} - z_k\|\Big) + H'\Big(2\|z_k - \bar{z}\| + M\|z_{k+1} - z_k\|\Big)\Big]\|s_k\| \\
&\le M\Big[\bar{H}\Big(3\|z_k - \bar{z}\| + \|z_{k+1} - \bar{z}\|\Big) + H'\Big((2+M)\|z_k - \bar{z}\| + M\|z_{k+1} - \bar{z}\|\Big)\Big]\|s_k\| \\
&\le M[4\bar{H} + 2H'(1+M)]\nu_k\|s_k\|,
\end{aligned}$$

where M is an upper bound of $\|\nabla F(z)\|$ on D (and hence an upper bound of $\|\nabla L(z)\|$ on D) and the second inequality follows from (5.3) and (5.5). Putting $C_3 = M[4\bar{H} + 2H'(1+M)]$, we get (5.12).

Next we show that $\tau_k = 1$ for all k sufficiently large. By the nonsingularity of $\nabla F(\bar{z})$, there is a positive constant c_3 such that when k is sufficiently large

$$\|\nabla F(z_k)p\| \ge c_3\|p\|, \qquad \|\nabla F(z_k)^T p\| \ge c_3\|p\|, \qquad \forall p \in R^{n+m+r}. \tag{5.13}$$

From (3.22), (5.12) and (5.13), we get

$$\begin{aligned} r_k^T s_k &= s_k^T \nabla F(z_k) \nabla F(z_k)^T s_k + s_k^T \bar{v}_k \\ &\geq \|\nabla F(z_k)^T s_k\|^2 - \|s_k\| \, \|\bar{v}_k\| \\ &\geq c_3^2 \|s_k\|^2 - C_3 \nu_k \|s_k\|^2 \\ &= (c_3^2 - C_3 \nu_k) \|s_k\|^2. \end{aligned}$$

Since $\nu_k \to 0$, it follows that when k is sufficiently large,

$$r_k^T s_k \geq \frac{1}{2} c_3^2 \|s_k\|^2 > 0,$$

which, by the definition (3.11) of τ_k, implies that $\tau_k = 1$ holds for all k sufficiently large. Moreover, we have

$$y_k^T s_k = r_k^T s_k + \|F(z_k)\| \, s_k^T s_k \geq r_k^T s_k \geq \frac{1}{2} c_3^2 \|s_k\|^2.$$

Putting $c_1 = \frac{1}{2} c_3^2$ yields the left-hand inequality of (5.8). We also get

$$\begin{aligned} \|y_k\| &= \left\| r_k + \|F(z_k)\| s_k \right\| \\ &\leq \|r_k\| + \|F(z_k)\| \, \|s_k\| \\ &\leq \|\nabla F(z_k) \nabla F(z_k)^T s_k\| + \|\bar{v}_k\| + \|F(z_k)\| \, \|s_k\| \\ &\leq (M^2 + C_3 \nu_k + \bar{M}) \|s_k\|, \end{aligned}$$

where $\bar{M}$ is an upper bound of $\|F(z)\|$ on D. The above inequalities show the right-hand inequality of (5.9). Moreover, we have

$$y_k^T s_k \leq \|y_k\| \, \|s_k\| \leq (M^2 + C_3 \nu_k + \bar{M}) \|s_k\|^2.$$

This implies the right-hand inequality of (5.8). Also, we have

$$\begin{aligned} \|y_k\| &= \left\| r_k + \|F(z_k)\| s_k \right\| \\ &\geq \|r_k\| - \|F(z_k)\| \, \|s_k\| \\ &\geq \|\nabla F(z_k) \nabla F(z_k)^T s_k\| - \|\bar{v}_k\| - \|F(z_k)\| \, \|s_k\| \\ &\geq (c_3^2 - C_3 \nu_k - \|F(z_k)\|) \|s_k\|, \end{aligned}$$

where the last inequality follows from (5.12) and (5.13). Since $\nu_k \to 0$ and $F(z_k) \to F(\bar{z}) = 0$, the last inequality implies that there is a constant $c_2 > 0$ such that the left-hand inequality of (5.9) holds for all k sufficiently large.

Now we verify (5.10). Observe that

$$\begin{aligned}\|y_k - \nabla F(z_k)\nabla F(z_k)^T s_k\| &= \left\|r_k + \|F(z_k)\|s_k - \nabla F(z_k)\nabla F(z_k)^T s_k\right\| \\ &\leq \|F(z_k)\|\,\|s_k\| + \|r_k - \nabla F(z_k)\nabla F(z_k)^T s_k\| \\ &= \|F(z_k)\|\,\|s_k\| + \|\bar{v}_k\| \\ &\leq (\|F(z_k)\| + C_3\nu_k)\|s_k\|,\end{aligned}$$

where the last equality follows from (3.22) and the last inequality follows from (5.12). Then (5.10) is obtained by letting $M_3 = \max\{1, C_3\}$. □

The next lemma follows from Theorem 4.2.

Lemma 5.4 *Let Assumptions A and B hold. Then for every $k \in \bar{K}$, either (3.16) holds or $\alpha_k \geq \tilde{\alpha}$ holds, where $\bar{K}$ is defined by (4.14) and $\tilde{\alpha} > 0$ is some constant.*

Proof By Lemma 4.3, for every k large enough, if $\alpha_k \neq 1$, then

$$\alpha_k \geq \frac{2\rho(p_k^T B_k p_k - t_k\|p_k\|\,\|F(z_k)\|)}{\sigma_1\|F_k(z_k)\|^2 + (\sigma_2 + M^2)\|p_k\|^2}.$$

Recall the definition (4.14) of $\bar{K}$. Then in a similar way to the proof for (4.16), we deduce that there is a positive constant M_0 such that for all $i \in \bar{K}$ sufficiently large, $\|F(z_i)\| \leq M_0\|p_i\|$. It then follows

$$\begin{aligned}\alpha_i &\geq \frac{2\rho(p_i^T B_i p_i - t_i\|p_i\|\,\|F(z_i)\|)}{\sigma_1\|\bar{F}(z_i)\|^2 + (\sigma_2 + M^2)\|p_i\|^2} \\ &\geq \frac{2\rho(\beta_2\|p_i\|^2 - M_0 t_i\|p_i\|^2)}{\sigma_1 M_2^2\|p_i\|^2 + (\sigma_2 + M^2)\|p_i\|^2} \\ &= \frac{2\rho(\beta_2 - M_0 t_i)}{\sigma_1 M_2^2 + (\sigma_2 + M^2)}.\end{aligned}$$

Since $t_i \to 0$, the above inequalities show that when $i \in \bar{K}$ is sufficiently large, α_i is bounded away from zero. In other words, there is a positive constant $\bar{\alpha}$ and an index $\bar{i}$ such that $\alpha_i \geq \min\{\bar{\alpha}, 1\}$ holds for all $i \in \bar{K}$ with $i \geq \bar{i}$. □

Lemma 5.5 *Let Assumptions A and B hold. Then we have*

$$\sum_{k=0}^{\infty} \|z_k - \bar{z}\| < \infty. \tag{5.14}$$

Proof It is easy to see by Algorithm 1 that for every i large enough, we have either

$$\|F_i(z_{i+1})\|^2 = \|F(z_{i+1})\|^2 \leq \beta^2\|F(z_i)\|^2, \qquad \text{for } i \in \bar{K}_2 \tag{5.15}$$

or

$$\begin{aligned} \|F(z_{i+1})\|^2 &\leq (1-\sigma_1\alpha_i^2+\eta_i)\|F(z_i)\|^2 - \sigma_2\|s_i\|^2 \\ &\leq (1-\sigma_1\alpha_i^2+\eta_i)\|F(z_i)\|^2, \quad \text{for } i \notin \bar{K}_2, \end{aligned} \tag{5.16}$$

where $\bar{K}_2$ is defined by (3.26). Let $\bar{K}$ be defined by (4.14). Then, from Lemma 5.4, for each $i \in \bar{K}$, either (5.15) holds or $\alpha_i \geq \tilde{\alpha}$ for some constant $\tilde{\alpha} > 0$. In the latter case, it follows from (5.16) that

$$\|F(z_{i+1})\|^2 \leq (1+\eta_i-\sigma_1\tilde{\alpha}^2)\|F(z_i)\|^2.$$

Since $\eta_i \to 0$, there are an index i_0 and a constant $\theta_1 \in (0,1)$ such that $1+\eta_i-\sigma_1\tilde{\alpha}^2 \leq \theta_1$ holds for all $i \geq i_0$. Let $\theta_2 = \max\{\beta^2, \theta_1\} < 1$. Then

$$\|F(z_{i+1})\|^2 \leq \theta_2\|F(z_i)\|^2 \tag{5.17}$$

holds for all $i \in \bar{K}$ with $i \geq i_0$. However, Lemma 4.2 shows that the number of elements in $\bar{K}$ is at least $\lceil \frac{k}{2} \rceil$. Therefore, for any $k > 2i_0$, there are at least $\lceil \frac{k}{2} \rceil - i_0$ of indices i such that (5.17) holds. Let K_3 denote the set of indices i for which (5.17) holds and N_k denote the number of indices in K_3 not exceeding k. Then we have $N_k \geq \lceil \frac{k}{2} \rceil - i_0$ for each k large enough. Since (5.15) or (5.16) holds for all i large enough, multiplying (5.16) for $i \notin K_3$ and (5.17) for $i \in K_3$ from $i = i_0$ to $i = k$ yields

$$\begin{aligned} \|F(z_{k+1})\|^2 &\leq \Big[\prod_{\substack{i=i_0 \\ i \notin K_3}}^{k} (1+\eta_i-\sigma_1\alpha_i^2) \Big] \theta_2^{N_k} \|F(x_{i_0})\|^2 \\ &\leq \Big[\prod_{\substack{i=i_0 \\ i \notin K_3}}^{k} (1+\eta_i) \Big] \theta_2^{N_k} \|F(x_{i_0})\|^2 \\ &\leq \Big[\prod_{i=0}^{k} (1+\eta_i) \Big] \theta_2^{\lceil \frac{k}{2} \rceil - i_0} \|F(x_{i_0})\|^2 \\ &\leq e^{\eta} \theta_2^{\frac{k}{2} - (i_0+1)} \|F(x_{i_0})\|^2 \\ &= c_0 \tilde{\theta}^k, \end{aligned}$$

where $c_0 = e^{\eta}\theta_2^{-(i_0+1)}\|F(x_{i_0})\|^2$ and $\tilde{\theta} = \theta_2^{\frac{1}{2}} \in (0,1)$. Since $z_k \to \bar{z}$, it follows from (5.4) and the above inequalities that $\|z_{k+1} - \bar{z}\|^2 \leq c^{-2} c_0 \tilde{\theta}^k$ holds for all k large enough. Hence we get (5.14) as desired. □

The subsequent analysis follows a similar line to that of [23]. First, we verify the following lemma.

Lemma 5.6 *Let Assumptions A and B hold. If*

$$\lim_{k\to\infty} \frac{\|(B_k - \nabla F(\bar{z})\nabla F(\bar{z})^T)p_k\|}{\|p_k\|} = 0, \tag{5.18}$$

then $\alpha_k = 1$ for all k sufficiently large. Moreover, $\{z_k\}$ converges superlinearly.

Proof Denote

$$\xi_k = \frac{\|(B_k - \nabla F(\bar{z})\nabla F(\bar{z})^T)p_k\|}{\|p_k\|}. \tag{5.19}$$

Then we have from (3.15) and (3.20)

$$\begin{aligned}
& \nabla F(\bar{z})\nabla F(\bar{z})^T(z_k + p_k - \bar{z}) \\
= \;& \nabla F(\bar{z})\nabla F(\bar{z})^T(z_k - \bar{z}) + \nabla F(\bar{z})\nabla F(\bar{z})^T p_k \\
= \;& \nabla F(\bar{z})\nabla F(\bar{z})^T(z_k - \bar{z}) - q_k + (\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)p_k \\
= \;& \nabla F(\bar{z})\nabla F(\bar{z})^T(z_k - \bar{z}) - \nabla F(z_k)(F(z_k) - F(\bar{z})) + (\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)p_k - u_k \\
= \;& (\nabla F(\bar{z})\nabla F(\bar{z})^T - \nabla F(z_k)\overline{A}_k)(z_k - \bar{z}) + (\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)p_k - u_k,
\end{aligned} \tag{5.20}$$

where u_k is defined by (3.6) and $\overline{A}_k = \int_0^1 \nabla F(\bar{z} + \tau(z_k - \bar{z}))^T d\tau$. From (3.21) and (5.3), we have

$$\begin{aligned}
\|u_k\| &\le \Big(\int_0^1 \|\nabla_x L(x_k + \tau\alpha_{k-1}L(z_k), \lambda_k, \mu_k) - \nabla_x L(\bar{z})\| d\tau + \|\nabla_x L(z_k) - \nabla_x L(\bar{z})\|\Big)\|F(z_k)\| \\
&\le H'\Big(\|z_k - \bar{z}\| + \|\alpha_{k-1}L(z_k)\| + \|z_k - \bar{z}\|\Big)\|F(z_k)\| \\
&\le H'\Big(2\|z_k - \bar{z}\| + \|L(z_k) - L(\bar{z})\|\Big)\|F(z_k)\| \\
&\le H'(2 + M)\|z_k - \bar{z}\|\,\|F(z_k)\| \\
&\le C_4\nu_k\|F(z_k)\|,
\end{aligned}$$

where $C_4 = H'(2+M)$ and the third inequality follows form the fact that $L(\bar{z}) = 0$ and $\alpha_{k-1} \le 1$ for every k. Thus we get from (3.15), (3.20) and (5.19)

$$\begin{aligned}
\|\nabla F(\bar{z})\nabla F(\bar{z})^T p_k\| &= \|(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)p_k - q_k\| \\
&\le \|(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)p_k\| + \|q_k\| \\
&\le \xi_k\|p_k\| + \|\nabla F(z_k)F(z_k)\| + \|u_k\| \\
&\le \xi_k\|p_k\| + \|\nabla F(z_k)F(z_k)\| + C_4\nu_k\|F(z_k)\|.
\end{aligned} \tag{5.21}$$

Since $\nabla F(\bar{z})$ is nonsingular, there exists a positive constant c_4 such that $\|\nabla F(\bar{z})p\| \ge c_4\|p\|$ and $\|\nabla F(\bar{z})^T p\| \ge c_4\|p\|$ holds for all $p \in R^n$, which in turn implies that $\|\nabla F(\bar{z})\nabla F(\bar{z})^T p_k\| \ge c_4^2\|p_k\|$. It then follows from (5.21) that

$$(c_4^2 - \xi_k)\|p_k\| \le (\|\nabla F(z_k)\| + C_4\nu_k)\|F(z_k)\|.$$

By the assumption that $\xi_k \to 0$ and the fact that $\nu_k \to 0$ and $\nabla F(z_k) \to \nabla F(\bar{z})$, we claim that there exists a positive constant M_4 such that for all k sufficiently large

$$\|p_k\| \le M_4\|F(z_k)\| = M_4\|F(z_k) - F(\bar{z})\| \le MM_4\|z_k - \bar{z}\|.$$

Therefore, taking the norm operation in (5.20) yields

$$\begin{aligned}
& \|\nabla F(\bar{z})\nabla F(\bar{z})^T(z_k + p_k - \bar{z})\| \\
\le\ & \|(\nabla F(\bar{z})\nabla F(\bar{z})^T - \nabla F(z_k)\overline{A}_k)(z_k - \bar{z})\| + \|(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)p_k\| + \|u_k\| \\
\le\ & \|\nabla F(\bar{z})\nabla F(\bar{z})^T - \nabla F(z_k)\overline{A}_k\|\,\|z_k - \bar{z}\| + \xi_k\|p_k\| + C_4\nu_k\|F(z_k) - F(\bar{z})\| \\
\le\ & \|\nabla F(\bar{z})\nabla F(\bar{z})^T - \nabla F(z_k)\overline{A}_k\|\|z_k - \bar{z}\| + \xi_k M_4\|z_k - \bar{z}\| + C_4M\nu_k\|z_k - \bar{z}\| \\
=\ & o(\|z_k - \bar{z}\|). \qquad (5.22)
\end{aligned}$$

This together with (5.13) yields

$$\frac{\|z_k + p_k - \bar{z}\|}{\|z_k - \bar{z}\|} \to 0. \qquad (5.23)$$

Moreover, we have

$$\begin{aligned}
\|F(z_k + p_k)\| &= \|F(z_k + p_k) - F(\bar{z})\| \\
&\le M\|z_k + p_k - \bar{z}\| \\
&= \frac{M}{c}\frac{\|z_k + p_k - \bar{z}\|}{\|z_k - \bar{z}\|}c\|z_k - \bar{z}\| \\
&\le \frac{M}{c}\frac{\|z_k + p_k - \bar{z}\|}{\|z_k - \bar{z}\|}\|F(z_k) - F(\bar{z})\| \\
&= \frac{M}{c}\frac{\|z_k + p_k - \bar{z}\|}{\|z_k - \bar{z}\|}\|F(z_k)\|,
\end{aligned}$$

where the last inequality follows from (5.4). This and (5.23) indicate that (3.16) is satisfied for all k sufficiently large. In other words, the unit steplength is accepted for all k sufficiently large. Consequently, (5.23) implies the superlinear convergence of $\{z_k\}$. □

Lemma 5.6 shows that to establish superlinear convergence of Algorithm 1, it suffices to verify that $\{z_k\}$ satisfies the Dennis-Moré condition (5.18). In the rest of this section, we devote ourselves to showing that this is true for Algorithm 1.

Denote $P = [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1/2}$. For an $n \times n$ matrix A, define a matrix norm $\|A\|_P = \|PAP\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. We let H_k and H_{k+1} stand for the inverse matrices of B_k and B_{k+1}, respectively. The following lemma is similar to Lemma 3.6 in [23], which shows that the BFGS formula (3.17) exhibits a similar property to that of the conventional BFGS formula. For completeness, we give a proof.

Lemma 5.7 *Under Assumptions A and B, there exist positive constants M_5, M_6, M_7 and $\beta \in (0,1)$ such that for all k sufficiently large*

$$\|B_{k+1} - \nabla F(\bar{z})\nabla F(\bar{z})^T\|_P \le \|B_k - \nabla F(\bar{z})\nabla F(\bar{z})^T\|_P + M_5\nu_k, \tag{5.24}$$

$$\|H_{k+1} - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}} \le (1 - \tfrac{1}{2}\beta\zeta_k^2 + M_6\nu_k)\|H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}} + M_7\nu_k, \tag{5.25}$$

where ν_k is defined by (5.11) and ζ_k is given by

$$\zeta_k = \frac{\|P^{-1}[H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}]y_k\|}{\|H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}}\|Py_k\|}. \tag{5.26}$$

In particular, $\{\|B_k\|\}$ and $\{\|H_k\|\}$ are bounded.

Proof From the update formula (3.17), we get

$$\begin{aligned} & P(B_{k+1} - \nabla F(\bar{z})\nabla F(\bar{z})^T)P \\ = \; & P(B_k - \nabla F(\bar{z})\nabla F(\bar{z}))P - \frac{[(PB_kP)(P^{-1}s_k)][(PB_kP)(P^{-1}s_k)]^T}{(P^{-1}s_k)^T(PB_kP)(P^{-1}s_k)} \\ & + \frac{(Py_k)(Py_k)^T}{(Py_k)^T(P^{-1}s_k)}. \end{aligned} \tag{5.27}$$

Denote $\tilde{B}_k = PB_kP$, $\tilde{s}_k = P^{-1}s_k$, $\tilde{y}_k = Py_k$, $Q_{k+1} = P(B_{k+1} - \nabla F(\bar{z})\nabla F(\bar{z})^T)P = PB_{k+1}P - I$, and $Q_k = P(B_k - \nabla F(\bar{z})\nabla F(\bar{z})^T)P = \tilde{B}_k - I$. Then Q_k, Q_{k+1} and $\tilde{B}_k$ are symmetric, and (5.27) is rewritten as

$$Q_{k+1} = Q_k - \frac{\tilde{B}_k\tilde{s}_k\tilde{s}_k^T\tilde{B}_k}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k} + \frac{\tilde{y}_k\tilde{y}_k^T}{\tilde{y}_k^T\tilde{s}_k}.$$

Taking the norm operation on the both sides, we get

$$\|Q_{k+1}\|_F \le \Big\|Q_k - \frac{\tilde{B}_k\tilde{s}_k\tilde{s}_k^T\tilde{B}_k}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k} + \frac{\tilde{s}_k\tilde{s}_k^T}{\|\tilde{s}_k\|^2}\Big\|_F + \Big\|\frac{\tilde{s}_k\tilde{s}_k^T}{\|\tilde{s}_k\|^2} - \frac{\tilde{y}_k\tilde{y}_k^T}{\tilde{y}_k^T\tilde{s}_k}\Big\|_F. \tag{5.28}$$

We estimate the two terms on the right-hand side of (5.28). For the first term, we have

$$\begin{aligned} & \Big\|Q_k - \frac{\tilde{B}_k\tilde{s}_k\tilde{s}_k^T\tilde{B}_k}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k} + \frac{\tilde{s}_k\tilde{s}_k^T}{\|\tilde{s}_k\|^2}\Big\|_F^2 \\ & = \mathrm{trace}\Big\{\Big(Q_k - \frac{\tilde{B}_k\tilde{s}_k\tilde{s}_k^T\tilde{B}_k}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k} + \frac{\tilde{s}_k\tilde{s}_k^T}{\|\tilde{s}_k\|^2}\Big)\Big(Q_k - \frac{\tilde{B}_k\tilde{s}_k\tilde{s}_k^T\tilde{B}_k}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k} + \frac{\tilde{s}_k\tilde{s}_k^T}{\|\tilde{s}_k\|^2}\Big)^T\Big\} \\ & = \mathrm{trace}\Big(Q_k^2 + \frac{\|\tilde{B}_k\tilde{s}_k\|^2\tilde{B}_k\tilde{s}_k\tilde{s}_k^T\tilde{B}_k}{(\tilde{s}_k^T\tilde{B}_k\tilde{s}_k)^2} + \frac{\tilde{s}_k\tilde{s}_k^T}{\|\tilde{s}_k\|^2} - \frac{Q_k\tilde{B}_k\tilde{s}_k\tilde{s}_k^T\tilde{B}_k + \tilde{B}_k\tilde{s}_k\tilde{s}_k^T\tilde{B}_kQ_k}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k} \\ & \quad - \frac{\tilde{B}_k\tilde{s}_k\tilde{s}_k^T + \tilde{s}_k\tilde{s}_k^T\tilde{B}_k}{\|\tilde{s}_k\|^2} + \frac{Q_k\tilde{s}_k\tilde{s}_k^T + \tilde{s}_k\tilde{s}_k^TQ_k}{\|\tilde{s}_k\|^2}\Big) \\ & = \|Q_k\|_F^2 + \frac{\|\tilde{B}_k\tilde{s}_k\|^4}{(\tilde{s}_k^T\tilde{B}_k\tilde{s}_k)^2} + 1 - 2\frac{\tilde{s}_k^T\tilde{B}_kQ_k\tilde{B}_k\tilde{s}_k}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k} - 2\frac{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k}{\|\tilde{s}_k\|^2} + 2\frac{\tilde{s}_k^TQ_k\tilde{s}_k}{\|\tilde{s}_k\|^2}. \end{aligned} \tag{5.29}$$

By the definition of Q_k and $\tilde{B}_k$, we have

$$\begin{aligned}\tilde{s}_k^T\tilde{B}_kQ_k\tilde{B}_k\tilde{s}_k &= \tilde{s}_k^T\tilde{B}_k(\tilde{B}_k-I)\tilde{B}_k\tilde{s}_k\\ &= \tilde{s}_k^T\tilde{B}_k^3\tilde{s}_k-\tilde{s}_k^T\tilde{B}_k^2\tilde{s}_k\\ &= \tilde{s}_k^T\tilde{B}_k^3\tilde{s}_k-\|\tilde{B}_k\tilde{s}_k\|^2\end{aligned}$$

and

$$\tilde{s}_k^TQ_k\tilde{s}_k=\tilde{s}_k^T\tilde{B}_k\tilde{s}_k-\|\tilde{s}_k\|^2.$$

So, we get from (5.29)

$$\begin{aligned}&\Big\|Q_k-\frac{\tilde{B}_k\tilde{s}_k\tilde{s}_k^T\tilde{B}_k}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k}+\frac{\tilde{s}_k\tilde{s}_k^T}{\|\tilde{s}_k\|^2}\Big\|_F^2\\ &= \|Q_k\|_F^2+\frac{\|\tilde{B}_k\tilde{s}_k\|^4}{(\tilde{s}_k^T\tilde{B}_k\tilde{s}_k)^2}+1-2\frac{\tilde{s}_k^T\tilde{B}_k^3\tilde{s}_k}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k}+2\frac{\|\tilde{B}_k\tilde{s}_k\|^2}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k}-2\frac{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k}{\|\tilde{s}_k\|^2}+2\frac{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k}{\|\tilde{s}_k\|^2}-2\\ &= \|Q_k\|_F^2+2\Big[\Big(\frac{\|\tilde{B}_k\tilde{s}_k\|^2}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k}\Big)^2-\frac{\tilde{s}_k^T\tilde{B}_k^3\tilde{s}_k}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k}\Big]-\Big[\Big(\frac{\|\tilde{B}_k\tilde{s}_k\|^2}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k}\Big)^2-2\frac{\|\tilde{B}_k\tilde{s}_k\|^2}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k}+1\Big]\\ &= \|Q_k\|_F^2+2\Big[\Big(\frac{\|\tilde{B}_k\tilde{s}_k\|^2}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k}\Big)^2-\frac{\tilde{s}_k^T\tilde{B}_k^3\tilde{s}_k}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k}\Big]-\Big[1+\frac{\|\tilde{B}_k\tilde{s}_k\|^2}{\tilde{s}_k^T\tilde{B}_k\tilde{s}_k}\Big]^2\\ &\le \|Q_k\|_F^2,\end{aligned}\tag{5.30}$$

where the inequality holds because

$$\|\tilde{B}_k\tilde{s}_k\|^2=[\tilde{B}_k^{\frac{3}{2}}\tilde{s}_k]^T[\tilde{B}_k^{\frac{1}{2}}\tilde{s}_k]\le\|\tilde{B}_k^{\frac{3}{2}}\tilde{s}_k\|\|\tilde{B}_k^{\frac{1}{2}}\tilde{s}_k\|=[\tilde{s}_k^T\tilde{B}_k^3\tilde{s}_k]^{\frac{1}{2}}[\tilde{s}_k^T\tilde{B}_k\tilde{s}_k]^{\frac{1}{2}}.$$

For the second term of (5.28) we have

$$\begin{aligned}\Big\|\frac{\tilde{s}_k\tilde{s}_k^T}{\|\tilde{s}_k\|^2}-\frac{\tilde{y}_k\tilde{y}_k^T}{\tilde{y}_k^T\tilde{s}_k}\Big\|_F &\le \Big\|\Big(\frac{1}{\|\tilde{s}_k\|^2}-\frac{1}{\tilde{y}_k^T\tilde{s}_k}\Big)\tilde{s}_k\tilde{s}_k^T\Big\|_F+\Big\|\frac{\tilde{s}_k\tilde{s}_k^T-\tilde{y}_k\tilde{y}_k^T}{\tilde{y}_k^T\tilde{s}_k}\Big\|_F\\ &\le \frac{|\tilde{y}_k^T\tilde{s}_k-\|\tilde{s}_k\|^2|}{\tilde{y}_k^T\tilde{s}_k}+\frac{\|\tilde{s}_k(\tilde{s}_k-\tilde{y}_k)^T\|_F+\|(\tilde{s}_k-\tilde{y}_k)\tilde{y}_k^T\|_F}{\tilde{y}_k^T\tilde{s}_k}\\ &\le \frac{\|\tilde{y}_k-\tilde{s}_k\|\,\|\tilde{s}_k\|+\|\tilde{s}_k\|\,\|\tilde{s}_k-\tilde{y}_k\|+\|\tilde{s}_k-\tilde{y}_k\|\,\|\tilde{y}_k\|}{\tilde{y}_k^T\tilde{s}_k}\\ &= \frac{(2\|P^{-1}s_k\|+\|Py_k\|)\|Py_k-P^{-1}s_k\|}{y_k^Ts_k}.\end{aligned}\tag{5.31}$$

By the definition of P, we get

$$\begin{aligned}&\|Py_k-P^{-1}s_k\|\\ \le\ &\|P\|\,\|y_k-P^{-2}s_k\|\\ =\ &\|P\|\,\|y_k-\nabla F(\bar{z})\nabla F(\bar{z})^Ts_k\|\\ \le\ &\|P\|\Big(\|y_k-\nabla F(z_k)\nabla F(z_k)^Ts_k\|+\|(\nabla F(z_k)\nabla F(z_k)^T-\nabla F(\bar{z})\nabla F(\bar{z})^T)s_k\|\Big)\\ \le\ &\Big(M_3(\nu_k+\|F(z_k)\|)+\|\nabla F(z_k)\nabla F(z_k)^T-\nabla F(\bar{z})\nabla F(\bar{z})^T\|\Big)\|P\|\,\|s_k\|,\end{aligned}$$

where the last inequality follows from (5.10). Since ∇F is Lipschitz continuous at $\bar{z}$, it follows from the last inequality that there is a positive constant C_5 such that

$$\begin{aligned}
\|Py_k - P^{-1}s_k\| &\leq \Big(M_3(\nu_k + \|F(z_k)\|) + C_5\|z_k - \bar{z}\|\Big)\|P\|\,\|s_k\| \\
&\leq \Big(M_3(\nu_k + \|F(z_k) - F(\bar{z})\|) + C_5\nu_k\Big)\|P\|\,\|s_k\| \\
&\leq \Big(M_3(\nu_k + M\|z_k - \bar{z}\|) + C_5\nu_k\Big)\|P\|\,\|s_k\| \\
&\leq C_6\nu_k\|s_k\|, \qquad (5.32)
\end{aligned}$$

where $C_6 = [M_3(1+M) + C_5]\|P\|$. Substituting this into (5.31) yields

$$\begin{aligned}
&\Big\|\frac{\tilde{s}_k\tilde{s}_k^T}{\|\tilde{s}_k\|^2} - \frac{\tilde{y}_k\tilde{y}_k^T}{\tilde{y}_k^T\tilde{s}_k}\Big\|_F \\
\leq\ & \frac{\Big(2\|P^{-1}s_k\| + \|Py_k\|\Big)C_6\nu_k\|s_k\|}{y_k^Ts_k} \\
\leq\ & \frac{\Big(2\|P^{-1}\|\,\|s_k\| + \|P\|\,\|y_k\|\Big)C_6\nu_k\|s_k\|}{c_1\|s_k\|^2} \\
\leq\ & c_1^{-1}\Big(2\|P^{-1}\| + C_2\|P\|\Big)C_6\nu_k \\
=\ & M_5\nu_k,
\end{aligned}$$

where $M_5 = c_1^{-1}C_6\Big(2\|P^{-1}\| + C_2\|P\|\Big)$, the second inequality follows from (5.8), and the last inequality follows from (5.9). Combining this with (5.28) and (5.30), we get (5.24).

Now we verify (5.25). The inverse update formula of BFGS method is represented as

$$\begin{aligned}
H_{k+1} &= H_k + \frac{(s_k - H_ky_k)s_k^T + s_k(s_k - H_ky_k)^T}{y_k^Ts_k} - \frac{y_k^T(s_k - H_ky_k)s_ks_k^T}{(y_k^Ts_k)^2} \\
&= \Big(I - \frac{s_ky_k^T}{y_k^Ts_k}\Big)H_k\Big(I - \frac{y_ks_k^T}{y_k^Ts_k}\Big) + \frac{s_ks_k^T}{y_k^Ts_k},
\end{aligned}$$

which is the dual form of DFP update formula in the sense that $H_k \leftrightarrow B_k$, $H_{k+1} \leftrightarrow B_{k+1}$ and $s_k \leftrightarrow y_k$. By (5.32), the condition of Lemma 3.1 in [9] is satisfied (with the identifications $s \leftrightarrow y_k$, $y \leftrightarrow s_k$, $B \leftrightarrow H_k$, $A \leftrightarrow [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}$ and $M \leftrightarrow P^{-1}$). Therefore there are constants $\alpha_1 > 0, \alpha_2 > 0$ and $\beta \in (0,1)$ such that

$$\begin{aligned}
&\|H_{k+1} - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}} \\
&\leq \Big(\sqrt{1-\beta\zeta_k^2} + \alpha_1\frac{\|P^{-1}s_k - Py_k\|}{\|Py_k\|}\Big)\|H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}} \\
&\quad + \alpha_2\frac{\|s_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}y_k\|}{\|Py_k\|}
\end{aligned}$$

$$= \Big(\sqrt{1-\beta\zeta_k^2} + \alpha_1 \frac{\|P^{-1}s_k - Py_k\|}{\|Py_k\|}\Big)\|H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}} + \alpha_2 \frac{\|s_k - P^2 y_k\|}{\|Py_k\|}, \tag{5.33}$$

where ζ_k is defined by (5.26). By the nonsingularity of P, there is a constant $c_4 > 0$ such that

$$\|Py_k\| \geq c_4\|y_k\| \geq c_4 c_2 \|s_k\|,$$

where the last inequality follows from (5.9). So, we get from (5.33)

$$\begin{aligned}
&\|H_{k+1} - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}} \\
&\quad \leq \Big(\sqrt{1-\beta\zeta_k^2} + \alpha_1 \frac{C_6\nu_k\|s_k\|}{c_4c_2\|s_k\|}\Big)\|H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}} + \alpha_2 \frac{C_6\nu_k\|P\|\,\|s_k\|}{c_4c_2\|s_k\|}, \\
&\quad \leq (1 - \frac{1}{2}\beta\zeta_k^2 + \alpha_1 c_2^{-1}c_4^{-1}C_6\nu_k)\|H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}} + \alpha_2 c_2^{-1}c_4^{-1}C_6\nu_k\|P\|,
\end{aligned}$$

where the first inequality follows from (5.32) and the last inequality holds since $\sqrt{1-\beta\zeta_k^2} \leq 1 - \frac{1}{2}\beta\zeta_k^2$. In view of (5.32), we get (5.25).

Finally, from Lemma 3.4 in [9] and (5.14), we see that $\{\|B_k - \nabla F(\bar{z})\nabla F(\bar{z})^T\|_P\}$ and $\{\|H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}}\}$ converge. In particular, $\{\|B_k\|\}$ and $\{\|H_k\|\}$ are bounded. □

Now we prove superlinear convergence of Algorithm 1.

Theorem 5.1 *Let Assumptions A and B hold. Then*

$$\lim_{k\to\infty} \frac{\|(B_k - \nabla F(\bar{z})\nabla F(\bar{z})^T)p_k\|}{\|p_k\|} = 0. \tag{5.34}$$

Moreover, $\{z_k\}$ converges to $\bar{z}$ superlinearly.

Proof We rewrite (5.25) as

$$\begin{aligned}
&\frac{1}{2}\beta\zeta_k^2\|H_k - \nabla F(\bar{z})\nabla F(\bar{z})^T\|_{P^{-1}} \\
&\quad \leq \|H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}} - \|H_{k+1} - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}} \\
&\qquad + \Big(M_6\|H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}} + M_7\Big)\nu_k.
\end{aligned}$$

Since $\{\|H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}}\}$ is bounded and $\nu_k \to 0$, this inequality together with the definition (5.26) of ζ_k yields

$$\lim_{k\to\infty} \zeta_k^2\|H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}} = \lim_{k\to\infty} \frac{\|P^{-1}(H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1})y_k\|^2}{\|H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}}\|Py_k\|^2} = 0. \tag{5.35}$$

Moreover, since $\|H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}\|_{P^{-1}}$ is bounded, (5.35) implies

$$\lim_{k\to\infty} \frac{\|P^{-1}(H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1})y_k\|}{\|Py_k\|} = 0.$$

By the nonsingularity of P and (5.9), there exist some constants $M_8 > 0$ and $c_5 > 0$ such that $\|Py_k\| \le M_8\|s_k\|$ for all k and $\|P^{-1}w\| \ge c_5 w$ for all $w \in R^n$. Hence we get

$$\lim_{k\to\infty} \frac{\|(H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1})y_k\|}{\|s_k\|} = 0. \tag{5.36}$$

On the other hand, we have

$$\begin{aligned}
& \|(H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1})y_k\| \\
= \; & \|H_k(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)[\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}y_k\| \\
\ge \; & \|H_k(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)s_k\| \\
& - \|H_k(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)([\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1}y_k - s_k)\| \\
= \; & \|H_k(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)s_k\| - \|H_k(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)P(Py_k - P^{-1}s_k)\| \\
\ge \; & \|H_k(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)s_k\| - \|H_k(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)P\|\,\|Py_k - P^{-1}s_k\| \\
\ge \; & \|H_k(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)s_k\| - C_6\nu_k\|H_k(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)P\|\|s_k\| \\
= \; & \|H_k(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)s_k\| + o(\|s_k\|),
\end{aligned}$$

where the last inequality follows from (5.32) and the last equality holds because $\nu_k \to 0$ and both $\{\|B_k\|\}$ and $\{\|H_k\|\}$ are bounded. Moreover the fact that $\{\|B_k\|\}$ and $\{\|H_k\|\}$ are bounded particularly implies that $\{\|B_k\|\}$ and $\{\|H_k\|\}$ are uniformly nonsingular. Therefore, there is a constant $c_6 > 0$ such that $\|H_k(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)s_k\| \ge c_6\|(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)s_k\|$ for all k. So we have

$$\|(H_k - [\nabla F(\bar{z})\nabla F(\bar{z})^T]^{-1})y_k\| \ge c_6\|(\nabla F(\bar{z})\nabla F(\bar{z})^T - B_k)s_k\| + o(\|s_k\|),$$

and hence (5.36) yields (5.34). In view of Lemma 5.6, the proof is complete. □

6. Discussion

We have proposed a BFGS method for solving the KKT system of a general constrained optimization problem by means of successive splitting of an equivalent system of nonsmooth equations. Under Assumptions A and B, we have established global and superlinear convergence of the method. Compared with an SQP method, the proposed method has two advantages: First, at

every iteration, the subproblem is always solvable and has a unique solution, so that no additional computational effort is required to restore the solvability of subproblems. Second, to ensure global and superlinear convergence of the method, we are free from the conventional but restrictive condition (1.5). The proposed method is applicable to nonconvex mathematical programming problems. A deficiency of the method is that the subproblems are of full dimension. That is, at each iteration, an $n + m + r$ dimensional linear equation has to be solved. It is well known that under some regularity conditions, the KKT point relies only on active constraints. Reducing the size of subproblems deserves to be considered. We also notice that superlinear convergence has been established under the strict complementarity condition (5.1). It is certainly important to investigate whether the method retains superlinear convergence without this assumption.

References

[1] Boggs, P. T., Tolle, J. W., and Wang, P., *On the local convergence of quasi-Newton methods for constrained optimization*, SIAM Journal on Control and Optimization, Vol. 20, pp. 161-171, 1982.

[2] Bonnans, J. F., and Launay, G., *Sequential quadratic programming with penalization of the displacement*, SIAM Journal on Optimization, Vol. 5, pp. 792-812, 1995.

[3] Burke, J. V., *A sequential quadratic programming method for potentially infeasible mathematical programs*, Journal of Mathematics Analysis and Applications, Vol. 139, pp. 319-351, 1989.

[4] Burke, J. V., and Han, S. P., *A robust sequential quadratic programming method*, Mathematical Programming, Vol. 43, pp. 277-303, 1989.

[5] Byrd, R., and Nocedal, J., *A tool for the analysis of quasi-Newton methods with application to unconstrained minimization*, SIAM Journal on Numerical Analysis, Vol. 26, pp 727-739, 1989.

[6] Byrd, R., Tapia, R. A., and Zhang, Y., *An SQP augmented Lagrangian BFGS algorithm for constrained optimization*, SIAM Journal on Optimization, Vol. 2, pp. 210-241, 1992.

[7] Chen, X., and Qi, L., *A parameterized Newton method and a quasi-Newton method for nonsmooth equations*, Computational Optimization and Applications, Vol. 3, pp. 157-179, 1994.

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

[9] Dennis, J. E. Jr., and Moré, J. J., *A characterization of superlinear convergence and its application to quasi-Newton methods*, Mathematics of Computation, Vol. 28, 549-560, 1974.

[10] Facchinei, F., and Lucidi, S., *Quadratically and superlinearly convergent algorithms for the solution of inequality constrained minimization problems*, Journal of Optimization Theory and Applications, Vol. 85, pp. 265-289, 1995.

[11] Facchinei, F., and Soares, J., *A new merit function for nonlinear complementarity problems and a related algorithm*, SIAM Journal on Optimization, Vol. 7, pp. 225-247, 1997.

[12] Fischer, A., *A special Newton-type optimization method*, Optimization, Vol. 24, pp. 269-284, 1992.

[13] Fischer, A., *A Newton-type method for positive semidefinite linear complementarity problems*, Journal of Optimization Theory and Applications, Vol. 86, pp. 585-608, 1995.

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

[15] Fukushima, M., *A successive quadratic programming algorithm with global and superlinear convergence properties*, Mathematical Programming, Vol. 35, pp. 253-264, 1986.

[16] Geiger, C., and Kanzow, A., *On the resolution of monotone complementarity problems*, Computational Optimization and Applications, Vol. 5, pp. 155-173, 1996.

[17] Han, S. P., *A globally convergent method for nonlinear programming*, Journal of Optimization Theory and Applications, Vol. 22, pp. 297-309, 1977.

[18] Han, S. P., Pang, J. S., and Rangaraj, N., *Globally convergent Newton method for nonsmooth equations*, Mathematics of Operations Research, Vol. 17, pp. 586-607, 1992.

[19] Heinkenschloss, M., *Projected sequential quadratic programming methods*, SIAM Journal on Optimization, Vol. 6 , pp. 373-417, 1996.

[20] Jiang, H., and Qi, L., *A new nonsmooth equations approach to nonlinear complementarity problems*,SIAM Journal on Control and Optimization, Vol. 35, pp. 178-193, 1997.

[21] Kanzow, C., and Fukushima, M., *Equivalence of the generalized complementarity problem to differentiable unconstrained minimization*, Journal of Optimization Theory and Applications, Vol. 90, pp. 581-603, 1996.

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

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

[24] Murray, W., and Prieto, F., *A sequential quadratic programming algorithm using an incomplete solution of the subproblem*, SIAM Journal on Optimization, Vol. 5, pp. 590-640, 1995.

[25] Pang, J. S., *Newton's method for B-differentiable equations*, Mathematics of Operations Research, Vol. 15, pp. 311-341, 1990.

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

[27] Pang, J. S., *Serial and parallel computation of Karush-Kuhn-Tucker points via nonsmooth equations*, SIAM Journal on Optimization, Vol.4, pp. 872-893, 1994.

[28] Powell, M. J. D., *The convergence of variable metric methods for nonlinearly constrained optimization calculations*, Nonlinear Programming 3, Edited by O.L. Mangasarian, R.R. Meyer and S.M. Robinson, Academic Press, pp. 27-63, 1978.

[29] Powell, M. J. D., *Variable metric methods for constrained optimization*, Mathematical Programming, the State of the Art, Edited by A. Bachem, M. Grötschel and B. Korte, Springer-Verlag, pp. 288-311, 1982.

[30] Powell, M. J. D., and Yuan, Y., *A recursive quadratic programming algorithm that uses differentiable exact penalty functions*, Mathematical Programming, Vol. 35, pp. 265-278, 1986.

[31] Qi, L., *Convergence analysis of some algorithms for solving nonsmooth equations*, Mathematics of Operations Research, Vol. 18, pp. 227-244, 1993.

[32] Qi, L., and Chen, X., *A globally convergent successive approximation method for severely nonsmooth equations*, SIAM Journal on Control and Optimization, Vol. 33, pp. 402-418, 1995.

[33] Sahba, M., *A globally convergent algorithm for nonlinearly constrained optimization problems*, Journal of Optimization Theory and Applications, Vol. 52, pp. 291-309, 1987.

[34] Yamashita, N., and Fukushima, M., *Modified Newton methods for solving a semismooth reformulation of monotone complementarity problems*, Mathematical Programming, Vol. 76, pp. 469-491, 1997.

[35] ZHOU, S. Z., LI, D. H., and ZENG, J. P., *A successive approximation quasi-Newton process for nonlinear complementarity problems*, Recent Advances in Nonsmooth Optimization, Edited by D.-Z. Du, L. Qi and R.S. Womersley, World Scientific, pp. 459-472, 1995.