A Derivative-Free Line Search and DFP Method for Symmetric Equations with Global and Superlinear Convergence

Donghui Li [1]

Department of Applied Mathematics

Hunan University

Changsha, China 410082

e-mail: dhli@dns.hunu.edu.cn

Masao Fukushima

Department of Applied Mathematics and Physics

Graduate School of Engineering

Kyoto University

Kyoto 606-8501, Japan

e-mail: fuku@kuamp.kyoto-u.ac.jp

April 7, 1998

Abstract In this paper, we propose a derivative-free line search suited to iterative methods for solving systems of nonlinear equations with symmetric Jacobian matrices. The proposed line search can be implemented conveniently by a backtracking process and has such an attractive property that any iterative method with this line search generates a sequence of iterates that is approximately norm descent. Moreover, if the Jacobian matrices are uniformly nonsingular, then the generated sequence converges to the unique solution. We incorporate this line search with a Gauss-Newton based DFP method for solving symmetric equations. Under appropriate conditions, we establish global and superlinear convergence of the proposed DFP method. The obtained results show, in particular, that the proposed DFP method with inexact line search converges globally and superlinearly even for nonconvex unconstrained optimization problems and equality constrained optimization problems.

Key words: Symmetric equations, DFP method, global convergence, superlinear convergence.

AMS Subject Classification: 65H10

Running Title: DFP Method for Symmetric Equations

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

1 Introduction

Consider the system of nonlinear equations

$$F(x) = 0, \quad x \in R^n, \tag{1.1}$$

where $F : R^n \to R^n$ is assumed to be continuously differentiable. In this paper, we restrict ourselves to the case where the Jacobian matrix $\nabla F(x)$ is symmetric for every $x \in R^n$. This class of problems particularly includes unconstrained optimization problems, saddle point problems and KKT systems for equality constrained optimization problems. Other problems such as the discretized two point boundary value problems and discretized elliptic boundary value problems also take the form of (1.1) with symmetric Jacobian (see Chapter 1 in [15]).

DFP method is a well known quasi-Newton method for solving (1.1). Like many other quasi-Newton methods, DFP method possesses local superlinear convergence property (see e.g. [5]). Moreover, when used to solve convex unconstrained optimization problems, it converges globally and superlinearly if exact line search is used (see [16]). A remarkable result due to Dixon [6] shows that when exact line search is used, Broyden's class of quasi-Newton methods including DFP method and the well known BFGS method generate the same sequence of iterates for an unconstrained optimization problem. However, when inexact line search is used, much difference appears among these methods. In particular, DFP method often performs quite badly. Theoretically, it was shown by Byrd, Nocedal and Yuan [3] that if Wolfe-type line search is used, then every method in the convex Broyden's class of quasi-Newton methods converges globally and superlinearly for convex unconstrained optimization problems except for DFP method. Whether DFP method with inexact line search retains global convergence is not known even for the problem of minimizing a uniformly convex function. This open problem has been mentioned several times (see e.g. [3], [8], [14]). On the other hand, the method has never been observed to fail. In fact no one has given an example to show that DFP method does not converge. The result established in [12] shows that when used to minimize a strictly convex quadratic function with Wolfe-type line search, DFP method converges globally and superlinearly, which suggests the possibility that DFP method with inexact line search converges globally for more general unconstrained minimization problems.

The purpose of this paper is to further study this open problem. Instead of giving a direct

answer to it, we will establish global convergence of DFP method under a different setting. More specifically, unlike the traditional DFP method for solving optimization problems, we will present a Gauss-Newton based DFP method that is related to Gauss-Newton method for (1.1). First, we propose a derivative-free line search for the least squares problem associated with (1.1). The proposed line search possesses the attractive property that any algorithm with this line search generates a sequence of iterates that is approximately norm descent. Moreover, if ∇F is uniformly nonsingular, then the generated sequence converges to the unique solution. Under appropriate conditions, DFP method with this line search is shown to have global and superlinear convergence properties. In particular, this method converges globally and superlinearly for unconstrained optimization problems and equality constrained problems even if the problems do not satisfy the convexity assumption.

In the next section, we first propose a derivative-free line search and show its useful property. We then present a Gauss-Newton based DFP method that uses the proposed line search. In Section 3, we show some useful properties for DFP method under suitable assumptions. In Section 4, we prove global and superlinear convergence of the proposed DFP method. Section 5 summarizes the results and gives some final remarks.

2 A Derivative-Free Line Search and DFP Method

Line search is useful in enlarging convergence domain of an iterative method. Generally, an iterative method with line search exhibits some descent property such as the function descent property for optimization problems and the norm descent property for nonlinear equations. Line search has commonly been used in iterative methods for solving optimization problems (see e.g. [7]). For nonlinear equations, however, the study on line search is relatively scarce. In Newton's method for solving nonlinear equations, an Armijo-type line search similar to that for optimization problems is often used, i.e., given the iterate x_k and the Newton direction $p_k = -\nabla F(x_k)^{-1}F(x_k)$, a steplength $\lambda_k > 0$ is chosen to be the largest element in $\{1, \rho, \rho^2, \ldots\}$ that satisfies

$$\|F(x_k + \lambda_k p_k)\|^2 - \|F(x_k)\|^2 \leq \sigma\lambda_k F(x_k)^T \nabla F(x_k) p_k, \tag{2.1}$$

or equivalently,

$$\|F(x_k + \lambda_k p_k)\|^2 \leq (1 - \sigma\lambda_k) F(x_k)^T \nabla F(x_k) p_k,$$

where $\rho \in (0,1)$ and $\sigma \in (0,1)$ are given constants. Here and throughout the paper, $\|\cdot\|$ denotes the Euclidean norm of vectors and the induced norm of matrices.

If we define the function $\theta : R^n \to R$ by $\theta(x) = \frac{1}{2}\|F(x)\|^2$, then (2.1) is just an Armijo-type line search for the minimization problem

$$\min \theta(x), \qquad x \in R^n.$$

Newton's method with Armijo-type line search is a norm descent method. For quasi-Newton methods, however, an Armijo-type line search like (2.1) does not seem to be directly applicable because quasi-Newton directions are not necessarily descent directions for the function θ. Therefore, globalization of quasi-Newton methods for solving nonlinear equations will require a new line search. The first line search used with quasi-Newton methods was proposed by Griewank [10] for Broyden's method, which determines a (positive or negative) steplength λ_k so as to satisfy

$$\frac{F(x_k)^T(F(x_k + \lambda_k p_k) - F(x_k))}{\|F(x_k + \lambda_k p_k) - F(x_k)\|^2} \geq \frac{1}{2} + \epsilon, \tag{2.2}$$

where $\epsilon \in (0, 1/6)$ is a given constant. We see from (2.2) that if $F(x_k)^T \nabla F(x_k) p_k \neq 0$ then (2.2) is satisfied for all (positive or negative) λ_k sufficiently close to zero. However, in the case $F(x_k)^T \nabla F(x_k) p_k = 0$, the existence of λ_k satisfying (2.2) is not guaranteed theoretically [10, pp .81-82]. Recently, the authors [13] proposed another line search to globalize BFGS method for the symmetric equation (1.1), which chooses a steplength λ_k satisfying

$$\|F(x_k + \lambda_k p_k)\|^2 - \|F(x_k)\|^2 \leq -\sigma_1 \|\lambda_k F(x_k)\|^2 - \sigma_2 \|\lambda_k p_k\|^2 + \epsilon_k \|F(x_k)\|^2, \tag{2.3}$$

where $\sigma_1, \sigma_2 > 0$ are given constants and $\{\epsilon_k\}$ is a given positive sequence. It is easy to see that (2.3) is satisfied for all sufficiently small $\lambda_k > 0$ because $\epsilon_k \|F(x_k)\|^2$ is positive and independent of λ_k. With this line search, global and superlinear convergence for a Gauss-Newton based BFGS method has been established under suitable conditions [13]. Yet it seems difficult to establish similar globalization results for DFP method by using the line search (2.3).

In this paper, we propose a new derivative-free line search. Let p_k be a direction generated by an iterative method for solving (1.1). We determine a steplength $\lambda_k > 0$ so as to satisfy

$$\|F(x_k + \lambda_k p_k)\| - \|F(x_k)\| \leq -\sigma_1 \|\lambda_k F(x_k)\| - \sigma_2 \|F(x_k + \lambda_k p_k) - F(x_k)\| + \epsilon_k, \tag{2.4}$$

where σ_1 and σ_2 are positive constants and $\{\epsilon_k\}$ is a positive sequence satisfying the following inequality with some constant $\epsilon > 0$:

$$\sum_{k=0}^{\infty} \epsilon_k \leq \epsilon < \infty. \tag{2.5}$$

It is easy to see that as $\lambda_k \to 0^+$, the left-hand side of (2.4) goes to zero, while the right-hand side goes to a positive constant ϵ_k. So, (2.4) holds for all $\lambda_k > 0$ sufficiently small, and hence we can find a steplength λ_k by a backtracking process. Moreover, (2.4) implies

$$\|F(x_k + \lambda_k p_k)\| \leq \|F(x_k)\| + \epsilon_k. \tag{2.6}$$

Therefore the sequence $\{x_k\}$ generated by $x_{k+1} = x_k + \lambda_k p_k$ is approximately norm descent because $\epsilon_k \to 0$ by (2.5). Another attractive property of this line search is shown in the following lemma.

Lemma 2.1 *Let $\{x_k\}$ be generated by $x_{k+1} = x_k + \lambda_k p_k$ with λ_k determined by (2.4). Then $\{\|F(x_k)\|\}$ converges. Moreover, if F is continuously differentiable and ∇F is uniformly nonsingular, then*

$$\sum_{k=0}^{\infty} \|x_{k+1} - x_k\| < \infty. \tag{2.7}$$

In particular, $\{x_k\}$ converges.

Proof Since (2.4) implies

$$\|F(x_{k+1})\| \leq \|F(x_k)\| + \epsilon_k,$$

and $\{\epsilon_k\}$ satisfies (2.5), it follows from Lemma 3.3 in [4] that $\{\|F(x_k)\|\}$ converges. If F is continuously differentiable and ∇F is uniformly nonsingular, then there exists a constant $m > 0$ such that $\|F(x) - F(y)\| \geq m\|x - y\|$ holds for all $x, y \in R^n$. Then (2.4) yields

$$\begin{aligned} \sigma_2 m\|x_{k+1} - x_k\| &\leq \sigma_2\|F(x_{k+1}) - F(x_k)\| \\ &\leq \|F(x_k)\| - \|F(x_{k+1})\| - \sigma_1\lambda_k\|F(x_k)\| + \epsilon_k \\ &\leq \|F(x_k)\| - \|F(x_{k+1})\| + \epsilon_k. \end{aligned}$$

Summing these inequalities, we get (2.7) as desired. The last part of the lemma is an easy consequence of (2.7). □

Lemma 2.1 shows a very interesting property of line search (2.4). That is, any iterative method with this line search generates a convergent sequence whenever the function F has uniformly nonsingular Jacobian matrices.

Now we present a DFP method with this line search for solving the system of equations (1.1).

Algorithm 1

Step 0 Choose constants $\rho \in (0,1)$, $\lambda_{-1} > 0$, $\sigma_1, \sigma_2 > 0$ such that $\sigma_1 + \sigma_2 < 1$, and a positive sequence $\{\epsilon_k\}$ satisfying (2.5). Choose an initial point $x_0 \in R^n$ and a symmetric positive definite matrix $B_0 \in R^{n\times n}$. Let $k := 0$.

Step 1. Compute the direction p_k by solving the linear equation

$$B_k p_k + q_k = 0, \tag{2.8}$$

where

$$q_k = \lambda_{k-1}^{-1}(F(x_k + \lambda_{k-1}F(x_k)) - F(x_k)). \tag{2.9}$$

Step 2. Let $\Delta_k : R \to R^n$ be defined by $\Delta_k(\lambda) = F(x_k + \lambda p_k) - F(x_k)$. Let i_k be the smallest nonnegative integer i satisfying

$$\|F(x_k + \rho^i p_k)\| - \|F(x_k)\| \le -\sigma_1\|\rho^i F(x_k)\| - \sigma_2\|\Delta_k(\rho^i)\| + \epsilon_k. \tag{2.10}$$

Let the steplength be $\lambda_k = \rho^{i_k}$.

Step 3. Determine the next iterate by $x_{k+1} = x_k + \lambda_k p_k$.

Step 4. Let $s_k = x_{k+1} - x_k$, $\delta_k = \Delta_k(\lambda_k) = F(x_{k+1}) - F(x_k)$ and $y_k = F(x_k + \delta_k) - F(x_k)$. If $y_k^T s_k \le 0$, then let $B_{k+1} = B_k$. Otherwise, update B_k by DFP formula:

$$B_{k+1} = \Big(I - \frac{y_k s_k^T}{y_k^T s_k}\Big) B_k \Big(I - \frac{s_k y_k^T}{y_k^T s_k}\Big) + \frac{y_k y_k^T}{y_k^T s_k}. \tag{2.11}$$

Step 5. Let $k := k + 1$. Go to Step 1.

We see that Step 4 ensures the positive definiteness of B_k for every k, which in turn ensures that the subproblem (2.8) has a unique solution for each k. Moreover, Step 2 can be executed finitely for each k. Therefore, Algorithm 1 is well defined.

Define

$$G_k = \int_0^1 \nabla F(x_k + \tau\lambda_{k-1}F(x_k))d\tau. \tag{2.12}$$

Then, by (2.9) and (2.12), we have $q_k = G_k F(x_k)$ and (2.8) can be rewritten as

$$B_k p_k + G_k F(x_k) = 0. \tag{2.13}$$

If $\lambda_{k-1}\|F(x_k)\|$ is small, then $G_k \approx \nabla F(x_k) = \nabla F(x_k)^T$ and hence (2.8) is regarded as an approximation of the following linear equation:

$$B_k p_k + \nabla F(x_k) F(x_k) = 0.$$

If in addition, B_k approximates $\nabla F(x_k)\nabla F(x_k)^T = \nabla F(x_k)^2$ along direction p_k, and $\nabla F(x_k)$ is nonsingular, then the direction p_k generated by (2.13) (or (2.8)) will provide a good approximation of the Gauss-Newton direction $-(\nabla F(x_k)\nabla F(x_k)^T)^{-1}\nabla F(x_k)F(x_k)$. In this sense, Algorithm 1 is a Gauss-Newton based DFP method for the symmetric equation (1.1).

In [13], the authors established global and superlinear convergence of a Gauss-Newton based BFGS method for solving (1.1) whose subproblem takes the same form as (2.8).

3 Basic Assumptions and Preliminary Lemmas

In this section, we show some important properties of Algorithm 1, which will turn out to be useful in analyzing its convergence. Throughout, we make the following assumptions:

Assumption A: (i) The function F is continuously differentiable on a convex set D containing the level set

$$\Omega = \{x \mid \|F(x)\| \le \|F(x_0)\| + \epsilon\},$$

where ϵ is a positive real number satisfying (2.5).

(ii) The Jacobian ∇F is symmetric, bounded and uniformly nonsingular on D, i.e., $\nabla F(x) = \nabla F(x)^T$ for all $x \in D$ and there exist constants $M \ge m > 0$ such that

$$m\|p\| \le \|\nabla F(x)p\| \le M\|p\|, \quad \forall x \in D, \forall p \in R^n.$$

(iii) ∇F is Lipschitz continuous on D, i.e., there is a constant $L > 0$ such that

$$\|\nabla F(x) - \nabla F(y)\| \le L\|x - y\|, \quad \forall x, y \in D. \tag{3.1}$$

Remarks (1) In view of (2.5), (2.6) and the definition of Ω, it is easy to see that $\{x_k\} \subset \Omega$.

(2) Assumption A ensures that there are constants $M \ge m > 0$ such that the following

inequalities hold:

$$m \leq \|\nabla F(x)\| \leq M, \quad \forall x \in D, \tag{3.2}$$

$$\frac{1}{M} \leq \|\nabla F(x)^{-1}\| \leq \frac{1}{m}, \quad \forall x \in D, \tag{3.3}$$

$$m\|p\| \leq \|\nabla F(x)p\| \leq M\|p\|, \quad \forall x \in D, \forall p \in R^n, \tag{3.4}$$

$$\frac{1}{M}\|p\| \leq \|\nabla F(x)^{-1}p\| \leq \frac{1}{m}\|p\|, \quad \forall x \in D, \forall p \in R^n, \tag{3.5}$$

$$m\|x-y\| \leq \|F(x)-F(y)\| \leq M\|x-y\|, \quad \forall x, y \in D. \tag{3.6}$$

Now let us define

$$\tilde{G}_k = \int_0^1 \nabla F(x_k + \tau s_k)d\tau \qquad \text{and} \qquad \bar{G}_k = \int_0^1 \nabla F(x_k + \tau \delta_k)d\tau.$$

Then we have

$$\delta_k = F(x_{k+1}) - F(x_k) = \tilde{G}_k s_k \qquad \text{and} \qquad y_k = \bar{G}_k \delta_k = \bar{G}_k \tilde{G}_k s_k. \tag{3.7}$$

Moreover we have the following results.

Lemma 3.1 *The following inequalities hold for every k:*

$$\|\delta_k\| \leq M\|s_k\| \quad \textit{and} \quad \|y_k\| \leq M\|\delta_k\| \leq M^2\|s_k\|, \tag{3.8}$$

$$\|G_k\| \leq M, \tag{3.9}$$

$$\|\bar{G}_k - \tilde{G}_k\| \leq L(M+1)\|s_k\|, \tag{3.10}$$

$$\|\bar{G}_k - \nabla F(x_k)\| \leq L\|\delta_k\| \leq LM\|s_k\|, \tag{3.11}$$

$$\|\tilde{G}_k - \nabla F(x_k)\| \leq L\|s_k\|. \tag{3.12}$$

Proof It is obvious that (3.8) follows from (3.7) and (3.4) directly. Similarly, (3.9) follows from (2.12) and (3.2). Also, by the mean value theorem and (3.1), we have

$$\begin{aligned} \|\bar{G}_k - \tilde{G}_k\| &\leq \int_0^1 \|\nabla F(x_k + \tau\delta_k) - \nabla F(x_k + \tau s_k)\|d\tau \\ &\leq L(\|s_k\| + \|\delta_k\|) \\ &\leq L(M+1)\|s_k\|. \end{aligned}$$

Similarly, we have

$$\|\bar{G}_k - \nabla F(x_k)\| \leq \int_0^1 \|\nabla F(x_k + \tau\delta_k) - \nabla F(x_k)\| d\tau \leq L\|\delta_k\| \leq LM\|s_k\|$$

and

$$\|\tilde{G}_k - \nabla F(x_k)\| \leq \int_0^1 \|\nabla F(x_k + \tau s_k) - \nabla F(x_k)\| d\tau \leq L\|s_k\|.$$

□

Lemma 3.2 *We have*

$$\sum_{k=0}^{\infty} \|\lambda_k F(x_k)\| < \infty. \tag{3.13}$$

Proof The inequality (3.13) can be obtained by summing (2.4) and by taking into account (2.5). □

Lemma 3.3 *There exists a constant $M_1 > 0$ such that for all k*

$$\|y_k - \nabla F(x_k)^2 s_k\| \leq M_1 \|s_k\|^2. \tag{3.14}$$

Proof By an elementary deduction we get

$$\begin{aligned}
\|y_k - \nabla F(x_k)^2 s_k\| &= \|[\bar{G}_k \tilde{G}_k - \nabla F(x_k)^2] s_k\| \\
&= \|[(\bar{G}_k - \nabla F(x_k))\tilde{G}_k + \nabla F(x_k)(\tilde{G}_k - \nabla F(x_k))] s_k\| \\
&\leq [\|\bar{G}_k - \nabla F(x_k)\| \, \|\tilde{G}_k\| + \|\nabla F(x_k)\| \, \|\tilde{G}_k - \nabla F(x_k)\|] \, \|s_k\| \\
&\leq (LM^2\|s_k\| + LM\|s_k\|)\|s_k\| \\
&= ML(M+1)\|s_k\|^2,
\end{aligned} \tag{3.15}$$

where the first equality follows from (3.7), and the last inequality follows from (3.2), (3.11) and (3.12). Putting $M_1 = ML(M+1)$, we get (3.14). □

Lemma 3.4 *There is a constant $m_1 > 0$ such that for all k sufficiently large*

$$y_k^T s_k \geq m_1 \|s_k\|^2. \tag{3.16}$$

Proof We have for every k

$$\begin{aligned} y_k^T s_k &= s_k^T \bar{G}_k \tilde{G}_k s_k \\ &= \|\tilde{G}_k s_k\|^2 + s_k^T (\bar{G}_k - \tilde{G}_k)\tilde{G}_k s_k \\ &\geq \|\tilde{G}_k s_k\|^2 - \|\bar{G}_k - \tilde{G}_k\|\|\tilde{G}_k\|\|s_k\|^2 \\ &= \|F(x_{k+1}) - F(x_k)\|^2 - \|\bar{G}_k - \tilde{G}_k\|\|\tilde{G}_k\|\|s_k\|^2 \\ &\geq m^2\|s_k\|^2 - ML(M+1)\|s_k\|^3, \end{aligned}$$

where the first and the last equalities follow from (3.7), and the last inequality follows from (3.2), (3.6) and (3.10). Since $\|s_k\| \to 0$, we get (3.16). □

Lemma 3.4 shows that under Assumption A, B_{k+1} will always be determined by the update formula (2.11) after a finite number of iterations.

4 Global and Superlinear Convergence of DFP Method

In this section, we prove global and superlinear convergence of Algorithm 1. First, we show some useful properties of DFP update formula. Let P be a symmetric nonsingular matrix. For any matrix A, we define a matrix norm $\|A\|_P = \|PAP\|_F$, where $\|\cdot\|_F$ stands for the Frobenius norm of a matrix.

Lemma 4.1 *Denote $P_k = \nabla F(x_k)$. Then there exist constants $c_1 > 0, c_2 > 0$ and $\alpha \in (0,1)$ such that for all k large enough*

$$\|B_{k+1} - P_k^2\|_{P_k^{-1}} \leq \Big(1 - \frac{1}{2}\alpha\mu_k^2 + c_1\|s_k\|\Big)\|B_k - P_k^2\|_{P_k^{-1}} + c_2\|s_k\|, \tag{4.1}$$

where

$$\mu_k = \frac{\|P_k^{-1}(B_k - P_k^2)s_k\|}{\|B_k - P_k^2\|_{P_k^{-1}}\|P_k s_k\|}.$$

Proof From (3.3) and (3.14), we get

$$\begin{aligned} \|P_k^{-1}y_k - P_k s_k\| &= \|P_k^{-1}(y_k - \nabla F(x_k)^2 s_k)\| \\ &\leq \frac{1}{m}\|y_k - \nabla F(x_k)^2 s_k\| \\ &\leq \frac{M_1}{m}\|s_k\|^2. \end{aligned} \tag{4.2}$$

Since $s_k \to 0$, the condition of Lemma 3.1 in [4] is satisfied for all sufficiently large k. Therefore, there are constants $\alpha \in (0,1)$ and $c_1' > 0, c_2' > 0$ such that

$$\begin{aligned}\|B_{k+1} - P_k^2\|_{P_k^{-1}} &\leq \left(\sqrt{1-\alpha\mu_k^2} + c_1' \frac{\|P_k^{-1}y_k - Ps_k\|)}{\|P_k s_k\|}\right)\|B_k - P_k^2\|_{P_k^{-1}} \\ &\quad + c_2' \frac{\|y_k - P_k^2 s_k\|}{\|P_k s_k\|}.\end{aligned}$$

Notice that $\|P_k s_k\| = \|\nabla F(x_k)s_k\| \geq m\|s_k\|$ by (3.4), and $\sqrt{1-\alpha\mu_k^2} \leq 1 - \frac{1}{2}\alpha\mu_k^2$. Hence, by letting $c_1 = c_1' M_1/m^2$ and $c_2 = c_2' M_1/m$, we get (4.1) from (3.14) and (4.2). □

Lemma 4.2 *There exist constants $\beta_1, \beta_2 > 0$ such that for all k sufficiently large*

$$\begin{aligned}\|B_{k+1} - P_k^2\|_{P_k^{-1}} &\leq \Big(1 - \frac{1}{2}\alpha\mu_k^2 + c_1\|s_k\| + \beta_1\|s_{k-1}\|\Big)\|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}} \\ &\quad + c_2\|s_k\| + \beta_2\|s_{k-1}\|. \qquad (4.3)\end{aligned}$$

Moreover, $\{\|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}}\}$ converges and $\{\|B_k\|\}$ is bounded.

Proof Denote $\tilde{Q}_k = P_{k-1}^{-1}(B_k - P_{k-1}^2)P_{k-1}^{-1} = P_{k-1}^{-1}B_k P_{k-1}^{-1} - I$. Then we can write

$$\begin{aligned}P_k^{-1}(B_k - P_k^2)P_k^{-1} &= P_k^{-1}B_k P_k^{-1} - I \\ &= P_k^{-1}P_{k-1}(P_{k-1}^{-1}B_k P_{k-1}^{-1})P_{k-1}P_k^{-1} - I \\ &= P_k^{-1}P_{k-1}\tilde{Q}_k P_{k-1}P_k^{-1} + P_k^{-1}P_{k-1}P_{k-1}P_k^{-1} - I \\ &= [(P_k^{-1}P_{k-1} - I) + I]\tilde{Q}_k[(P_{k-1}P_k^{-1} - I) + I] \\ &\quad + P_k^{-1}(P_{k-1}^2 - P_k^2)P_k^{-1} \\ &= \tilde{Q}_k + (P_k^{-1}P_{k-1} - I)\tilde{Q}_k(P_{k-1}P_k^{-1} - I) \\ &\quad + (P_k^{-1}P_{k-1} - I)\tilde{Q}_k + \tilde{Q}_k(P_{k-1}P_k^{-1} - I) \\ &\quad + P_k^{-1}(P_{k-1}^2 - P_k^2)P_k^{-1}. \qquad (4.4)\end{aligned}$$

We have

$$\begin{aligned}\|P_k^{-1}P_{k-1} - I\| &\leq \|P_k^{-1}\|\,\|P_k - P_{k-1}\| \\ &= \|\nabla F(x_k)^{-1}\|\,\|\nabla F(x_k) - \nabla F(x_{k-1})\| \\ &\leq \frac{L}{m}\|x_k - x_{k-1}\| \\ &= M_2\|s_{k-1}\|, \qquad (4.5)\end{aligned}$$

where $M_2 = L/m$ and the last inequality follows from (3.1) and (3.3). Similarly, we have

$$\begin{aligned}
\|P_k^{-1}(P_{k-1}^2 - P_k^2)P_k^{-1}\| &\leq \|P_k^{-1}P_{k-1}(P_{k-1} - P_k)P_k^{-1}\| + \|P_k^{-1}(P_{k-1} - P_k)\| \\
&\leq \|P_k^{-1}\|\,\|P_{k-1}\|\,\|P_{k-1} - P_k\|\,\|P_k^{-1}\| + \|P_k^{-1}\|\,\|P_{k-1} - P_k\| \\
&\leq \frac{ML}{m^2}\|s_{k-1}\| + \frac{L}{m}\|s_{k-1}\| \\
&= \frac{L(m+M)}{m^2}\|s_{k-1}\| \\
&= M_3\|s_{k-1}\|, \qquad (4.6)
\end{aligned}$$

where $M_3 = L(m+M)/m^2$. Since any two norms are equivalent, there exist positive constants C and c such that $\|A\|_F \leq C\|A\|$ and $\|A\| \leq c\|A\|_F$ hold for every matrix $A \in R^{n\times n}$. Taking norm operation in (4.4), we have for every k large enough

$$\begin{aligned}
\|B_k - P_k^2\|_{P_k^{-1}} &= \|P_k^{-1}(B_k - P_k^2)P_k^{-1}\|_F \\
&\leq \|\tilde{Q}_k\|_F + \|(P_k^{-1}P_{k-1} - I)\tilde{Q}_k(P_k^{-1}P_{k-1} - I)\|_F \\
&\quad + \|(P_k^{-1}P_{k-1} - I)\tilde{Q}_k\|_F + \|\tilde{Q}_k(P_k^{-1}P_{k-1} - I)\|_F \\
&\quad + \|P_k^{-1}(P_{k-1}^2 - P_k^2)P_k^{-1}\|_F \\
&\leq \|\tilde{Q}_k\|_F + C\|(P_k^{-1}P_{k-1} - I)\tilde{Q}_k(P_k^{-1}P_{k-1} - I)\| \\
&\quad + C\|(P_k^{-1}P_{k-1} - I)\tilde{Q}_k\| + C\|\tilde{Q}_k(P_k^{-1}P_{k-1} - I)\| \\
&\quad + C\|P_k^{-1}(P_{k-1}^2 - P_k^2)P_k^{-1}\| \\
&\leq \|\tilde{Q}_k\|_F + C\|P_k^{-1}P_{k-1} - I\|^2\|\tilde{Q}_k\| + 2C\|P_k^{-1}P_{k-1} - I\|\|\tilde{Q}_k\| \\
&\quad + CM_3\|s_{k-1}\| \\
&\leq \|\tilde{Q}_k\|_F + CM_2^2\|s_{k-1}\|^2\|\tilde{Q}_k\| + 2CM_2\|s_{k-1}\|\,\|\tilde{Q}_k\| \\
&\quad + CM_3\|s_{k-1}\| \\
&\leq \|\tilde{Q}_k\|_F + cCM_2^2\|s_{k-1}\|^2\|\tilde{Q}_k\|_F + 2cCM_2\|s_{k-1}\|\|\tilde{Q}_k\|_F \\
&\quad + CM_3\|s_{k-1}\| \\
&= \|\tilde{Q}_k\|_F(1 + cCM_2^2\|s_{k-1}\|^2 + 2cCM_2\|s_{k-1}\|) + CM_3\|s_{k-1}\| \\
&= \|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}}(1 + cCM_2^2\|s_{k-1}\|^2 + 2cCM_2\|s_{k-1}\|) \\
&\quad + CM_3\|s_{k-1}\|, \qquad (4.7)
\end{aligned}$$

where the third and the fourth inequalities follow from (4.6) and (4.5), respectively. Since $s_k \to 0$, it follows that there is a constant $M_4 > 0$ such that

$$1 + cCM_2^2\|s_{k-1}\|^2 + 2cCM_2\|s_{k-1}\| \le 1 + M_4\|s_{k-1}\|$$

holds for all k large enough. We therefore get from (4.7)

$$\|B_k - P_k^2\|_{P_k^{-1}} \le (1 + M_4\|s_{k-1}\|)\|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}} + CM_3\|s_{k-1}\|. \tag{4.8}$$

Substituting this inequality to (4.1), we obtain

$$\begin{aligned}
\|B_{k+1} - P_k^2\|_{P_k^{-1}} &\le \Big(1 - \frac{1}{2}\alpha\mu_k^2 + c_1\|s_k\|\Big)\|B_k - P_k^2\|_{P_k^{-1}} + c_2\|s_k\| \\
&\le \Big(1 - \frac{1}{2}\alpha\mu_k^2 + c_1\|s_k\|\Big)\Big[(1 + M_4\|s_{k-1}\|)\|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}} \\
&\quad + CM_3\|s_{k-1}\|\Big] + c_2\|s_k\| \\
&\le \Big[1 + M_4\|s_{k-1}\| - \frac{1}{2}\alpha\mu_k^2 \\
&\quad + c_1(1 + M_4\|s_{k-1}\|)\|s_k\|\Big]\|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}} \\
&\quad + CM_3(1 + c_1\|s_k\|)\|s_{k-1}\| + c_2\|s_k\| \\
&= \Big[1 - \frac{1}{2}\alpha\mu_k^2 + c_1\|s_k\| \\
&\quad + (c_1M_4\|s_k\| + M_4)\|s_{k-1}\|\Big]\|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}} \\
&\quad + CM_3(1 + c_1\|s_k\|)\|s_{k-1}\| + c_2\|s_k\|.
\end{aligned}$$

Since $s_k \to 0$, we get (4.3).

Finally, since $\sum_{k=0}^{\infty}\|s_k\| < \infty$ by Lemma 2.1, Lemma 3.3 in [4] indicates that $\{\|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}}\}$ converges. In particular it is bounded. Moreover, since the sequence $\{P_k\} = \{\nabla F(x_k)\}$ is bounded and uniformly nonsingular by Assumption A (ii), the boundedness of $\{\|B_k\|\}$ follows as desired. □

The following technical lemma will be needed later on.

Lemma 4.3 *Let $\{u_k\}$, $\{v_k\}$ and $\{w_k\}$ be positive sequences such that $w_k \le bu_k + v_k$ holds for some constant $b > 0$ and all k large enough. Suppose that the sequence $\{v_k\}$ converges to zero. Suppose further that for some positive constants b_1 and b_2, and a positive sequence $\{t_k\}$ converging to zero,*

$$\lim_{k\to\infty} \frac{u_k}{u_k(1 + b_1t_k) + b_2t_k} w_k^2 = 0. \tag{4.9}$$

Then we have

$$\lim_{k\to\infty} w_k = 0.$$

Proof Denote $J = \{k \mid u_k \geq t_k\}$. Then, for every $k \in J$ large enough, we have

$$\begin{aligned}\frac{u_k}{u_k(1+b_1t_k)+b_2t_k} &= \frac{1}{1+b_1t_k+b_2(t_k/u_k)} \\ &\geq \frac{1}{1+b_1t_k+b_2} \\ &\geq \frac{1}{1+b_1+b_2},\end{aligned}$$

where the last inequality follows from the assumption that $t_k \to 0$. Therefore (4.9) implies $\lim_{k\in J, k\to\infty} w_k = 0$.

On the other hand, from the assumption that both $\{v_k\}$ and $\{t_k\}$ tend to zero and the inequality $w_k \leq bu_k + v_k$, it is easy to see that $\lim_{k\notin J, k\to\infty} w_k = 0$. The proof is complete. □

In the following, it will be convenient to denote

$$\eta_k = \frac{\|(B_k - \nabla F(x_k)^2)p_k\|}{\|p_k\|}. \tag{4.10}$$

The next lemma will play an important role in the convergence analysis for the proposed algorithm.

Lemma 4.4 *Let η_k be defined by (4.10). Then we have*

$$\lim_{k\to\infty} \eta_k = 0. \tag{4.11}$$

Proof We rewrite (4.3) as

$$\begin{aligned}\frac{1}{2}\alpha\mu_k^2\|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}} \leq\ & \|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}} - \|B_{k+1} - P_k^2\|_{P_k^{-1}} \\ & +(c_1\|s_k\| + \beta_1\|s_{k-1}\|)\|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}} + c_2\|s_k\| + \beta_2\|s_{k-1}\|.\end{aligned}$$

Since $\{\|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}}\}$ converges by Lemma 4.2 and $s_k \to 0$ by Lemma 2.1, the right-hand side of the above inequality tends to zero. Then we obtain

$$\mu_k^2\|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}} = \frac{\|P_k^{-1}(B_k - P_k^2)s_k\|^2}{\|B_k - P_k^2\|_{P_k^{-1}}^2\|P_k s_k\|^2}\|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}} \to 0. \tag{4.12}$$

By (3.4) and (3.5), we have

$$\|P_k^{-1}(B_k - P_k^2)s_k\| \geq \frac{1}{M}\|(B_k - P_k^2)s_k\| \quad \text{and} \quad \|P_k s_k\| \leq M\|s_k\|.$$

Since $\{\|B_k - P_k^2\|_{P_k^{-1}}\}$ is bounded, (4.12) implies

$$\frac{\|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}}}{\|B_k - P_k^2\|_{P_k^{-1}}} \frac{\|(B_k - P_k^2)s_k\|^2}{\|s_k\|^2} \to 0. \tag{4.13}$$

Denote again $\tilde{Q}_k = P_{k-1}^{-1}(B_k - P_{k-1}^2)P_{k-1}^{-1}$ as in the proof of Lemma 4.2. Then (4.8) can be rewritten as

$$\|B_k - P_k^2\|_{P_k^{-1}} \leq \Big(1 + M_4\|s_{k-1}\|\Big)\|\tilde{Q}_k\|_F + CM_3\|s_{k-1}\|.$$

Therefore, we get

$$\begin{aligned} \frac{\|B_k - P_{k-1}^2\|_{P_{k-1}^{-1}}}{\|B_k - P_k^2\|_{P_k^{-1}}} \frac{\|(B_k - P_k^2)s_k\|^2}{\|s_k\|^2} &= \frac{\|\tilde{Q}_k\|_F}{\|B_k - P_k^2\|_{P_k^{-1}}}\eta_k^2 \\ &\geq \frac{\|\tilde{Q}_k\|_F}{\|\tilde{Q}_k\|_F(1 + M_4\|s_{k-1}\|) + CM_3\|s_{k-1}\|}\eta_k^2. \end{aligned} \tag{4.14}$$

Moreover, since $\|P_{k-1}\| \leq M$ and $\|\cdot\| \leq c\|\cdot\|_F$ for some constant $c > 0$, we have

$$\begin{aligned} \eta_k &= \frac{\|(B_k - P_k^2)s_k\|}{\|s_k\|} \\ &\leq \|B_k - P_k^2\| \\ &\leq \|B_k - P_{k-1}^2\| + \|P_{k-1}^2 - P_k^2\| \\ &\leq M^2\|P_{k-1}^{-1}(B_k - P_{k-1}^2)P_{k-1}^{-1}\| + \|P_{k-1}^2 - P_k^2\| \\ &\leq M^2 c\|P_{k-1}^{-1}(B_k - P_{k-1}^2)P_{k-1}^{-1}\|_F + \|P_{k-1}^2 - P_k^2\| \\ &= M^2 c\|\tilde{Q}_k\|_F + \|P_{k-1}^2 - P_k^2\|. \end{aligned}$$

Therefore (4.11) follows from (4.13), (4.14) and Lemma 4.3 with $u_k = \|\tilde{Q}_k\|_F$,$v_k = \|P_{k-1}^2 - P_k^2\|$ $w_k = \eta_k$ and $t_k = \|s_{k-1}\|$. □

We are now ready to prove a global convergence theorem for Algorithm 1.

Theorem 4.1 *Let Assumption A hold. Then the sequence $\{x_k\}$ generated by Algorithm 1 converges to the unique solution of (1.1).*

Proof We have already shown in Lemma 2.1 that $\{x_k\}$ converges. Thus it suffices to show that

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

If $\{\lambda_k\}$ contains a subsequence that is bounded away from zero, then it is not difficult to see that Lemma 3.2 yields (4.15). So, we suppose that $\lim_{k\to\infty} \lambda_k = 0$.

From (2.13), we get for every k

$$\begin{aligned}\|\nabla F(x_k)^2 p_k\| &= \|(\nabla F(x_k)^2 - B_k)p_k - G_k F(x_k)\| \\ &\le \|(\nabla F(x_k)^2 - B_k)p_k\| + \|G_k F(x_k)\| \\ &\le \eta_k \|p_k\| + M\|F(x_k)\|, \end{aligned} \tag{4.16}$$

where η_k is defined by (4.10) and the last inequality follows from (3.9) in Lemma 3.1. Since by (3.4) we have $\|\nabla F(x_k)^2 p_k\| \ge m\|\nabla F(x_k) p_k\| \ge m^2\|p_k\|$, it then follows from (4.16) that

$$\|p_k\| \le \frac{\eta_k}{m^2}\|p_k\| + \frac{M}{m^2}\|F(x_k)\|.$$

This together with (4.11) shows that $\{p_k\}$ is bounded. In particular, there is a subsequence $\{p_k\}_{k\in K}$ converging to a vector $\bar{p}$. Again, from (2.13) we have

$$\|\nabla F(x_k)^2 p_k + G_k F(x_k)\| = \|(\nabla F(x_k)^2 - B_k)p_k\| = \eta_k\|p_k\|. \tag{4.17}$$

Since $\eta_k \to 0$ by Lemma 4.4, taking the limit on both sides of (4.17) as $k \to \infty$ with $k \in K$ yields

$$\nabla F(\bar{x})^2 \bar{p} + \nabla F(\bar{x}) F(\bar{x}) = 0,$$

or equivalently

$$\nabla F(\bar{x})\bar{p} + F(\bar{x}) = 0, \tag{4.18}$$

where $\bar{x}$ is the limit of the sequence $\{x_k\}$, which exists as mentioned at the beginning of the proof.

On the other hand, since $\lim_{k\to\infty} \lambda_k = 0$, the line search rule implies that $\lambda_k' \stackrel{\triangle}{=} \lambda_k/\rho = \rho^{i_k - 1}$ does not satisfy (2.10) for all sufficiently large k, which means

$$\begin{aligned}\|F(x_k + \lambda_k' p_k)\| - \|F(x_k)\| &\ge -\lambda_k'\sigma_1\|F(x_k)\| - \sigma_2\|\Delta_k(\lambda_k')\| + \epsilon_k \\ &\ge -\lambda_k'\sigma_2\|F(x_k)\| - \sigma_2\|F(x_k + \lambda_k' p_k) - F(x_k)\|. \end{aligned}$$

Multiplying the both sides by $\|F(x_k + \lambda_k' p_k)\| + \|F(x_k)\|$, we get

$$\begin{aligned}&\|F(x_k + \lambda_k' p_k)\|^2 - \|F(x_k)\|^2 \\ \ge\ & -\Big(\lambda_k'\sigma_1\|F(x_k)\| + \sigma_2\|F(x_k + \lambda_k' p_k) - F(x_k)\|\Big)\Big(\|F(x_k + \lambda_k' p_k)\| + \|F(x_k)\|\Big). \end{aligned}$$

Dividing the both sides by λ'_k and then taking the limit as $k \to \infty$ with $k \in K$, we get

$$2F(\bar{x})^T \nabla F(\bar{x})\bar{p} \geq -2\Big(\sigma_1 \|F(\bar{x})\| + \sigma_2 \|\nabla F(\bar{x})\bar{p}\|\Big)\|F(\bar{x})\|.$$

By (4.18), this inequality can be rewritten as

$$-2\|F(\bar{x})\|^2 \geq -2(\sigma_1 + \sigma_2)\|F(\bar{x})\|^2.$$

Since $0 < \sigma_1 + \sigma_2 < 1$, we get $F(\bar{x}) = 0$. This completes the proof. □

We now turn to establishing a superlinear convergence theorem for Algorithm 1. We first prove the following lemma.

Lemma 4.5 *We have*

$$\lim_{k\to\infty} \frac{\|x_k + p_k - \bar{x}\|}{\|x_k - \bar{x}\|} = 0, \tag{4.19}$$

where $\bar{x}$ is the unique solution of (1.1).

Proof Let G_k be defined by (2.12). Then it follows from (3.1) and (3.6) that

$$\begin{aligned}
\|G_k - \nabla F(x_k)\| &= \Big\| \int_0^1 \Big[\nabla F(x_k + \tau\lambda_{k-1}F(x_k)) - \nabla F(x_k)\Big] d\tau \Big\| \\
&\leq \int_0^1 \|\nabla F(x_k + \tau\lambda_{k-1}F(x_k)) - \nabla F(x_k)\| d\tau \\
&\leq \frac{1}{2} L\lambda_{k-1}\|F(x_k)\| \\
&\leq \frac{1}{2} L\|F(x_k) - F(\bar{x})\| \\
&\leq \frac{1}{2} LM\|x_k - \bar{x}\|. \qquad (4.20)
\end{aligned}$$

By (3.4) and (2.13), we get

$$\begin{aligned}
& m^2\|x_k + p_k - \bar{x}\| \\
&\leq \|\nabla F(x_k)^2(x_k + p_k - \bar{x})\| \\
&= \|(\nabla F(x_k)^2 - B_k)p_k - G_k F(x_k) + \nabla F(x_k)^2(x_k - \bar{x})\| \\
&\leq \|(\nabla F(x_k)^2 - B_k)p_k\| + \|G_k F(x_k) - \nabla F(x_k)^2(x_k - \bar{x})\| \\
&= \eta_k\|p_k\| + \|G_k[F(x_k) - F(\bar{x})] - \nabla F(x_k)^2(x_k - \bar{x})\| \\
&\leq \eta_k(\|x_k + p_k - \bar{x}\| + \|x_k - \bar{x}\|) \\
&\qquad + \Big\| \Big[G_k \int_0^1 \nabla F(\bar{x} + \tau(x_k - \bar{x}))d\tau - \nabla F(x_k)^2\Big](x_k - \bar{x})\Big\|
\end{aligned}$$

$$
\begin{aligned}
&= \eta_k(\|x_k + p_k - \bar{x}\| + \|x_k - \bar{x}\|) \\
&\quad + \Big\| \Big[(G_k - \nabla F(x_k)) \int_0^1 \nabla F(\bar{x} + \tau(x_k - \bar{x}))d\tau \\
&\quad + \nabla F(x_k) \int_0^1 [\nabla F(\bar{x} + \tau(x_k - \bar{x})) - \nabla F(x_k)]d\tau \Big](x_k - \bar{x}) \Big\| \\
&\leq \eta_k(\|x_k + p_k - \bar{x}\| + \|x_k - \bar{x}\|) \\
&\quad + \Big[\|G_k - \nabla F(x_k)\| \int_0^1 \|\nabla F(\bar{x} + \tau(x_k - \bar{x}))\|d\tau \\
&\quad + \|\nabla F(x_k)\| \int_0^1 \|\nabla F(\bar{x} + \tau(x_k - \bar{x})) - \nabla F(x_k)\|d\tau \Big] \|x_k - \bar{x}\| \\
&\leq \eta_k(\|x_k + p_k - \bar{x}\| + \|x_k - \bar{x}\|) \\
&\quad + \frac{1}{2}(LM^2\|x_k - \bar{x}\| + ML\|x_k - \bar{x}\|)\|x_k - \bar{x}\| \\
&= \eta_k(\|x_k + p_k - \bar{x}\| + \|x_k - \bar{x}\|) + \frac{1}{2}LM(M+1)\|x_k - \bar{x}\|^2, \qquad (4.21)
\end{aligned}
$$

where the last inequality follows from (4.20), (3.2) and (3.1). Since $\eta_k \to 0$, (4.21) implies (4.19). □

The following theorem shows the superlinear convergence of Algorithm 1.

Theorem 4.2 *Let Assumption A hold. Then the sequence $\{x_k\}$ generated by Algorithm 1 converges superlinearly to the solution $\bar{x}$ of (1.1).*

Proof From Lemma 4.5, it suffices to verify that $\lambda_k = 1$ holds for all k large enough. First note that, by Lemma 2.1 in [4], (4.19) implies

$$
\lim_{k \to \infty} \frac{\|p_k\|}{\|x_k - \bar{x}\|} = 1. \qquad (4.22)
$$

In particular, we have $p_k \to 0$. Also, by (2.12) and (2.13), we get

$$
\begin{aligned}
0 &= B_k p_k + G_k F(x_k) \\
&= B_k p_k + \nabla F(x_k)F(x_k) + (G_k - \nabla F(x_k))F(x_k) \\
&= B_k p_k + \nabla F(x_k)F(x_k) + (G_k - \nabla F(x_k))(F(x_k) - F(\bar{x})) \\
&= B_k p_k + \nabla F(x_k)F(x_k) + o(\|x_k - \bar{x}\|) \\
&= (B_k - \nabla F(x_k)^2)p_k + \nabla F(x_k)(F(x_k) + \nabla F(x_k)p_k) + o(\|x_k - \bar{x}\|).
\end{aligned}
$$

This together with (3.3) implies that

$$
\|F(x_k) + \nabla F(x_k)p_k\| \leq \|\nabla F(x_k)^{-1}\| \left(\|(B_k - \nabla F(x_k)^2)p_k\| + o(\|x_k - \bar{x}\|) \right)
$$

$$
\begin{aligned}
&= \|\nabla F(x_k)^{-1}\|(\eta_k\|p_k\| + o(\|x_k - \bar{x}\|)) \\
&= o(\|p_k\|) + o(\|x_k - \bar{x}\|) \\
&= o(\|p_k\|),
\end{aligned}
$$

where the second equality follows from (3.3) and Lemma 4.4, and the last equality follows from (4.22). So we get

$$
\begin{aligned}
&\|F(x_k)\| - \sigma_1\|F(x_k)\| - \sigma_2\|F(x_k + p_k) - F(x_k)\| \\
&\quad = (1 - \sigma_1)\|F(x_k)\| - \sigma_2\|\nabla F(x_k)p_k\| + o(\|p_k\|) \\
&\quad \geq (1 - \sigma_1)\|F(x_k)\| - \sigma_2\|F(x_k)\| - \sigma_2\|F(x_k) + \nabla F(x_k)p_k\| + o(\|p_k\|) \\
&\quad = (1 - \sigma_1 - \sigma_2)\|F(x_k) - F(\bar{x})\| + o(\|x_k - \bar{x}\|) \\
&\quad \geq (1 - \sigma_1 - \sigma_2)m\|x_k - \bar{x}\| + o(\|x_k - \bar{x}\|), \qquad (4.23)
\end{aligned}
$$

where the last inequality follows from (3.6). On the other hand, by (4.19) and (3.6), we have

$$
\|F(x_k + p_k)\| \leq M\|x_k + p_k - \bar{x}\| = o(\|x_k - \bar{x}\|). \qquad (4.24)
$$

Therefore, it follows from (4.23) and (4.24) that for all k sufficiently large

$$
\begin{aligned}
\|F(x_k + p_k)\| &\leq \|F(x_k)\| - \sigma_1\|F(x_k)\| - \sigma_2\|F(x_k + p_k) - F(x_k)\| \\
&\leq \|F(x_k)\| - \sigma_1\|F(x_k)\| - \sigma_2\|F(x_k + p_k) - F(x_k)\| + \epsilon_k.
\end{aligned}
$$

This means that the unit steplength $\lambda_k = 1$ is accepted for all k large enough. □

The proof of Theorem 4.2 reveals the fact that $\|F(x_k + p_k)\| \leq (1 - \sigma_1)\|F(x_k)\| < \|F(x_k)\|$ holds for all k sufficiently large. In other words, when k is sufficiently large, Algorithm 1 exhibits a norm descent property.

5 Conclusion

In this paper, we have proposed a Gauss-Newton based DFP method with a derivative-free line search for solving symmetric equations and proved its global and superlinear convergence. Since the method can be applied to the problem of finding a stationary point of an unconstrained optimization problem, we have given, as a consequence, an affirmative answer to the question whether DFP method with inexact line search converges globally for uniformly convex

unconstrained optimization problems. However, whether DFP method with Wolfe-Powell line search converges globally for unconstrained optimization problem still remains open.

The results obtained in this paper are expected to pave the way to developing globally convergent damped quasi-Newton methods for solving systems of nonlinear equations. Moreover, global and superlinear convergence theorems may be established in a similar manner for BFGS method and the restricted Broyden's class of quasi-Newton methods. An interesting subject for further research is to extend the obtained results to general nonlinear equations with asymmetric Jacobian matrices.

References

[1] C.G. Broyden, J.E. Dennis, Jr. and J.J. Moré, On the local and superlinear convergence of quasi-Newton methods, *J. Inst. Math. Appl.*, 12 (1973) 223-246.

[2] R. Byrd and J. Nocedal, A tool for the analysis of quasi-Newton methods with application to unconstrained minimization, *SIAM J. Numer. Anal.*, 26 (1989) 727-739.

[3] R. Byrd, J. Nocedal and Y. Yuan, Global convergence of a class of quasi-Newton methods on convex problems, *SIAM J. Numer. Anal.*, 24 (1987) 1171-1189.

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

[5] J.E. Dennis and J.J. Moré, Quasi-Newton methods, motivation and theory, *SIAM Rev.*, 19 (1977) 46-89.

[6] L.C.W. Dixon, Variable metric algorithms: Necessary and sufficient conditions for identical behavior on nonquadratic functions, *J. Optimization Theory Appl.*, 10 (1972) 34-40.

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

[8] R. Fletcher, An overview of unconstrained optimization, in *Algorithms for Continuous Optimization: The State of the Art,* E. Spedicato, ed., Kluwer Academic Publishers, Boston, 1994, pp. 109-143.

[9] A. Griewank and Ph.L. Toint, Local convergence analysis for partitioned quasi-Newton updates, *Numer. Math.*, 39 (1982) 429-448.

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

[11] A. Griewank, The global convergence of partitioned BFGS on problems with convex decompositions and Lipschitzian gradients, *Math. Program.*, 50 (1991) 141-175.

[12] D. Li, On the global convergence of DFP method, *J. Hunan Univ., Nat. Sci.*, 20 (1993) 16-20.

[13] D. Li and M. Fukushima, 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.

[14] J. Nocedal, Theory of algorithms for unconstrained optimization, *Acta Numer.*, 1 (1992) 199-242.

[15] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[16] M.J.D. Powell, On the convergence of the variable metric algorithm, *J. Inst. Math. Appl.*, 7 (1971) 21-36.

[17] M.J.D. Powell, Some global convergence properties of a variable metric algorithm for minimization without exact line searches, in *Nonlinear Programming, SIAM-AMS Proceedings, Vol. IX,* R.W. Cottle, and C.E. Lemke, eds., SIAM, 1976, pp. 53-72.

[18] Ph.L. Toint, Global convergence of the partitioned BFGS algorithm for convex partially separable optimization, *Math. Program.*, 36 (1986) 290-306.

[19] Y. Zhang and R.P. Tewarson, Quasi-Newton algorithms with updates from the pre-convex part of Broyden's family, *IMA J. Numer. Anal.*, 8 (1988) 487-509.