A Derivative-Free Line Search and Global Convergence of Broyden-Like Method for Nonlinear Equations

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

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

March 4, 1999

Abstract

In this paper, by using derivative-free line search, we propose quasi-Newton methods for smooth nonlinear equations. Under appropriate conditions, we show that the proposed quasi-Newton methods converge globally and superlinearly. In particular, for nonlinear equations involving a mapping with positive definite Jacobian matrices, we propose a norm descent quasi-Newton method and establish its global and superlinear convergence.

Key words: Derivative-free line search, quasi-Newton method, global convergence, superlinear convergence

Abbreviated title: A Global Broyden-like Method

AMS subject classification: 65H10, 90C26

[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 $F = (F_1, F_2, \ldots, F_n)$ be a continuously differentiable mapping from R^n into itself. Consider the nonlinear equation

$$F(x) = 0. \tag{1.1}$$

Among various numerical methods for solving (1.1), quasi-Newton methods are regarded as one of the most effective classes of methods. Since the first quasi-Newton method for solving nonlinear equations was proposed by Broyden [1] in 1965, there has been significant progress in the theoretical study on quasi-Newton methods, especially in the local convergence analysis. Early work on local convergence analysis for quasi-Newton methods can be found in [2] and [3]. A systematic and comprehensive review is given by Dennis and Moré [4]. On the contrary, the study on global convergence of quasi-Newton methods is scarce. The main difficulty seems to lie in the lack of effective line searches. Line searches that require the calculation of derivatives are inappropriate for quasi-Newton methods. Therefore, to develop a global quasi-Newton method, derivative-free line search is desirable. To the authors' knowledge, the earliest global convergence result is due to Griewank [5], where a derivative-free line search is proposed. Griewank shows that if Broyden's method with the proposed derivative-free line search is applied to a nonlinear equation with uniformly nonsingular Jacobian matrices, and if the Jacobian matrices are Lipschitz continuous, then the method converges to the unique solution of the nonlinear equation. The result obtained by Griewank [5] makes it possible to construct global quasi-Newton methods for nonlinear equations, though as pointed out in [5], there is a certain difficulty with the line search. In fact the example given in the next section indicates the difficulty. Recently, the authors have proposed some derivative-free line searches that do not suffer from such difficulty, and developed global quasi-Newton methods for symmetric nonlinear equations and a class of nonsmooth equations arising from nonlinear programming problems, nonlinear complementarity problems, mixed complementarity problems and variational inequality problems [6, 7, 8, 9, 10]. However, those line searches possess only approximate norm descent property even for a smooth symmetric equation that involves a mapping with positive definite Jacobian matrices.

The purpose of this paper is to propose a quasi-Newton method with a well-defined derivative-free line search and establish global and superlinear convergence of the proposed

method. Moreover, for a nonlinear equation involving a mapping with positive definite Jacobian matrices, we propose a norm descent quasi-Newton method and show its global and superlinear convergence.

We organize the paper as follows. In the next section, we first introduce a derivative-free line search that is approximately norm descent. Based on this line search, we propose a quasi-Newton method and show its global and superlinear convergence. In Section 3, we develop a norm descent quasi-Newton method for solving nonlinear equations involving a mapping with positive definite but not necessary symmetric Jacobian matrices, and prove its global and superlinear convergence.

2. Approximate Norm Descent Quasi-Newton Method

In this section, we propose a quasi-Newton method based on an approximate norm descent derivative-free line search. We first review the derivative-free line search due to Griewank [5]. Define

$$q_k(\lambda) = -\frac{F(x_k)^T(F(x_k + \lambda p_k) - F(x_k))}{\|F(x_k + \lambda p_k) - F(x_k)\|^2}. \tag{2.1}$$

Then the derivative-free line search in [5] determines a steplength λ_k (positive or negative) such that the following inequality holds:

$$q_k(\lambda_k) \geq \frac{1}{2} + \epsilon, \tag{2.2}$$

where $\epsilon \in (0, 1/6)$ is a given constant. A favorable property of this line search is that it makes an iterative method possess a norm descent property. Indeed, it follows from (2.2) that $2q_k(\lambda_k) - 1 \geq 2\epsilon$, which implies

$$\|F(x_{k+1})\|^2 = \|F(x_k + \lambda_k p_k)\|^2 \leq \|F(x_k)\|^2 - 2\epsilon\|F(x_k + \lambda p_k) - F(x_k)\|^2 < \|F(x_k)\|^2.$$

It is easy to see from the definition (2.1) of q_k that if $F(x_k)^T\nabla F(x_k)p_k \neq 0$, then (2.2) is satisfied for all λ_k sufficiently close to zero, and we can determine a steplength λ_k by using a backtracking process finitely. However, when $F(x_k)^T\nabla F(x_k)p_k = 0$ holds, the existence of λ_k satisfying (2.2) is not guaranteed. Indeed, the following example shows the possibility that this line search fails when $F(x_k)^T\nabla F(x_k)p_k = 0$.

Example. Let $F : R^2 \to R^2$ be defined by

$$F(x) = \begin{pmatrix} \frac{1}{3}(x_1 - x_2)^3 + x_1 - x_2 \\ x_2 \end{pmatrix}.$$

The unique solution of the equation $F(x) = 0$ is $x^* = (0,0)^T$. By direct calculation, we get the Jacobian matrix

$$\nabla F(x) = \begin{pmatrix} (x_1 - x_2)^2 + 1 & 0 \\ -(x_1 - x_2)^2 - 1 & 1 \end{pmatrix}.$$

It is clear that $\nabla F(x)$ is uniformly nonsingular on any bounded set. Let $x_k = (a,0)^T$ be the kth iterate, where $a \neq 0$ is a given constant, and let $p_k = (1,1)^T$ be the direction generated. Then we have $F(x_k) = (\frac{1}{3}a^3 + a, 0)^T$ and $F(x_k + \lambda p_k) - F(x_k) = (0, \lambda)^T$. It then follows that $F(x_k)^T(F(x_k + \lambda p_k) - F(x_k)) = 0$ for all λ. However, we have $\|F(x_k + \lambda p_k) - F(x_k)\|^2 = \lambda^2$. Therefore, there is no λ satisfying (2.2).

The above example reveals that the derivative-free line search (2.2) may not be well-defined. In the following, we propose a new derivative-free line search that is well-defined and easy to implement. Let $\eta > 0$ be a given constant and $\{\eta_k\}$ be a given positive sequence satisfying

$$\sum_{k=0}^{\infty} \eta_k \leq \eta < \infty. \tag{2.3}$$

Let $\sigma_1 > 0$ be a constant. It is not difficult to see that the inequality

$$\|F(x_k + \lambda p_k)\| \leq \|F(x_k)\| - \sigma_1 \|\lambda p_k\|^2 + \eta_k \|F(x_k)\| \tag{2.4}$$

holds for all sufficiently small $\lambda > 0$, because as λ goes to zero, the left-hand side of (2.4) tends to $\|F(x_k)\|$ but the right-hand side tends to $(1 + \eta_k)\|F(x_k)\|$. We call (2.4) an approximate norm descent line search since we have for each k

$$\|F(x_k + \lambda p_k)\| \leq (1 + \eta_k)\|F(x_k)\| \tag{2.5}$$

and η_k satisfies (2.3). The line search (2.4) can be implemented by a backtracking process.

Procedure 1 (Approximate Norm Descent Line Search):
Let $\beta \in (0,1)$ be given. For $i = 0, 1, \ldots$, check the inequality (2.4) with $\lambda = \beta^i$ successively. Let i_k be the smallest nonnegative integer i such that (2.4) holds for $\lambda = \beta^i$. Let $\lambda_k := \beta^{i_k}$.

By using the approximate norm descent line search, we propose a quasi-Newton method for solving (1.1) as follows.

Algorithm 1 (Approximate Norm Descent Quasi-Newton Method):

Step 0. Choose constants $\rho, \beta, \bar{\theta} \in (0,1)$, and $\sigma_1, \sigma_2 > 0$. Select a positive sequence $\{\eta_k\}$ satisfying (2.3). Choose an initial point $x_0 \in R^n$ and a nonsingular matrix $B_0 \in R^{n \times n}$. Let $k := 0$.

Step 1. Stop, if $F(x_k) = 0$. Otherwise, solve the linear equation

$$B_k p + F(x_k) = 0 \tag{2.6}$$

to get p_k.

Step 2. If

$$\|F(x_k + p_k)\| \leq \rho \|F(x_k)\| - \sigma_2 \|p_k\|^2, \tag{2.7}$$

then let $\lambda_k := 1$ and go to Step 4.

Step 3. Let λ_k be determined by Procedure 1.

Step 4. Let $x_{k+1} := x_k + \lambda_k p_k$.

Step 5. Update B_k by Broyden-like formula

$$B_{k+1} := B_k + \theta_k \frac{(y_k - B_k s_k) s_k^T}{\|s_k\|^2}, \tag{2.8}$$

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

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

Now, we show some useful properties of Algorithm 1. First, it is clear from (2.4) and (2.7) that for each k

$$\sigma_0 \|s_k\|^2 = \sigma_0 \|x_{k+1} - x_k\|^2 \leq \|F(x_k)\| - \|F(x_{k+1})\| + \eta_k \|F(x_k)\|, \tag{2.9}$$

where $\sigma_0 = \min\{\sigma_1, \sigma_2\}$. Next, we show that $\{x_k\}$ is contained in the level set

$$\Omega = \{x \in R^n \mid \|F(x)\| \leq e^{\eta} \|F(x_0)\|\}, \tag{2.10}$$

where η is specified in (2.3).

Lemma 2.1 *The sequence $\{x_k\}$ generated by Algorithm 1 is contained in Ω.*

Proof It follows from (2.5) that for any k

$$\begin{aligned}
\|F(x_{k+1})\| &\leq (1+\eta_k)\|F(x_k)\| \\
&\vdots \\
&\leq \|F(x_0)\| \prod_{i=0}^{k}(1+\eta_i) \\
&\leq \|F(x_0)\| \Big(\frac{1}{k+1}\sum_{i=0}^{k}(1+\eta_i)\Big)^{k+1} \\
&= \|F(x_0)\| \Big(1+\frac{1}{k+1}\sum_{i=0}^{k}\eta_i\Big)^{k+1} \\
&\leq \|F(x_0)\| \Big(1+\frac{\eta}{k+1}\Big)^{k+1} \\
&\leq e^{\eta}\|F(x_0)\|.
\end{aligned}$$

This completes the proof. □

The following lemma comes from Lemma 3.3 in [3].

Lemma 2.2 *Let $\{a_k\}$ and $\{r_k\}$ be positive sequences satisfying*

$$a_{k+1} \leq (1+r_k)a_k + r_k$$

and

$$\sum_{k=0}^{\infty} r_k < \infty. \tag{2.11}$$

Then $\{a_k\}$ converges.

Lemma 2.3 *Let the level set Ω be bounded and $\{x_k\}$ be generated by Algorithm 1. Then we have*

$$\sum_{k=0}^{\infty} \|s_k\|^2 < \infty. \tag{2.12}$$

Proof By the boundedness of Ω and Lemma 2.1, the sequence $\{\|F(x_k)\|\}$ is bounded. Since η_k satisfies (2.3), we get (2.12) by summing both sides of (2.9). □

By using (2.5) and Lemma 2.2, it is easy to prove the following lemma.

Lemma 2.4 *Let the level set Ω be bounded and $\{x_k\}$ be generated by Algorithm 1. Then the sequence $\{\|F(x_k)\|\}$ converges.*

To establish global convergence of Algorithm 1, we need the following assumption.

Assumption A (i) The level set Ω defined by (2.10) is bounded.

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

$$\|\nabla F(x) - \nabla F(y)\| \leq L\|x - y\|, \qquad \forall x, y \in \Omega. \tag{2.13}$$

(iii) $\nabla F(x)$ is nonsingular for every $x \in \Omega$.

Although Assumption A may look strong, it is not as strong as the assumption adopted in [5], where the uniform nonsingularity of $\nabla F(x)$ is assumed.

Before proving global convergence of Algorithm 1, we show two lemmas.

Lemma 2.5 *Let $\{a_k\}$, $\{b_k\}$ and $\{\zeta_k\}$ be positive sequences satisfying*

$$a_{k+1}^2 \leq (a_k + b_k)^2 - \alpha\zeta_k^2, \qquad k = 0, 1, \ldots, \tag{2.14}$$

where $\alpha > 0$ is a constant. Then the condition

$$\sum_{k=0}^{\infty} b_k^2 < \infty \tag{2.15}$$

implies

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \zeta_i^2 = 0, \tag{2.16}$$

and the condition

$$\sum_{k=0}^{\infty} b_k < \infty \tag{2.17}$$

implies

$$\sum_{k=0}^{\infty} \zeta_k^2 < \infty. \tag{2.18}$$

Proof The inequality (2.14) particularly implies

$$a_{k+1} \leq a_k + b_k \leq \cdots \leq a_0 + \sum_{i=0}^{k} b_i. \tag{2.19}$$

So, we get again from (2.14)

$$\alpha\zeta_k^2 \le (a_k + b_k)^2 - a_{k+1}^2 = a_k^2 - a_{k+1}^2 + 2a_k b_k + b_k^2 \le a_k^2 - a_{k+1}^2 + 2\Big(a_0 + \sum_{i=0}^{k-1} b_i\Big)b_k + b_k^2,$$

where the last inequality follows from (2.19). Summing these inequalities yields

$$\begin{aligned}
\alpha \sum_{k=0}^{l-1} \zeta_k^2 &\le a_0^2 + 2\sum_{k=0}^{l-1} b_k\Big(a_0 + \sum_{i=0}^{k-1} b_i\Big) + \sum_{k=0}^{l-1} b_k^2 \\
&= a_0^2 + 2a_0 \sum_{k=0}^{l-1} b_k + 2\sum_{k=0}^{l-1} b_k \sum_{i=0}^{k-1} b_i + \sum_{k=0}^{l-1} b_k^2 \\
&= a_0^2 + 2a_0 \sum_{k=0}^{l-1} b_k + \Big(\sum_{k=0}^{l-1} b_k\Big)^2 \\
&= \Big(a_0 + \sum_{k=0}^{l-1} b_k\Big)^2. \qquad (2.20)
\end{aligned}$$

If (2.17) holds, then (2.18) follows from (2.20) immediately. Now, suppose (2.15) holds. Let t be an arbitrary positive integer. It follows from (2.20) that for any $l > t$

$$\begin{aligned}
\alpha \sum_{k=0}^{l-1} \zeta_k^2 &\le \Big(a_0 + \sum_{k=0}^{t-1} b_k + \sum_{k=t}^{l-1} b_k\Big)^2 \\
&\le 2\Big(a_0 + \sum_{k=0}^{t-1} b_k\Big)^2 + 2\Big(\sum_{k=t}^{l-1} b_k\Big)^2 \\
&= 2\Big(a_0 + \sum_{k=0}^{t-1} b_k\Big)^2 + 2(l-t)\sum_{k=t}^{l-1} b_k^2.
\end{aligned}$$

Dividing the both sides of the last inequality by l and letting l go to infinity yield

$$\alpha \lim_{l\to\infty} \frac{1}{l} \sum_{k=0}^{l-1} \zeta_k^2 \le 2\sum_{k=t}^{\infty} b_k^2. \qquad (2.21)$$

Since t is arbitrary, we get (2.16) from (2.15) and (2.21). □

For the subsequent analysis, it will be convenient to define

$$A_{k+1} = \int_0^1 \nabla F(x_k + ts_k)dt. \qquad (2.22)$$

Then we have $y_k = A_{k+1}s_k$, and by the update formula (2.8), we have

$$B_{k+1} = B_k + \theta_k \frac{(A_k - B_k)s_k s_k^T}{\|s_k\|^2}. \qquad (2.23)$$

Moreover, we define

$$\zeta_k = \frac{\|y_k - B_k s_k\|}{\|s_k\|}. \tag{2.24}$$

Since $y_k = A_{k+1} s_k$, we can write

$$\zeta_k = \frac{\|(A_{k+1} - B_k)s_k\|}{\|s_k\|} = \frac{\|(A_{k+1} - B_k)p_k\|}{\|p_k\|}. \tag{2.25}$$

Lemma 2.6 *Let the conditions (i) and (ii) in Assumption A hold and $\{x_k\}$ be generated by Algorithm 1. If*

$$\sum_{k=0}^{\infty} \|s_k\|^2 < \infty, \tag{2.26}$$

then we have

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \zeta_i^2 = 0. \tag{2.27}$$

In particular, there is a subsequence of $\{\zeta_k\}$ tending to zero. If

$$\sum_{k=0}^{\infty} \|s_k\| < \infty, \tag{2.28}$$

then we have

$$\sum_{k=0}^{\infty} \zeta_k^2 < \infty. \tag{2.29}$$

In particular, the whole sequence $\{\zeta_k\}$ converges to zero.

Proof Note that, by the Lipschitz continuity of ∇F, we get

$$\begin{aligned} \|A_{k+1} - A_k\|_F &\leq \int_0^1 \|\nabla F(x_k + ts_k) - \nabla F(x_{k-1} + ts_{k-1})\|_F dt \\ &\leq L \int_0^1 \|(x_k + ts_k) - (x_{k-1} + ts_{k-1})\| dt \\ &= L \int_0^1 \|(1-t)s_{k-1} + ts_k\| dt \\ &\leq \frac{1}{2} L(\|s_k\| + \|s_{k-1}\|). \end{aligned} \tag{2.30}$$

Let

$$a_k = \|B_k - A_k\|_F \quad \text{and} \quad b_k = \|A_{k+1} - A_k\|_F.$$

Then by (2.23) and (2.25), we deduce that

$$\begin{aligned} a_{k+1}^2 &= \|(B_k - A_{k+1})(I - \theta_k s_k s_k^T/\|s_k\|^2)\|_F^2 \\ &= \|B_k - A_{k+1}\|_F^2 - \theta_k(2-\theta_k)\|(B_k - A_{k+1})s_k s_k^T\|_F^2/\|s_k\|^2 \\ &= \|B_k - A_{k+1}\|_F^2 - \theta_k(2-\theta_k)\|y_k - B_k s_k\|^2/\|s_k\|^2 \\ &= \|B_k - A_{k+1}\|_F^2 - \theta_k(2-\theta_k)\zeta_k^2. \end{aligned} \tag{2.31}$$

Since θ_k satisfies $|\theta_k - 1| \leq \bar{\theta}$, it follows that $\theta_k(2-\theta_k) \geq (1-\bar{\theta}^2) > 0$. So, from (2.31) and the triangle inequality, we obtain

$$a_{k+1}^2 \leq (a_k + b_k)^2 - (1-\bar{\theta}^2)\zeta_k^2. \tag{2.32}$$

From (2.30), the conditions (2.28) and (2.26) imply that $\{b_k\}$ satisfies (2.15) and (2.17), respectively. Therefore, by Lemma 2.5, we complete the proof. □

Griewank [5] proved a property similar to Lemma 2.6 for Broyden's rank one method, i.e., $\theta_k \equiv 1$ in (2.8). Lemma 2.6 can be regarded as a slight extension of Griewank's result.

Now we show global convergence of Algorithm 1.

Theorem 2.1 *Let Assumption A hold. Then the sequence $\{x_k\}$ generated by Algorithm 1 converges to the unique solution of (1.1).*

Proof By Lemma 2.4, the sequence $\{\|F(x_k)\|\}$ converges. Thus, it suffices to show that there exists an accumulation point of $\{x_k\}$ that is a solution of (1.1). If there are infinitely many k for which λ_k is determined by (2.7), then $\|F(x_{k+1})\| \leq \rho\|F(x_k)\|$ holds for infinitely many k. This implies $\liminf_{k\to\infty} \|F(x_k)\| = 0$, and hence the conclusion is true.

Now assume that λ_k are determined by (2.4) for all k sufficiently large. Let ζ_k be defined by (2.24) and A_{k+1} be defined by (2.22). By (2.12) and Lemma 2.6, there is a subsequence $\{\zeta_k\}_{k\in K}$ of $\{\zeta_k\}$ converging to zero. Since $\{x_k\}_{k\in K} \subset \Omega$ is bounded, without loss of generality, we may assume that it converges to some point $\bar{x}$. Since (2.12) particularly implies $s_k = x_{k+1} - x_k \to 0$, it is clear that $\{A_{k+1}\}_{k\in K}$ tends to $\nabla F(\bar{x})$ and hence there exists a constant $M_1 > 0$ such that $\|A_{k+1}^{-1}\| \leq M_1$ for all $k \in K$ sufficiently large. Thus, we get by (2.6) and (2.25)

$$\begin{aligned} \|p_k\| &= \|A_{k+1}^{-1}((A_{k+1} - B_k)p_k - F(x_k))\| \\ &\leq \|A_{k+1}^{-1}\| \left(\|(A_{k+1} - B_k)p_k\| + \|F(x_k)\|\right) \\ &\leq M_1(\zeta_k\|p_k\| + \|F(x_k)\|). \end{aligned}$$

This implies that there is a constant $C_1 > 0$ such that for all $k \in K$ sufficiently large

$$\|p_k\| \leq C_1 \|F(x_k)\|. \tag{2.33}$$

Without loss of generality, we assume that $\{p_k\}_{k\in K}$ converges to some vector $\bar{p}$. Since we have $\|B_k p_k - A_{k+1} p_k\| = \zeta_k \|p_k\|$ by (2.25), it follows that $B_k p_k \to \nabla F(\bar{x})\bar{p}$ as $k \to \infty$ with $k \in K$. Taking the limit in (2.6) as $k \to \infty$ with $k \in K$ yields

$$\nabla F(\bar{x})\bar{p} + F(\bar{x}) = 0. \tag{2.34}$$

Let $\bar{\lambda} = \limsup_{k\in K, k\to\infty} \lambda_k$. Then we have $\bar{\lambda} \geq 0$ and $\bar{\lambda}\bar{p} = 0$. If $\bar{\lambda} > 0$, then $\bar{p} = 0$ and hence it follows from (2.34) that $F(\bar{x}) = 0$. If $\bar{\lambda} = 0$, i.e., $\lim_{k\in K, k\to\infty} \lambda_k = 0$, then by Procedure 1, when $k \in K$ is sufficiently large, $\lambda'_k \stackrel{\triangle}{=} \lambda_k/\beta$ does not satisfy (2.4). This implies

$$\|F(x_k + \lambda'_k p_k)\| - \|F(x_k)\| \geq -\sigma_1 \|\lambda'_k p_k\|^2. \tag{2.35}$$

Multiplying the both sides of (2.35) by $(\|F(x_k + \lambda'_k p_k)\| + \|F(x_k)\|)/\lambda'_k$ and letting $k \to \infty$ with $k \in K$ yield

$$F(\bar{x})^T \nabla F(\bar{x})\bar{p} \geq 0.$$

This together with (2.34) implies $F(\bar{x}) = 0$. The proof is then complete. □

Now, we are going to show superlinear convergence of Algorithm 1. First, we prove the following lemma.

Lemma 2.7 *Let Assumption A hold and $\{x_k\}$ be generated by Algorithm 1. Then there exist a constant $\zeta > 0$ and an index $\bar{k}$ such that $\lambda_k = 1$ whenever $\zeta_k \leq \zeta$ and $k \geq \bar{k}$. Moreover, the inequality*

$$\|F(x_k + p_k)\| \leq \rho \|F(x_k)\| - \sigma_2 \|p_k\|^2 < \rho \|F(x_k)\| \tag{2.36}$$

holds for all $k \geq \bar{k}$ such that $\zeta_k \leq \zeta$.

Proof By Step 2 of Algorithm 1, it suffices to show that there is a constant ζ such that (2.36) holds when $\zeta_k \leq \zeta$ and k is sufficiently large. Since, by Theorem 2.1, $\{x_k\}$ converges to the unique solution of (1.1), say, x^*, there exists a constant $M_2 > 0$ such that $\|A_{k+1}^{-1}\| \leq M_2$ for all k sufficiently large. In a way similar to the proof of (2.33), we can

show that there exist constants $\zeta' > 0$ and $C_2 > 0$ such that when $\zeta_k \leq \zeta'$ and k is large enough

$$\|p_k\| \leq C_2 \|F(x_k)\|. \tag{2.37}$$

We also have from (2.6)

$$\begin{aligned} A_{k+1}(x_k + p_k - x^*) &= A_{k+1}(x_k - x^*) + (A_{k+1} - B_k)p_k - F(x_k) \\ &= (A_{k+1} - \nabla F(x^*))(x_k - x^*) + (A_{k+1} - B_k)p_k \\ &\quad - F(x_k) + F(x^*) + \nabla F(x^*)(x_k - x^*). \end{aligned}$$

This implies

$$\begin{aligned} \|x_k + p_k - x^*\| &\leq \|A_{k+1}^{-1}\| \left(\|A_{k+1} - \nabla F(x^*)\| \, \|x_k - x^*\| + \|(A_{k+1} - B_k)p_k\| \right. \\ &\quad \left. + \|F(x_k) - F(x^*) - \nabla F(x^*)(x_k - x^*)\| \right) \\ &\leq M_2(\zeta_k \|p_k\| + o(\|x_k - x^*\|)) \\ &\leq M_2(C_2 \zeta_k \|F(x_k) - F(x^*)\| + o(\|x_k - x^*\|)) \\ &\leq M_2(C_2 M_3 \zeta_k \|x_k - x^*\| + o(\|x_k - x^*\|)), \end{aligned} \tag{2.38}$$

where $M_3 > 0$ stands for an upper bound of $\nabla F(x)$ on Ω, the second inequality follows from (2.25), and the third inequality follows from (2.37). It then follows that

$$\begin{aligned} \|F(x_k + p_k)\| &= \|F(x_k + p_k) - F(x^*)\| \\ &\leq M_3 \|x_k + p_k - x^*\| \\ &\leq M_2 M_3 (C_2 M_3 \zeta_k \|x_k - x^*\| + o(\|x_k - x^*\|)). \end{aligned} \tag{2.39}$$

On the other hand, by the nonsingularity of $\nabla F(x^*)$ and the fact that $x_k \to x^*$, there is a constant $m > 0$ such that

$$\|F(x_k)\| = \|F(x_k) - F(x^*)\| \geq m \|x_k - x^*\| \tag{2.40}$$

holds for all k sufficiently large. Therefore, we deduce from (2.37), (2.39) and (2.40) that when $\zeta_k \leq \zeta'$

$$\begin{aligned} &\|F(x_k + p_k)\| - \rho \|F(x_k)\| + \sigma_2 \|p_k\|^2 \\ &\quad \leq M_2 M_3 (C_2 M_3 \zeta_k \|x_k - x^*\| + o(\|x_k - x^*\|)) - \rho m \|x_k - x^*\| + \sigma_2 C_2 \|F(x_k)\|^2 \\ &\quad \leq M_2 M_3 (C_2 M_3 \zeta_k \|x_k - x^*\| + o(\|x_k - x^*\|)) - \rho m \|x_k - x^*\| + \sigma_2 C_2 M_3^2 \|x_k - x^*\|^2 \\ &\quad = -(\rho m - C_3 \zeta_k) \|x_k - x^*\| + o(\|x_k - x^*\|), \end{aligned} \tag{2.41}$$

where $C_3 = C_2 M_2 M_3^2$. Let $\zeta = \min\{\zeta', \frac{1}{2}\rho m C_3^{-1}\}$. Then the conclusion follows from (2.41). □

The following theorem establishes superlinear convergence of Algorithm 1.

Theorem 2.2 *Let Assumption A hold. Then the sequence $\{x_k\}$ generated by Algorithm 1 converges to the unique solution of (1.1) superlinearly.*

Proof Let ζ and $\bar{k}$ be as specified in Lemma 2.7. By Lemmas 2.3 and 2.6, (2.27) holds. Hence, there is an index $\tilde{k}$ such that when $k \geq \tilde{k}$

$$\frac{1}{k}\sum_{i=0}^{k-1} \zeta_i^2 \leq \frac{1}{2}\zeta^2.$$

This shows that for any $k \geq \tilde{k}$, there are at least $\lceil \frac{k}{2} \rceil$ many indices $i \leq k$ such that $\zeta_i^2 \leq \zeta^2$, or equivalently $\zeta_i \leq \zeta$. Let $k' = \max\{\bar{k}, \tilde{k}\}$. Then, by Lemma 2.7, for any $k \geq 2k'$, there are at least $\lceil \frac{k}{2} \rceil - k'$ many indices $i \leq k$ such that $\lambda_i = 1$ and

$$\|F(x_{i+1})\| = \|F(x_i + p_i)\| < \rho \|F(x_i)\|. \tag{2.42}$$

Let I_k be the set of indices for which (2.42) holds and let i_k be the number of elements in I_k. Then $i_k \geq \frac{k}{2} - k' - 1$. On the other hand, it is clear from Algorithm 1 that for each $i \notin I_k$, we have

$$\|F(x_{i+1})\| \leq \|F(x_i)\|(1 + \eta_i). \tag{2.43}$$

Multiplying inequalities (2.42) for $i \in I_k$ and (2.43) for $i \notin I_k$ from $i = k'$ to k yields

$$\|F(x_{k+1})\| \leq \rho^{i_k} \|F(x_{k'})\| \prod_{i=k'}^{k} (1 + \eta_i) \leq \|F(x_{k'})\| \rho^{\frac{1}{2}k - k' - 1} e^{\eta},$$

where the last inequality follows from (2.3). Therefore, we obtain

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

This together with (2.40) implies (2.28). Hence, by Lemma 2.6, we deduce $\zeta_k \to 0$. The superlinear convergence of $\{x_k\}$ then follows from (2.38). □

3. Norm Descent Quasi-Newton Method

The derivative-free line search proposed in the previous section is well defined as long as F is continuous. However, it does not possess norm descent property even if p_k provides a descent direction for the norm function $\frac{1}{2}\|F\|^2$ at x_k. This feature may deteriorate the computational effectiveness of iterative methods, though the results obtained in the previous section have shown its usefulness in globalizing quasi-Newton methods. In this section, we propose a derivative-free norm descent line search that is suitable for an iterative method for solving nonlinear equations involving a mapping with positive definite but not necessarily symmetric Jacobian matrices. It is clear that if $\nabla F(x)$ is positive definite, then $p = -F(x)$ is a descent direction of the norm function

$$\varphi(x) = \frac{1}{2}\|F(x)\|^2. \tag{3.1}$$

This suggests that if at some step k, the steplength λ_k determined by Procedure 1 does not satisfy $\|F(x_k + \lambda_k p_k)\| < \|F(x_k)\|$, then it would be beneficial to use $p_k = -F(x_k)$ as an alternative search direction instead of quasi-Newton direction, and determine a steplength $\lambda > 0$ such that

$$\|F(x_k + \lambda p_k)\| \leq \|F(x_k)\| - \sigma_3\|\lambda p_k\|^2, \tag{3.2}$$

where $\sigma_3 > 0$ is a given constant. Since p_k is a descent direction of φ at x_k, it is not difficult to show that the inequality (3.2) holds for all $\lambda > 0$ sufficiently small. Such a λ can be obtained by a backtracking process finitely (see Proposition 3.1 below).

Procedure 2 (Norm Descent Line Search):
Let $\beta \in (0,1)$ be given. Let $\lambda_k := \beta^{i_k}$ be determined by Procedure 1. If $\|F(x_k + \lambda_k p_k)\| \geq \|F(x_k)\|$ holds, then let $p_k := -F(x_k)$. For $i = 0, 1, \ldots$, check the inequality (3.2) with $\lambda = \beta^i$. Let $\bar{i}_k$ be the smallest nonnegative integer i such that (3.2) holds for $\lambda = \beta^i$. Let $\lambda_k := \beta^{\bar{i}_k}$.

The following proposition shows that Procedure 2 is well-defined.

Proposition 3.1 *Let F be continuously differentiable and $\nabla F(x)$ is positive definite. If $F(x_k) \neq 0$, then Procedure 2 terminates finitely.*

Proof It suffices to verify the existence of λ satisfying (3.2). Since $\nabla F(x_k)$ is positive definite, $\bar{p}_k = -F(x_k)$ is a descent direction of the function φ defined by (3.1) at x_k

whenever $F(x_k) \neq 0$. Let $\sigma \in (0,1)$ be a given constant. Then the inequality

$$\varphi(x_k + \lambda p_k) \leq \varphi(x_k) + \sigma\lambda\varphi'(x_k, p_k) \tag{3.3}$$

holds for all $\lambda > 0$ sufficiently small, where $\varphi'(x_k, p_k) = F(x_k)^T \nabla F(x_k) p_k$ denotes the directional derivative of φ at x_k along direction p_k. By the definition of φ, (3.3) is rewritten as

$$\|F(x_k + \lambda p_k)\|^2 - \|F(x_k)\|^2 \leq -2\sigma\lambda F(x_k)^T \nabla F(x_k) F(x_k).$$

Dividing the both sides by $\|F(x_k + \lambda p_k)\| + \|F(x_k)\|$ yields

$$\begin{aligned} & \|F(x_k + \lambda p_k)\| - \|F(x_k)\| \qquad (3.4) \\ \leq & -\frac{2\sigma\lambda F(x_k)^T \nabla F(x_k) F(x_k)}{\|F(x_k + \lambda p_k)\| + \|F(x_k)\|} \\ = & -\sigma_3 \|\lambda p_k\|^2 - \lambda\Big(\frac{2\sigma\lambda F(x_k)^T \nabla F(x_k) F(x_k)}{\|F(x_k + \lambda p_k)\| + \|F(x_k)\|} - \sigma_3\lambda\|p_k\|^2\Big). \qquad (3.5) \end{aligned}$$

It is easy to see that, when $\lambda > 0$ is sufficiently small, the last term of (3.5) is negative. This completes the proof. □

By using Procedure 2 for line search, we propose a norm descent quasi-Newton method as follows.

Algorithm 2 (Norm Descent Quasi-Newton Method):

Step 0. Choose constants $\rho, \beta \in (0,1)$, $\sigma_1, \sigma_2, \sigma_3 > 0$. Select a positive sequence $\{\eta_k\}$ satisfying (2.3). Choose an initial point $x_0 \in R^n$ and a nonsingular matrix $B_0 \in R^{n \times n}$. Let $k := 0$.

Step 1. Stop, if $F(x_k) = 0$. Otherwise, solve the linear equation (2.6) to get p_k.

Step 2. If (2.7) holds, then let $\lambda_k := 1$ and go to Step 4.

Step 3. Let λ_k be determined by Procedure 2.

Step 4. Let $x_{k+1} := x_k + \lambda_k p_k$.

Step 5. Update B_k by Broyden-like formula (2.8).

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

We note that the only difference between Algorithms 1 and 2 lies in the line search step. Note that Lemma 2.6 remains true for Algorithm 2. Moreover, since Algorithm 2

is a descent algorithm, it is obvious that the generated sequence $\{x_k\}$ is contained in the level set

$$\bar{\Omega} = \{x \in R^n \mid \|F(x)\| \leq \|F(x_0)\|\}. \tag{3.6}$$

It is also not difficult to see that for each k

$$\bar{\sigma}_0\|s_k\|^2 = \bar{\sigma}_0\|x_{k+1} - x_k\|^2 \leq \|F(x_k)\| - \|F(x_{k+1})\| + \eta_k\|F(x_k)\|, \tag{3.7}$$

where $\bar{\sigma}_0 = \min\{\sigma_1, \sigma_2, \sigma_3\}$. Since η_k satisfies (2.3), (3.7) implies

$$\sum_{k=0}^{\infty} \|x_{k+1} - x_k\|^2 < \infty. \tag{3.8}$$

In order to establish global and superlinear convergence of Algorithm 2, we need the following assumption similar to but stronger than Assumption A.

Assumption A′ (i) The level set $\bar{\Omega}$ defined by (3.6) is bounded.

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

(iii) $\nabla F(x)$ is positive definite for every $x \in \bar{\Omega}$.

Assumption A′ is obtained by replacing the condition that $\nabla F(x)$ is nonsingular with the condition that $\nabla F(x)$ is positive definite in Assumption A.

By Lemma 2.6 and (3.8), we immediate have the following lemma.

Lemma 3.1 *Let Assumption A′ hold and $\{x_k\}$ be generated by Algorithm 2. Let ζ_k be defined by (2.24). Then we have*

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=0}^{k-1} \zeta_i^2 = 0. \tag{3.9}$$

In particular, there is a subsequence of $\{\zeta_k\}$ converging to zero.

Now, we show global convergence of Algorithm 2.

Theorem 3.1 *Let Assumption A′ hold. Then the sequence $\{x_k\}$ generated by Algorithm 2 converges to the unique solution of (1.1).*

Proof Since $\{\|F(x_k)\|\}$ is a decreasing sequence, it suffices to verify

$$\liminf_{k\to\infty} \|F(x_k)\| = 0. \tag{3.10}$$

We consider three cases. Case I: There are infinitely many k such that (2.7) holds; Case II: There are infinitely many k such that λ_k is determined by the backtracking process with $p_k = -B_k^{-1}F(x_k)$; and Case III: λ_k is determined by the backtracking process with $p_k = -F(x_k)$ for all k sufficiently large.

Case I is trivial. Case II is a particular case of Theorem 2.1. So, it suffices to verify (3.10) for Case III. Let $\tilde{\lambda} = \limsup_{k\to\infty} \lambda_k$. If $\tilde{\lambda} > 0$, then there is a subsequence of $\{p_k\}$ converging to zero, or equivalently (3.10) holds. Now, suppose $\tilde{\lambda} = 0$. That is, $\lim_{k\to\infty} \lambda_k = 0$. Then for all k sufficiently large, $\lambda_k' \stackrel{\triangle}{=} \lambda_k/\beta$ does not satisfy (3.2). This means

$$\|F(x_k + \lambda_k' p_k)\| - \|F(x_k)\| \geq -\sigma_3 \|\lambda_k' p_k\|^2 = -\sigma_3 \|\lambda_k' F(x_k)\|^2. \tag{3.11}$$

Since $\{x_k\} \subset \bar{\Omega}$ is bounded, there is a subsequence $\{x_k\}_{k\in\tilde{K}}$ converging to some point $\tilde{x}$. Multiplying the both sides of (3.11) by $(\|F(x_k + \lambda_k' p_k)\| + \|F(x_k)\|)/\lambda_k'$ and letting $k \to \infty$ with $k \in \tilde{K}$ yield

$$(\nabla F(\bar{x})F(\bar{x}))^T \bar{p} = -F(\bar{x})^T \nabla F(\bar{x})^T F(\bar{x}) \geq 0.$$

By the positive definiteness of $\nabla F(\bar{x})$, this implies $F(\bar{x}) = 0$. The proof is then complete. □

In a way similar to Lemma 2.7, we can prove the following lemma.

Lemma 3.2 *Let Assumption A′ hold and $\{x_k\}$ be generated by Algorithm 2. Then there exist a constant $\zeta > 0$ and an index $\bar{k}$ such that $\lambda_k = 1$ whenever $\zeta_k \leq \zeta$ and $k \geq \bar{k}$. Moreover, for those k, we have*

$$\|F(x_k + p_k)\| \leq \rho \|F(x_k)\| - \sigma_2 \|p_k\|^2 < \rho \|F(x_k)\|.$$

Lemma 3.2 reveals the fact that for all $k \geq \bar{k}$ such that $\zeta_k \leq \zeta$, the steplength λ_k is unity and is determined in Step 2. Therefore, if $\zeta_k \leq \zeta$ for all k sufficiently large, Algorithm 2 reduces to Algorithm 1. In a way similar to Theorem 2.2, we can show the following superlinear convergence theorem for Algorithm 2. We omit the proof.

Theorem 3.2 *Let Assumption A′ hold. Then the sequence $\{x_k\}$ generated by Algorithm 2 converges to the unique solution of (1.1) superlinearly.*

4. Final Remark

We have proposed a quasi-Newton method for solving nonlinear equations. Moreover, for a nonlinear equation involving a mapping with positive definite Jacobian matrices, we have proposed a descent quasi-Newton method. We have shown that both methods are well-defined and converge globally and superlinearly under appropriate conditions. To relax those somewhat restrictive assumptions is certainly an interesting and important subject for future research.

References

[1] C.G. Broyden, *A class of methods for solving nonlinear simultaneous equations*, Math. Comput., 19 (1965), pp. 577-593.

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

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

[4] J.E. Dennis, Jr. and J.J. Moré, *Quasi-Newton methods, Motivation and the theory*, SIAM Rev., 19 (1977), pp. 46-89.

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

[6] D.H. Li and M. Fukushima, *A globally and superlinearly convergent Gauss-Newton based BFGS method for symmetric equations*, to appear in SIAM J. Numer. Anal..

[7] D.H. Li and M. Fukushima, *A derivative-free line search and DFP method for symmetric equations with global and superlinear convergence*, Numer. Funct. Anal. Optim. 20 (1999), No. 2.

[8] D.H. Li and M. Fukushima, *Broyden-like methods for variational inequality and nonlinear complementarity problems with global and superlinear convergence*, Techni-

cal Report 98016, Department of Applied Mathematics and Physics, Kyoto University, September 1998.

[9] D.H. Li and M. Fukushima, *Smoothing Newton and quasi-Newton methods for mixed complementarity problems*, Technical Report 99001, Department of Applied Mathematics and Physics, Kyoto University, January 1999.

[10] D.H. Li, N. Yamashita and M. Fukushima, *A nonsmooth equation based BFGS method for solving KKT systems in mathematical programming*, Technical Report 98012, Department of Applied Mathematics and Physics, Kyoto University, July 1998.