Hokkaido Mathematical Journal Vol. 36 (2007) p. 729–743

Recent progress in the global convergence of quasi-Newton methods for nonlinear equations

Dong-Hui Li and Wanyou Cheng

(Received February 1, 2007; Revised May 11, 2007)

Abstract. The global convergence theory of quasi-Newton methods for optimization problems has well been established. Related work to the globalization of quasi-Newton methods for nonlinear equations is relatively less. The major difficulty in globalizing quasi-Newton methods for nonlinear equations lies in the lack of efficient line search technique. Recently, there have been proposed some derivative-free line searches. The study in the global convergence of some quasi-Newton methods has taken good progress. In this paper, we summarize some recent progress in the global convergence of quasi-Newton methods for solving nonlinear equations.

Key words: Nonlinear equation, quasi-Newton method, derivative-free line search, global convergence.

1. Introduction

Consider the problem of finding a solution to the nonlinear equation

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

where $F\colon R^n \to R^n$. Suppose $F(x)$ is continuously differentiable whose Jacobian is denoted by $J(x)$. If $F(x)$ is the gradient of some continuously differentiable function $f\colon R^n \to R$, then equation (1.1) is the first order necessary optimality condition of the following unconstrained optimization problem

$$\min f(x), \quad x \in R^n. \tag{1.2}$$

Function F is the gradient of some real-valued function if and only if $J(x)$ is symmetric for all x [22]. In general, if we let $\theta(x) = (1/2)\|F(x)\|^2$, problem (1.1) can be converted into the following global optimization problem

$$\min \theta(x), \quad x \in R^n. \tag{1.3}$$

2000 Mathematics Subject Classification : 65H10, 90C53.
Supported by the 973 project (2004CB719402) and the NSF (10471036) of China

Here and throughout the paper, we use $\|\cdot\|$ to denote the Euclidean norm of vectors. Quasi-Newton methods form an important class of iterative methods for solving nonlinear equations and optimization problems. A quasi-Newton method for solving (1.1) generates a sequence of iterates $\{x_k\}$ by the iterative process

$$x_{k+1} = x_k + d_k, \quad k = 0, 1, \ldots,$$

starting from some initial point x_0. The direction d_k is called a quasi-Newton direction, which is a solution of the following system of linear equations

$$B_k d_k + F(x_k) = 0, \tag{1.4}$$

where B_k is some matrix which is an approximation to $J(x_k)$. When $B_k = J(x_k)$, the iterative scheme is the well known Newton's method.

As an approximation to $J(x_k)$, B_k satisfies the so-called secant equation

$$B_{k+1} s_k = y_k, \tag{1.5}$$

where $s_k = x_{k+1} - x_k$ and $y_k = F(x_{k+1}) - F(x_k)$. By the mean-value theorem, we have

$$y_k = \Big(\int_0^1 J(x_{k+1} - \tau s_k)d\tau\Big)s_k \approx J(x_{k+1})s_k.$$

In this sense, B_{k+1} is an approximation to $J(x_{k+1})$. Suppose we have had B_k. Matrix B_{k+1} is obtained by updating B_k with some lower rank matrix. Well known update formulae include the Broyden's rank one formula,

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)s_k^T}{\|s_k\|^2}, \tag{1.6}$$

the BFGS formula,

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} \tag{1.7}$$

and the DFP formula

$$B_{k+1} = \Big(I - \frac{y_k s_k^T}{y_k^T s_k}\Big) B_k \Big(I - \frac{y_k s_k^T}{y_k^T s_k}\Big)^T + \frac{y_k y_k^T}{y_k^T s_k}, \tag{1.8}$$

where $I \in R^{n\times n}$ is the identity matrix.

If $F(x_k)$ is replaced by $\nabla f(x_k)$, the gradient of f at x_k, then the methods are quasi-Newton methods for solving unconstrained optimization problem (1.2).

It is easy to see from (1.7) and (1.8) that matrix B_{k+1} inherits the symmetry of B_k. In addition, B_{k+1} will be positive definite if $y_k^T s_k > 0$ and B_k is positive definite. Thus, when applied to solve optimization problem (1.2), the corresponding quasi-Newton direction is a descent direction of f at x_k. The BFGS and DFP methods are widely used in optimization or symmetric equations.

Matrix B_{k+1}in Broyden's rank one update formula (1.6) is generally not symmetric even if B_k is symmetric. We note that B_{k+1} may be singular even if B_k is nonsingular. As a result, the quasi-Newton direction may not exist. As a remedy, the so-called Broyden-like formula was proposed by Powell [24]

$$B_{k+1} = B_k + \mu_k \frac{(y_k - B_k s_k)s_k^T}{\|s_k\|^2}, \tag{1.9}$$

where $\mu_k \in (0,\, \mu)$ with some constant $\mu \in (0,\, 1)$ is chosen so that B_{k+1} is nonsingular. We refer to [21] for more details.

The following theorem gives a necessary and sufficient condition for a quasi-Newton method to be superlinearly convergent. The proof can be found in [7].

Theorem 1.1 *Let $F\colon R^n \to R^n$ be continuously differentiable with Lipschitz continuous Jacobian. Let $\{B_k\}$ be a sequence of nonsingular matrices. Suppose that x^* is a solution of* (1.1) *at which $J(x^*)$ is nonsingular and that the sequence*

$$x_{k+1} = x_k - B_k^{-1} F(x_k), \quad k = 0,\, 1,\, \ldots$$

converges to x^*. Then $\{x_k\}$ converges superlinearly to x^* if and only if

$$\lim_{k\to\infty} \frac{\|[B_k - F'(x^*)](x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0. \tag{1.10}$$

Condition (1.10) is called Dennis-Moré condition. It has become a fundamental criterion to check if a quasi-Newton method is superlinearly convergent. Most existing quasi-Newton methods including the Broyden-like methods, the BFGS and DFP methods satisfy the Dennis-Moré condition and hence are superlinearly convergent. We refer to [7] for a very good review in the local convergence of quasi-Newton methods.

The purpose of this paper is to emphasize the global convergence of the quasi-Newton methods. In the next section, we summarize some derivative-free line searches and related globally convergent quasi-Newton methods for general nonlinear equations. In Section 3, we pay particular attention to the quasi-Newton methods for monotone equations. We conclude the paper by giving some further research topics.

2. Derivative-free line searches

2.1. Globally convergent quasi-Newton methods for unconstrained optimization

Let us first recall global convergence of quasi-Newton methods for unconstrained optimization problems. A widely used strategy to globalize a quasi-Newton method is to adopt a line search procedure. This results in the following damped iterative process:

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, \ldots,$$

where d_k is a quasi-Newton direction and α_k is called steplength that is determined by some line search procedure. For optimization problems, if d_k is a descent direction of f at x_k, i.e., $\nabla f(x_k)^T d_k < 0$, then α_k can be obtained by Armijo line search or Wolfe line search. Specifically, in Armijo line search, α_k satisfies the condition

$$f(x_k + \alpha_k d_k) \le f(x_k) + \sigma_1 \alpha_k \nabla f(x_k)^T d_k, \tag{2.1}$$

where $\sigma_1 \in (0, 1)$ is a constant. In Wolfe line search, α_k satisfies the following two inequalities

$$\begin{cases} f(x_k + \alpha_k d_k) \le f(x_k) + \sigma_1 \alpha_k \nabla f(x_k)^T d_k, \\ \nabla f(x_k + \alpha_k d_k)^T d_k \ge \sigma_2 \nabla f(x_k)^T d_k, \end{cases} \tag{2.2}$$

where constants σ_1, σ_2 satisfy $0 < \sigma_1 < \sigma_2 < 1$. It is clear that if B_k is positive definite, the quasi-Newton direction is a descent direction of f at x_k. For most existing quasi-Newton methods including the BFGS, DFP and the restricted Broyden's class of quasi-Newton methods, which is a convex combination of BFGS and DFP method, if the objective function f is uniformly convex, then the positive definiteness of B_k is guaranteed. The global convergence of Broyden's class of quasi-Newton methods except for DFP method for solving uniformly convex optimization problems has been well established [3, 4, 25]. In the case where f is not uniformly convex, the Broyden's class of quasi-Newton methods still generate descent directions if Wolfe line search is used. However, in this case, an example given by Dai [6] shows that the BFGS method is not necessary globally convergent. To make the BFGS method be globally convergent for nonconvex minimization problems, some modification is necessary [16, 17].

2.2. Derivative-free line searches for nonlinear equations

The study in the global convergence of quasi-Newton methods for optimization has taken good successes. On the other hand, however, for nonlinear equations, the related work is relatively less. Early in 1970, Powell [23] proposed a Newton-Broyden type hybrid method. This method is globally and R-superlinearly convergent [21].

In general, the globalization of quasi-Newton methods is difficult. The major difficulty is the lack of practical line search technique. As pointed out in the previous section, the nonlinear equation (1.1) can be converted into the unconstrained optimization problem (1.2) or (1.3). However, the quasi-Newton methods for solving (1.1) and (1.2) or (1.3) are quite different. In (1.3), the gradient of θ relies on the Jacobian of F which should be avoided in quasi-Newton methods. Even for the case where $J(x)$ is symmetric, it is not desirable to get a solution of (1.1) through (1.2) because the computation of the objective function of (1.2) is very cost.

Observe that for nonlinear equations, the quasi-Newton direction d_k is generally not a descent direction of θ at x_k even if B_k is positive definite. Existing line searches such as the Armijo or Wolfe line search depend on the computation of derivatives. They are not available for the quasi-Newton method for solving nonlinear equations. In order to globalize a quasi-Newton method for solving nonlinear equation (1.1), we need some derivative-free line search technique. The first derivative-free line search was due to Griewank [8] in which α_k satisfies the following condition

$$-\frac{(F(x_k+\alpha_k d_k)-F(x_k))^T F(x_k)}{\|F(x_k+\alpha_k d_k)-F(x_k)\|^2} \geq \frac{1}{2}+\epsilon, \tag{2.3}$$

where $\epsilon \in (0,\, 1/6)$ is a constant. It is clear that if $F(x_k)^T J(x_k) d_k \neq 0$, then inequality (2.3) is satisfied for all α_k (positive or negative) sufficient small. Under certain conditions, Griewank [8] established the global convergence of Broyden's rank one method. However, as pointed out by Griewank [8], there is a difficult case for this line search. That is, if $F(x_k)^T J(x_k)d_k = 0$, then the existence of α_k satisfying (2.3) is not guaranteed.

Line search (2.3) is a monotone line search in the sense that $\|F(x_{k+1})\| < \|F(x_k)\|$. Specifically, we have

$$\|F(x_{k+1})\|^2 - \|F(x_k)\|^2 \leq -2\epsilon\|F(x_{k+1}) - F(x_k)\|^2.$$

Another monotone derivative-free line search was proposed by Bellavia, Gasparo and Macconi [1] in the so called switching method. The line search condition there is

$$\|F(x_k + \alpha_k d_k)\|^2 \leq (1 - 2\alpha)\|F(x_k)\|^2, \tag{2.4}$$

where $0 < \alpha < 1/2$ and d_k is generated by finite-difference Newton method. This line search seems not suitable for quasi-Newton method because the existence of α_k in (2.4) is not guaranteed if d_k is some quasi-Newton direction.

Recently, a well defined derivative-free line search was proposed by Li and Fukushima [12]. In this line search, α_k satisfies the following inequality

$$\begin{aligned} &\|F(x_k + \alpha_k d_k)\|^2 - \|F(x_k)\|^2 \\ &\qquad \leq -\sigma_1\|\alpha_k F(x_k)\|^2 - \sigma_2\|\alpha_k d_k\|^2 + \epsilon_k\|F(x_k)\|^2, \end{aligned} \tag{2.5}$$

where σ_1 and σ_2 are positive constants, and the positive sequence $\{\epsilon_k\}$ satisfies

$$\sum_{k=0}^{\infty} \epsilon_k < \infty. \tag{2.6}$$

It is not difficult to see that inequality (2.5) holds for all $\alpha_k > 0$ sufficiently small as long as $F(x_k) \neq 0$. Consequently, it is well defined and can be implemented by some backtracking process. Line search condition (2.5) is not monotone. In other words, the sequence $\{\theta(x_k)\}$ is not necessarily decreasing. However, as x_k goes to a solution of (1.1) the last term in (2.5) becomes very small. In this sense, it is approximately norm descent. This nonmonotone line search enjoys some good properties. It is easy to show that if $\{\|F(x_k)\|\}$ is bounded, then we have

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

Moreover, the function value sequence $\{\theta(x_k)\}$ is convergent. Consequently, if there is an accumulation point of $\{x_k\}$ that solves (1.1), then every accumulation point of $\{x_k\}$ will be a solution of (1.1).

With derivative-free line search (2.5), Li and Fukushima [14] established the global convergence of Broyden-like methods.

Theorem 2.1 *Let F be continuously differentiable and $J(x)$ be Lipschitz continuous. Suppose further that $J(x)$ is uniformly nonsingular, i.e., there is a constant $M > 0$ such that $\|J(x)^{-1}\| \le M$. Then the sequence $\{x_k\}$ generated by the Broyden-like method with line search* (2.5) *converges to the unique solution of* (1.1). *Moreover, the convergence rate is superliner.*

There are some other derivative-free line searches [5, 9, 13, 31]. In [13], Li and Fukushima developed a derivative-free line search similar to (2.5). The condition of α_k in [13] is

$$\begin{aligned}&\|F(x_k+\alpha_k d_k)\| - \|F(x_k)\| \\ &\qquad \le -\sigma_1\|\alpha_k F(x_k)\| - \sigma_2\|F(x_k+\alpha_k d_k) - F(x_k)\| + \epsilon_k, \end{aligned} \tag{2.7}$$

where σ_1 and σ_2 are positive constants and ϵ_k satisfies (2.6). An attractive property of this line search is that the sequence $\{x_k\}$ satisfies

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

Consequently, the whole sequence $\{x_k\}$ converges to some point. Under the same conditions as those in Theorem 2.1, Li and Fukushima [13] proved the global convergence of the DFP method for symmetric nonlinear equations.

Theorem 2.2 *Let the conditions in* Theorem 2.1 *hold. Suppose that $J(x)$ is symmetric for any x. Then the sequence $\{x_k\}$ generated by DFP with line search* (2.7) *converges to the unique solution of* (1.1). *Moreover, the convergence rate is superlinear if at the limit x^*, $J(x^*)$ is positive definite.*

Another interesting derivative-free line search was proposed by Birgin, Krejić and Martínez [2] in which α_k satisfies

$$\|F(x_k+\alpha_k d_k)\| \le (1+\alpha_k\sigma(\beta-1))\|F(x_k)\| + \epsilon_k, \tag{2.8}$$

where $\sigma \in (0, 1)$ and $\beta \in [0, 1)$ are constants and ϵ_k satisfies (2.6). A more general global convergence theorem for inexact quasi-Newton methods has been established in [2]. Quite recently, by combining the nonmonotone line search of Grippo, Lampariello and Lucidi [10] and Li and Fukushima line search [16], Cruz, Martínez and Raydon [5] proposed a spectral residual method where the derivative-free line search condition is

$$f(x_k + \alpha_k d_k) \leq \max_{0 \leq j \leq M-1} f(x_{k-j}) + \epsilon_k - \gamma \alpha_k^2 f(x_k),$$

where M is a positive integer, $\gamma > 0$ is a constant and ϵ_k satisfies (2.6). The reported numerical results show that this line search works very well.

2.3. Symmetric equations

By symmetric equation, we mean that the Jacobian $J(x)$ is symmetric for all x. Symmetric equations come from the first order necessary condition of the unconstrained optimization problem, the KKT system of the equality constrained optimization problem, the discretization for some differential equation and so on. When we apply a quasi-Newton method to solve a symmetric equation, it is reasonable to use a symmetric quasi-Newton matrix B_k to approximate $J(x_k)$. Li and Fukushima [12] proposed a Gauss-Newton based BFGS method for solving symmetric equations. Unlike (1.4), the quasi-Newton direction in [12] is an approximate quasi-Newton direction for solving unconstrained optimization problem (1.3). Let

$$g_k = \alpha_{k-1}^{-1}(F(x_k + \alpha_{k-1} F(x_k)) - F(x_k)),$$

where α_{k-1} is the steplength obtained in the previous iteration. If α_{k-1} or $F(x_k)$ is small, then we have $g_k \approx \nabla\theta(x_k) = J(x_k)F(x_k)$. Based on this observation, Li and Fukushima [12] proposed a Gauss-Newton based BFGS method in which d_k is a solution of the following system of linear equations:

$$B_k d + g_k = 0. \tag{2.9}$$

Matrix B_k in (2.9), as an approximation to $J(x_k)^T J(x_k) = J(x_k)^2$ is updated by (1.7) but the meaning of y_k is different. Specifically, the vector y_k in Li and Fukushima's method is defined by

$$\delta_k = F(x_{k+1}) - F(x_k),$$
$$y_k = F(x_k + \delta_k) - F(x_k) \approx J(x_k)\delta_k \approx J(x_k)^2 s_k.$$

This Gauss-Newton based BFGS method with line search (2.5) is also globally and superlinear convergent under the condition that $J(x)$ is uniformly nonsingular [12].

All the above mentioned line searches are not monotone. Indeed, the quasi-Newton direction d_k is generally not a descent direction of θ at x_k. However, for Gauss-Newton method, it is possible to get a norm descent quasi-Newton direction. Note that if we replace g_k in (2.9) by $q_k = \nabla\theta(x_k) = J(x_k)F(x_k)$, then the unique solution d_k of (2.9) satisfies

$$\nabla\theta(x_k)^T d_k = -F(x_k)^T J(x_k)^T B_k^{-1} J(x_k) F(x_k).$$

If B_k is positive definite, then we have $\nabla\theta(x_k)^T d_k < 0$. This means that d_k is a descent direction of θ at x_k. Since g_k is continuous in α_{k-1}, when α_{k-1} is sufficiently small, the solution of (2.9) will also provides a descent direction of θ at x_k. This observation motivates us to find some small scalar λ_k instead of α_{k-1} in the definition of g_k so that the corresponding quasi-Newton direction is a descent direction of θ at x_k. In what follows, we gives some details in [11] to find a descent quasi-Newton direction.

Let

$$g_k(\lambda) = (F(x_k + \lambda F(x_k)) - F(x_k))/\lambda. \tag{2.10}$$

Suppose that B_k is positive definite. Consider the system of linear equations with parameter λ:

$$B_k d + g_k(\lambda) = 0. \tag{2.11}$$

Let $d(\lambda)$ be the solution of (2.11). It is not difficult to show that when $\lambda > 0$ is sufficiently small, every solution of (2.11) is a descent direction of θ at x_k. Specifically, we have the following lemma [11].

Lemma 2.1 *Let σ_1 and σ_2 be positive constants and B_k be a symmetric and positive definite matrix. If x_k is not a stationary point of* (1.3)*, then there exists a constant $\bar\lambda > 0$ depending on k such that when $\lambda \in (0, \bar\lambda)$, the unique solution $d(\lambda)$ of* (2.11) *satisfies $\nabla\theta(x_k)^T d(\lambda) < 0$. Moreover, the inequality*

$$\theta(x_k + \lambda d(\lambda)) - \theta(x_k) \le -\sigma_1\|\lambda d(\lambda)\|^2 - \sigma_2\|\lambda F(x_k)\|^2 \tag{2.12}$$

holds for all $\lambda > 0$ sufficiently small.

Lemma 2.1 provides a way to find a descent quasi-Newton direction by adjusting parameter λ.

Procedure 1 Let constant $\rho \in (0, 1)$ be given. Let i_k be the smallest nonnegative integer such that inequality (2.12) holds with $\lambda = \rho^i$, $i = 0, 1, \ldots$. Let $d_k = d(\rho^{i_k})$ and $g_k = g_k(\rho^{i_k})$.

To enlarge steplength, [11] adopted the following forward procedure.

Procedure 2 Let i_k and d_k be determined by Procedure 1. If $i_k = 0$, we let $\lambda_k = 1$. Otherwise, let j_k be the largest positive integer $j \in \{0, 1, 2, \ldots, i_k - 1\}$ satisfying

$$\theta(x_k + \rho^{i_k - j} d_k) - \theta(x_k) \leq -\sigma_1 \|\rho^{i_k - j} d_k\|^2 - \sigma_2 \|\rho^{i_k - j} F(x_k)\|^2.$$

Let $\alpha_k = \rho^{i_k - j_k}$.

It is not difficult to see that Procedures 1 and 2 are well-defined. Procedures 1 and 2 give a way to find a descent quasi-Newton direction as long as B_k is positive definite. Recall that in BFGS formula (1.7), if B_k is positive definite, then B_{k+1} is positive definite if and only if $y_k^T s_k > 0$. While solving unconstrained optimization problem, the condition $y_k^T s_k > 0$ is satisfied if Wolfe line search (2.2) is used. For nonlinear equations, however, the Wolfe line search is no longer available because it depends on the computation of the gradient. To ensure the positive definiteness of B_k, similar to [17], Gu, Li, Qi and Zhou [11] proposed a modified BFGS formula (1.7) in which y_k is defined by

$$y_k = \gamma_k + \Big(\max\Big\{0, -\frac{\gamma_k^T s_k}{\|s_k\|^2}\Big\} + \phi(\|F(x_k)\|)\Big) s_k,$$

where $\gamma_k = F(x_k + \delta_k) - F(x_k)$, $\delta_k = F(x_{k+1}) - F(x_k)$, and function $\phi\colon R \to R$ satisfies (i) $\phi(t) > 0$ for all $t > 0$, (ii) $\phi(t) = 0$ if and only if $t = 0$, (iii) $\phi(t)$ is bounded if t is in a bounded set. In particular, $\phi(t) = \nu t$ with constant $\nu > 0$ meets the requirements. It is not difficult to show that y_k satisfies

$$y_k^T s_k \geq \max\{\gamma_k^T s_k, \phi(\|F(x_k)\|)\} \|s_k\|^2 > 0.$$

Consequently, B_{k+1} inherits the positive definiteness of B_k.

The following theorem establishes the global convergence of the descent BFGS method [11].

Theorem 2.3 *Let the level set*

$$\Omega = \{x \in R^n \mid \theta(x) \leq \theta(x_0)\}$$

be bounded, function F be continuously differentiable on Ω, and $J(x)$ be symmetric for every $x \in \Omega$. Let $\{x_k\}$ be generated by the descent BFGS method. Then we have

$$\liminf_{k\to\infty} \|\nabla\theta(x_k)\| = 0.$$

Moreover, suppose that there is a subsequence of $\{x_k\}$ converging to x^* at which $J(x^*)$ is nonsingular. Then x^* is a solution of (1.1). *Moreover, the whole sequence $\{x_k\}$ converges to x^*.*

It is worth mentioning that the global convergence of the descent BFGS method does not need the nonsingularity of $J(x)$.

The superlinear convergence of the descent BFGS method is stated as follows [11].

Theorem 2.4 *Let the conditions in* Theorem 2.3 *hold. Suppose that $J(x)$ is Lipschitz continuous. Then the sequence $\{x_k\}$ converges to x^* superlinearly.*

3. Monotone Equations

In this section, we pay particular attention to the monotone equations. By monotone equation (1.1), we mean that function F is monotone, namely,

$$(F(x) - F(y))^T(x - y) \geq 0, \quad \forall x, y \in R^n.$$

If F is continuously differentiable, then F is monotone if and only if $J(x)$ is positive semidefinite. The solution set S of monotone equation (1.1), if not empty, is convex.

Solodov and Svaiter [26] proposed a nice derivative-free line search and developed an inexact Newton method for solving monotone equations. At each iteration, the method uses a direction d_k satisfying

$$\|(J(x_k) + \mu_k I)d_k + F(x_k)\| \leq \rho_k \mu_k \|d_k\|,$$

where $\mu_k > 0$ and $\rho_k \in [0, 1)$ are parameters. The derivative-free line search is to find $\alpha_k = \beta^{m_k}$ with $\beta \in (0, 1)$ such that m_k is the smallest nonnegative integer m satisfying

$$-F(x_k + \beta^m d_k)^T d_k \geq \lambda(1 - \rho_k)\mu_k \|\|d_k\|, \tag{3.1}$$

where $\lambda \in (0, 1)$ is a constant. Let $z_k = x_k + \alpha_k d_k$. The next iterate is set to the projection of x_k to the hyperplane

$$\mathcal{H}_k = \{x \in R^n \mid F(z_k)^T(x - z_k) = 0\}.$$

That is,

$$x_{k+1} = x_k - \frac{F(z_k)^T(x_k - z_k)}{\|F(z_k)\|^2} F(z_k). \tag{3.2}$$

Note that the monotonicity of F implies that

$$F(z_k)^T(x_k - z_k) > 0.$$

This means that the hyperplane $\mathcal{H}$ strictly separates the current iterate x_k from zeros of the equation (1.1). Therefore, the project step can be regarded as an acceleration step. A nice property of this line search procedure is that the sequence $\{x_k\}$ satisfies for any $\bar{x} \in S$,

$$\|x_{k+1} - \bar{x}\|^2 \le \|x_k - \bar{x}\|^2 - \|x_{k+1} - x_k\|^2.$$

In particular, the distance from x_k to the solution set S decreases with k. Without the requirement of the nonsingularity of $J(x)$, Solodov and Svaiter [26] proved that the whole sequence $\{x_k\}$ converges to a solution of (1.1).

Zhou and Toh [30] extended Solodov and Svaiter's result and obtained the superlinear convergence of a Newton-type method even if the equation has singular solution.

Solodov and Svaiter's line search was applied to the BFGS method and limited memory BFGS method for solving monotone equation by Zhou and Li [28, 29]. Compared with Gauss-Newton based BFGS method where B_k is an approximation to $J(x_k)^T J(x_k)$, the quasi-Newton matrix in the methods of [28] and [29] is an approximation to $J(x_k)$. It is reasonable to believe that the subproblem in the latter methods is better conditioned than the former one. Theoretically, the methods are proved to be globally convergent without nonsingularity requirement of $J(x)$. In addition, the BFGS method in [29] retains superlinear convergence property. The reported preliminary numerical results in [28] and [29] showed that the limited memory BFGS method in [28] even performed better than the inexact Newton method did, and the performance of the BFGS method in [29] was much better than the Gauss-Newton based BFGS method.

4. Final remark

This review focused on the global convergence issue of quasi-Newton methods for solving nonlinear equations. We refer to a complete comprehensive review paper [20] on practical quasi-Newton methods for nonlinear equations. In particular, there listed 10 interesting open problems. To conclude, we give some other research topics related to the global convergence issue.

- The existing global convergence theory for Broyden's rank one method and Broyden-like methods needs the condition that the Jacobian $J(x)$ is uniformly nonsingular. From the theoretical point of view, it is important to remove this restriction.
- Is it possible to construct a globally convergent BFGS method for (nonmonotone) symmetric equations in which B_k is an approximation to $J(x_k)$?
- There are some works related to the global convergence of quasi-Newton methods for solving some nonsmooth equations arising from the nonlinear complementarity problem and the KKT system of the variational inequality [15, 18, 19, 27] etc. However, the strict complementarity condition is necessary in the superlinear convergence of these methods. This condition implicitly implies that the nonsmooth equation is locally smooth. It is interesting to release this condition.

References

[1] Bellavia S., Gasparo M.G. and Macconi M., *A switching-method for nonlinear systems*. Journal of Computational and Applied Mathematics **71** (1996), 83–93.

[2] Brigin E.G., Krejic N.K. and Martínez J.M., *Globally convergent inexact quasi-Newton methods for solving nonliear systems*. Numerical Algorithms **32** (2003), 249–260.

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

[4] Byrd R., Nocedal J. and Yuan Y., *Global convergence of a class of quasi-Newton methods on convex problems*. SIAM Journal on Numerical Analysis **24** (1987), 1171–1190.

[5] Cruz W., Martínez J. and Raydan M., *Special residual method without gradient information for solving large scale nonlinear systems of equations*. Mathematics of Computation **75** (2006), 1429–1448.

[6] Dai Y., *Convergence properties of the BFGS algorithm*. SIAM Journal on Optimization **13** (2002), 693–701.

[7] Dennis J.E. Jr. and Moré J.J., *Quasi-Newton methods, motivation and theory* SIAM Review **19** (1977), 46–89.

[8] Griewank A., *The 'global' convergence of Broyden-like methods with a suitable line search*. Journal of Australia Mathematics Society Ser. B **28** (1986), 75–92.

[9] Grippo L. and Sciandrone M., *Nonmonotone derivative-free methods for nonlinear equations*. to appear in Computational Optimization and Applications.

[10] Grippo L., Lampariello F. and Lucidi S., *A nonmonotone line search technique for Newton's method*. SIAM Journal on Numerical Analysis **23** (1986), 707–716.

[11] Gu G.Z., Li D.H., Qi L. and Zhou S.Z., *Descent directions of quasi-Newton methods for symmetric nonlinear equations*. SIAM Journal on Numerical Analysis **40** (2003), 1763–1774.

[12] Li D.H. and Fukushima M., *A globally and superlinear convergent Gauss-Newton-based BFGS method for symmetric nonlinear equations*. SIAM Journal on Numerical Analysis **37** (1999), 152–172.

[13] Li D.H. and Fukushima M., *A derivative-free line search and DFP method for symmetric equations with global and superlinear convergence*. Numerical Functional Analysys and Optimization **20** (1999), 59–77.

[14] Li D.H. and Fukushima M., *A derivative-free line search and global convergence of Broyden-like method for nonlinear equations*. Optimizaiton Methods Software **13** (2000), 181–201.

[15] Li D.H. and Fukushima M., *Smoothing Newton and quasi-Newton methods for mixed complementarity problems*. Computational Optimization and Applications **17** (2000), 203–230.

[16] Li D.H. and Fukushima M., *On the global convergence of BFGS method for nonconvex unconstrained optimization problems*. SIAM Journal on Optimization **11** (2001), 1054–1064.

[17] Li D.H. and Fukushima M., *A modified BFGS method and its global convergence in nonconvex minimization*. Journal of Computational and Applied Mathematics **129** (2001), 15–35.

[18] Li D.H. and Fukushima M., *Globally convergent Broyden-like methods for semismooth equations and applications to VIP, NCP and MCP*. Annals of Operations Research **103** (2001), 71–79.

[19] Li D.H., Yamashita N. and Fukushima M., *A nonsmooth equation based BFGS method for solving KKT systems in mathematical programming*. Journal of Optimization Theory and Applications **109** (2001), 123–167.

[20] Martínez J.M., *Practical quasi-Newton methods for solving nonlinear systems*. Journal of Computational and Applied Mathematics **124** (2000), 97–121.

[21] Moré J.J. and Trangenstein J.A., *On the global convergence of Broyden's method*. Mathematics of Computation **30** (1976), 523–540.

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

[23] Powell M.J.D., *A hybrid method for nonlinear equations.* in *"Numerical Methods for Nonlinear Algebraic Equations"* (P. Rabinowitz ed.), Gordon and Breach, London, 1970, pp. 87–114.

[24] Powell M.J.D., *A FORTRAN subroutine for solving systems of nonlinear algebraic equations.* in *"Numerical Methods for Nonlinear Algebraic Equations"*, (P. Rabinowitz ed.), Gordon and Breach, London, 1970, pp. 115–161.

[25] Powell M.J.D., *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.), Amer. Math. Soc. Providence. RI. (1976), 55–92.

[26] Solodov M.V. and Svaiter B.F., *A globally convergent inexact Newton method for systems of monotone equations.* in *"Reformulation: Nonsmooth, Piecewise smooth, Semismooth and Smoothing Methods"*, (M. Fukushima and L. Qi eds.), Kluwer Academic Publishers, 1998, pp. 355–369.

[27] Zhang J.W. and Li D.H., *A norm descent BFGS method for solving the KKT system of symmetric variational inequality problems.* Optimization Methods and Software **22** (2007), 237–252.

[28] Zhou W.J. and Li D.H., *Limited memory BFGS method for nonlinear monotone equations.* Journal of Computational Mathematics **25** (2007), 89–96.

[29] Zhou W.J. and Li D.H., *A globally convergent BFGS method for nonlinear monotone equations without any merit functions.* manuscrpts.

[30] Zhou G. and Toh K.C., *Superlinear convergence of a Newton-type algorithm for monotone equations.* Journal of Optimization Theory and Applications **125** (2005), 205–221.

[31] Zhu D.T., *Nonmonotone backtracking inexact quasi-Newton algorithms for solving smooth nonlinear equations.* Applied Mathematics and Computation **161** (2005), 875–895.

D.H. Li
College of Mathematics and Econometrics
Hunan University
Changsha, 410082 China
E-mail: dhli@hnu.cn

W. Cheng
College of Mathematics and Econometrics
Hunan University
Changsha 410082 China
E-mail: chengwanyou421@yahoo.com.cn