

## A NEW QUASI-NEWTON METHOD BASED ON ADJOINT BROYDEN UPDATES FOR SYMMETRIC NONLINEAR EQUATIONS

HUIPING CAO

ABSTRACT. In this paper, we propose a new rank two quasi-Newton method based on adjoint Broyden updates for solving symmetric nonlinear equations, which can be seen as a class of adjoint BFGS method. The new rank two quasi-Newton update not only can guarantee that  $B_{k+1}$  approximates Jacobian  $F'(x_{k+1})$  along direction  $s_k$  exactly, but also shares some nice properties such as positive definiteness and least change property with BFGS method. Under suitable conditions, the proposed method converges globally and superlinearly. Some preliminary numerical results are reported to show that the proposed method is effective and competitive.

### 1. Introduction

In this paper, we consider the problem of solving a system of nonlinear equation

$$(1) \quad F(x) = 0, \quad x \in \mathbb{R}^n,$$

where  $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a nonlinear mapping and continuously differentiable. When  $F$  is the gradient mapping of some function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , equation (1) is the first-order necessary condition for the following unconstrained optimization problem

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

For the nonlinear equations and unconstrained optimization problems, quasi-Newton methods have formed an important class of iterative methods for solving small and medium-scale problems. At each iteration of a quasi-Newton

---

Received September 5, 2015; Revised December 25, 2015.

2010 *Mathematics Subject Classification*. Primary 65K05, 65H10, 90C53.

*Key words and phrases*. adjoint Broyden update, quasi-Newton method, symmetric nonlinear equations, global convergence, superlinear convergence.

This work is supported by the National Natural Science Foundation of China No.11371154, 61502159 and the Fundamental Research Funds for the Central Universities 531107040846.

method, the quasi-Newton direction  $d_k$  is computed by solving the following system of linear equations

$$B_k d + F(x_k) = 0,$$

where  $B_k$  is an approximation to the Jacobian  $F'(x_k)$  that normally satisfies the following quasi-Newton condition (secant condition):

$$(2) \quad B_{k+1} s_k = y_k,$$

where  $s_k = x_{k+1} - x_k$  and  $y_k = F(x_{k+1}) - F(x_k)$ . The quasi-Newton matrix  $B_k$  can be updated by different quasi-Newton update formulae.

In this paper, we concentrate on symmetric nonlinear equations, which means that the Jacobian  $F'(x)$  of  $F$  defined by (1) is always symmetric for all  $x \in \mathbb{R}^n$ . The symmetric nonlinear equations have many practical backgrounds such as in the computation of the stationary points of unconstrained optimization problems, saddle points, large-scale scientific and engineering computing. For symmetric nonlinear equations, there have been many methods [3, 6, 8, 16] proposed for solving them, where BFGS method performs much better. The BFGS update formula takes the following form

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

For the study in the global convergence of nonlinear equations, the early work is due to [7], where Griewank established the global convergence of Broyden-like method. We can refer to a survey paper [9] for a summary about the global convergence of quasi-Newton methods for nonlinear equations.

There also has been some progress in the study of the numerical methods for solving symmetric nonlinear equations. Li and Fukushima [10] proposed a Gauss-Newton-based BFGS method for solving symmetric nonlinear equations and established the global and superlinear convergence. Based on the Gauss-Newton-based BFGS method, Gu et al. [8] proposed a norm descent BFGS method for solving symmetric nonlinear equations. The authors in [16] also studied quasi-Newton methods for solving symmetric nonlinear equations. Conjugate gradient type methods have also been applied to solve symmetric nonlinear equations. Li and Wang [11] extended the modified Fletcher-Reeves (FR) nonlinear conjugate gradient method proposed by Zhang et al. [17] to solve symmetric nonlinear equations. Zhou and Shen [18] also proposed a derivative-free Polak-Ribière-Polyak (PRP) method for solving symmetric nonlinear equations without the need of exact gradient and Jacobian, where the derivative-free method can be seen as a generalization of the classical PRP method used for solving unconstrained optimization problems [18].

Recently, a quasi-Newton method based on adjoint Broyden updates [13, 14] has been proposed for solving nonlinear equation. The adjoint Broyden update formula is

$$B_{k+1} = B_k + \frac{\sigma_k \sigma_k^T}{\sigma_k^T \sigma_k} (F'(x_{k+1}) - B_k),$$

where  $\sigma_k \in \mathbb{R}^n$  and  $\sigma_k \neq 0$ . Typical choices for  $\sigma_k$  have been given in [13]. Unlike existing quasi-Newton methods where  $B_{k+1}$  satisfies the secant equation (2) the matrix  $B_{k+1}$  in the adjoint Broyden method satisfies the so-called adjoint tangent condition

$$\sigma_k^T B_{k+1} = \sigma_k^T F'(x_{k+1}).$$

The adjoint Broyden method shares some nice properties with Broyden's method such as the least change property and the local superlinear convergence [13]. Moreover, it enjoys an affine invariant property respect to the scaling of the domain of nonlinear equations and possesses heredity property on affine problems when  $\sigma_k = (F'(x_{k+1}) - B_k)s_k$ . Extensive numerical results reported in [13] have shown that the adjoint Broyden method usually outperforms Newton's and Broyden's method in terms of runtime and iterations count, respectively.

Based on the satisfying performance of adjoint Broyden method, we propose a rank two quasi-Newton method for solving symmetric nonlinear equations, where the update formula is

$$(3) \quad B_{k+1} = B_k - \frac{B_k \sigma_k \sigma_k^T B_k}{\sigma_k^T B_k \sigma_k} + \frac{F'(x_{k+1}) \sigma_k \sigma_k^T F'(x_{k+1})}{\sigma_k^T F'(x_{k+1}) \sigma_k},$$

where  $\sigma_k \in \mathbb{R}^n$  and  $B_{k+1}$  is symmetric if  $B_k$  is symmetric. Firstly, we will give several lemmas about the rank two update formula: When  $B_k$  is positive definite, then  $B_{k+1}$  also is positive definite if and only if  $\sigma_k^T F'(x_{k+1}) \sigma_k > 0$ ; The matrix  $B_{k+1}$  is the unique solution of a variational problem. In this paper we set  $\sigma_k = s_k$ , then the above update formula fulfills the following condition

$$B_{k+1} s_k = F'(x_{k+1}) s_k,$$

which implies that  $B_{k+1}$  approximates  $F'(x_{k+1})$  along the direction  $s_k$  exactly. Moreover, we will present some results about the update formula, which show that our method possesses some favorable properties as the BFGS method [10, 16]: (a) the sequence generated by proposed method is norm descent, (b) the quasi-Newton matrices are positive definite, (c) the new method can obtain global and superlinear convergence. We also report some numerical results to verify the efficiency of the proposed method.

It can be seen that the update formula (3) includes  $F'(x_{k+1})$ , but it does not need to compute  $F'(x_{k+1})$  in practice. Because the forward and reverse mode of automatic differentiation provide the possibility to compute  $F'(x)s$  and  $\sigma^T F'(x)$  exactly within machine accuracy for given vectors  $x$ ,  $s$  and  $\sigma$ . According to [13], we can know that for complex function evaluations without special structure of  $F$ , since the computational effort for  $F'(x)s$  or  $\sigma^T F'(x)$  is equal to the evaluation of  $F$  times a constant  $c \leq 4$  independent of the dimension  $n$  of the state space. So the proposed method requires a computational effort in terms of function evaluations independent of the dimension  $n$  in each iteration.

The article is organized as follows. In Section 2, we derive a new rank two quasi-Newton update and give some nice properties. In Section 3, we propose a rank two quasi-Newton method for solving symmetric nonlinear equations. In Section 4, we establish the global and superlinear convergence of the proposed method under suitable conditions. Numerical results are presented in Section 5. Finally, we give the remarks.

## 2. The rank two quasi-Newton update

In this section, we will propose a new rank two quasi-Newton update based on adjoint tangent condition. Throughout the paper, we use  $\|\cdot\|$  to denote the Euclidean norm of vectors. We consider the following update

$$B_{k+1} = B_k + \Delta_k,$$

where  $\Delta_k$  is a rank two matrix and  $B_{k+1}$  satisfying the adjoint tangent condition

$$(4) \quad \sigma_k^T B_{k+1} = \sigma_k^T F'(x_{k+1}),$$

and  $\sigma_k \in \mathbb{R}^n$  and  $\sigma_k \neq 0$ . We let  $\Delta_k = a_k u_k u_k^T + b_k v_k v_k^T$ , where  $a_k, b_k$  are unknown constants and  $u_k, v_k \in \mathbb{R}^n$  are unknown vectors. According to the adjoint tangent condition (4), one has

$$\sigma_k^T B_k + a_k (\sigma_k^T u_k) u_k^T + b_k (\sigma_k^T v_k) v_k^T = \sigma_k^T F'(x_{k+1}),$$

which is equivalent to

$$(5) \quad a_k (\sigma_k^T u_k) u_k^T + b_k (\sigma_k^T v_k) v_k^T = \sigma_k^T F'(x_{k+1}) - \sigma_k^T B_k.$$

Vectors  $u_k$  and  $v_k$  are not unique, if we let

$$u_k^T = \beta_k \sigma_k^T B_k, \quad v_k^T = \gamma_k \sigma_k^T F'(x_{k+1}),$$

then we have

$$\Delta_k = a_k \beta_k^2 \cdot B_k \sigma_k \sigma_k^T B_k + b_k \gamma_k^2 \cdot F'(x_{k+1}) \sigma_k \sigma_k^T F'(x_{k+1}).$$

According to (5), it is easy to deduce

$$\left( a_k \beta_k^2 (\sigma_k^T B_k \sigma_k) + 1 \right) \sigma_k^T B_k + \left( b_k \gamma_k^2 (\sigma_k^T F'(x_{k+1}) \sigma_k) - 1 \right) \sigma_k^T F'(x_{k+1}) = 0.$$

So we can let

$$a_k \beta_k^2 = -\frac{1}{\sigma_k^T B_k \sigma_k}, \quad b_k \gamma_k^2 = \frac{1}{\sigma_k^T F'(x_{k+1}) \sigma_k},$$

i.e.,

$$\Delta_k = -\frac{B_k \sigma_k \sigma_k^T B_k}{\sigma_k^T B_k \sigma_k} + \frac{F'(x_{k+1}) \sigma_k \sigma_k^T F'(x_{k+1})}{\sigma_k^T F'(x_{k+1}) \sigma_k}.$$

Then we can obtain the following rank two update

$$(6) \quad B_{k+1} = B_k - \frac{B_k \sigma_k \sigma_k^T B_k}{\sigma_k^T B_k \sigma_k} + \frac{F'(x_{k+1}) \sigma_k \sigma_k^T F'(x_{k+1})}{\sigma_k^T F'(x_{k+1}) \sigma_k}.$$

It's obvious that  $B_{k+1}$  is symmetric when  $B_k$  is symmetric. If we denote  $H_k = B_k^{-1}$  and  $H_{k+1} = B_{k+1}^{-1}$ , then (6) can be written as

$$(7) \quad \begin{aligned} H_{k+1} = H_k - & \frac{H_k F'(x_{k+1}) \sigma_k \sigma_k^T + \sigma_k \sigma_k^T F'(x_{k+1}) H_k}{\sigma_k^T F'(x_{k+1}) \sigma_k} \\ & + \left( 1 + \frac{\sigma_k^T F'(x_{k+1}) H_k F'(x_{k+1}) \sigma_k}{\sigma_k^T F'(x_{k+1}) \sigma_k} \right) \cdot \frac{\sigma_k \sigma_k^T}{\sigma_k^T F'(x_{k+1}) \sigma_k}. \end{aligned}$$

*Remark a.* The update formula (6) can approximate the directional derivatives of  $F$  without the need of the Jacobian. And the forward and reverse mode of automatic differentiation provide the possibility to compute  $F'(x)s$  and  $\sigma^T F'(x)$  exactly within machine accuracy for given vectors  $x, s$  and  $\sigma$ . So it does not need to compute the Jacobian in practice for the proposed method.

For an  $n \times n$  matrix  $A$ ,  $A > 0$  indicates that  $A$  is positive definite. In what follows, we will give several lemmas to show some nice properties of (6), which can be proved similarly to that in [5, 6, 11].

**Lemma 2.1.** *We suppose that  $B_k > 0$  and  $B_{k+1}$  is updated by (6), then  $B_{k+1} > 0$  if and only if  $\sigma_k^T F'(x_{k+1}) \sigma_k > 0$ .*

*Proof.* According to the adjoint tangent condition (4), one has

$$\sigma_k^T F'(x_{k+1}) \sigma_k = \sigma_k^T B_{k+1} \sigma_k.$$

Then if  $B_{k+1}$  is positive definite, we have  $\sigma_k^T F'(x_{k+1}) \sigma_k > 0$ .

We assume that  $\sigma_k^T F'(x_{k+1}) \sigma_k > 0$  and  $B_k > 0$ . Then for  $\forall d_k \in \mathbb{R}^n$  and  $d_k \neq 0$ , we have from (6) that

$$d_k^T B_{k+1} d_k = d_k^T B_k d_k - \frac{(d_k^T B_k \sigma_k)^2}{\sigma_k^T B_k \sigma_k} + \frac{(d_k^T F'(x_{k+1}) \sigma_k)^2}{\sigma_k^T F'(x_{k+1}) \sigma_k}.$$

By the positive definiteness of  $B_k$ , there exists a symmetric and positive definite matrix  $B_k^{1/2}$  such that  $B_k = B_k^{1/2} B_k^{1/2}$ . It can be deduced by Cauchy-Schwarz inequality that

$$(8) \quad \begin{aligned} (d_k^T B_k \sigma_k)^2 &= \left( (B_k^{1/2} d_k)^T (B_k^{1/2} \sigma_k) \right)^2 \\ &\leq \|B_k^{1/2} d_k\|^2 \cdot \|B_k^{1/2} \sigma_k\|^2 \\ &= (d_k^T B_k d_k) (\sigma_k^T B_k \sigma_k), \end{aligned}$$

where the equality holds if and only if  $d_k = \lambda_k \sigma_k$ ,  $\lambda_k \neq 0$ .

When inequality (8) holds strictly, then we have

$$d_k^T B_{k+1} d_k > d_k^T B_k d_k - d_k^T B_k d_k + \frac{(d_k^T F'(x_{k+1}) \sigma_k)^2}{\sigma_k^T F'(x_{k+1}) \sigma_k} \geq 0.$$

When (8) is an equality, i.e., there exists a  $\lambda_k \neq 0$  such that  $d_k = \lambda_k \sigma_k$ , then we can get from (8) that

$$d_k^T B_{k+1} d_k \geq \frac{(d_k^T F'(x_{k+1}) \sigma_k)^2}{\sigma_k^T F'(x_{k+1}) \sigma_k} = \lambda_k^2 \sigma_k^T F'(x_{k+1}) \sigma_k > 0.$$

In conclusion, we have  $d_k^T B_{k+1} d_k > 0$  for  $\forall d_k \in \mathbb{R}^n$  and  $d_k \neq 0$ . □

*Remark b.* It is noticed that the proposed rank-two quasi-Newton update (6) shares the nice property of positive definiteness with BFGS update. When  $B_0 > 0$ , the matrices  $\{B_k\}$  updated by update formula (6) for solving uniformly convex unconstrained optimization problems also are symmetric and positive definite.

**Lemma 2.2.** *If formula (7) is written as  $H_{k+1} = H_k + E$ , where  $H_k$  is symmetric and satisfies  $\sigma_k^T = \sigma_k^T F'(x_{k+1}) H_k$ , then  $E$  solves the variational problem*

$$(9) \quad \min_E \|E\|_W \quad \text{s.t.} \quad E^T = E, \quad \sigma_k^T F'(x_{k+1}) E = \eta^T,$$

where  $\eta = \sigma_k^T - \sigma_k^T F'(x_{k+1}) H_k$  and  $W$  satisfies  $\sigma_k^T W = \sigma_k^T F'(x_{k+1})$ .

*Proof.* Since the problem is a convex programming problem, so we can solve its first order conditions. A suitable Lagrangian function is

$$\varphi = \frac{1}{4} \text{trace}(W E^T W E) + \text{trace}(\Lambda^T (E^T - E)) - \lambda^T W (E F'(x_{k+1}) \sigma_k - \eta),$$

where  $\Lambda$  and  $\lambda$  are the corresponding Lagrange multipliers for the two constraints. For derivatives respect to  $E$ ,  $\partial B / \partial B_{i,j} = e_i e_j^T$ , so in the case

$$\begin{aligned} \frac{\partial \varphi}{\partial E_{ij}} &= \frac{1}{4} (\text{trace}(W e_j e_i^T W E) + \text{trace}(W E^T W e_i e_j^T)) \\ &\quad + \text{trace}(\Lambda (e_j e_i^T - e_i e_j^T)) - \lambda^T W e_i e_j^T F'(x_{k+1}) \sigma_k = 0, \end{aligned}$$

or, using the symmetry and invariance of the trace to cyclic permutations, we have

$$\frac{1}{2} [W E W]_{ij} + \Lambda_{ij} - \Lambda_{ji} = [W \lambda \sigma_k^T F'(x_{k+1})]_{ij}.$$

Transposing and adding eliminates  $\Lambda$  to give

$$W E W = W \lambda \sigma_k^T F'(x_{k+1}) + F'(x_{k+1}) \sigma_k \lambda^T W,$$

and using  $\sigma_k^T W = \sigma_k^T F'(x_{k+1})$  and the nonlinearity of  $W$  it follows that

$$(10) \quad E = \lambda \sigma_k^T + \sigma_k \lambda^T.$$

Thus the result that the correction is of rank two is seen to arise naturally out of the analysis. Substituting (10) into  $\sigma_k^T F'(x_{k+1}) E = \eta^T$  and rearranging gives

$$\lambda = \frac{\eta - \sigma_k \lambda^T F'(x_{k+1}) \sigma_k}{\sigma_k^T F'(x_{k+1}) \sigma_k}.$$

Postmultiplying by  $\sigma_k^T F'(x_{k+1})$  gives

$$\lambda^T F'(x_{k+1})\sigma_k = \frac{1}{2} \frac{\sigma_k^T F'(x_{k+1})\eta}{\sigma_k^T F'(x_{k+1})\sigma_k},$$

so one has

$$\lambda = \frac{\eta - \frac{1}{2} \frac{\sigma_k^T \sigma_k F'(x_{k+1})\eta}{\sigma_k^T F'(x_{k+1})\sigma_k}}{\sigma_k^T F'(x_{k+1})\sigma_k} = \frac{H_k F'(x_{k+1})\sigma_k - \frac{1}{2} \frac{\sigma_k^T \sigma_k F'(x_{k+1})H_k F'(x_{k+1})\sigma_k}{\sigma_k^T F'(x_{k+1})\sigma_k}}{\sigma_k^T F'(x_{k+1})\sigma_k}.$$

Substituting this into (10) gives the correction formula defined by (7).  $\square$

**Lemma 2.3.** *If  $H_k = B_k^{-1}$  is positive definite and  $\sigma_k^T F'(x_{k+1})\sigma_k > 0$ , then the variational problem*

$$(11) \quad \min_{B>0} \psi(H_k^{1/2} B H_k^{1/2})$$

$$s.t. \quad B^T = B,$$

$$(12) \quad \sigma_k^T B = \sigma_k^T F'(x_{k+1}).$$

is solved by  $B_{k+1}$  given by (6).

*Proof.* By the definition of function  $\psi$  given by [2] and matrix product we have

$$(13) \quad \begin{aligned} \psi(H_k^{1/2} B H_k^{1/2}) &= \text{trace}(H_k B) - \ln(\det H_k \det B) \\ &= \psi(H_k B) = \psi(B H_k). \end{aligned}$$

A suitable Lagrangian function for the constrained optimization problem is

$$\begin{aligned} L(B, \Lambda, \lambda) &= \frac{1}{2} \psi(H_k^{1/2} B H_k^{1/2}) + \text{trace}(\Lambda^T (B^T - B)) \\ &\quad + (\sigma_k^T B - \sigma_k^T F'(x_{k+1}))\lambda_k \\ &= \frac{1}{2} \left( \psi(H_k B) - \ln(\det H_k) - \ln(\det B) \right) + \text{trace}(\Lambda^T (B^T - B)) \\ &\quad + (\sigma_k^T B - \sigma_k^T F'(x_{k+1}))\lambda_k, \end{aligned}$$

where  $\Lambda$  and  $\lambda$  are the corresponding Lagrange multipliers for (11) and (12).

Using the identity  $\partial B / \partial B_{i,j} = e_i e_j^T$  and Lemma 1.4 in [2], it follows that

$$(14) \quad \begin{aligned} \frac{\partial L}{\partial B_{ij}} &= \frac{1}{2} \text{trace} \left( H_k e_i e_j^T - (B^{-1})_{ji} \right) + \text{trace} \left( \Lambda^T (e_k e_i^T - e_i e_j^T) \right) + \sigma_k^T e_i e_j^T \lambda \\ &= \frac{1}{2} (H_k)_{ji} - (B^{-1})_{ji} + \Lambda_{ji} - \Lambda_{ij} + (\sigma_k^T \lambda)_{ij} = 0. \end{aligned}$$

It can be derived by transposing and adding from (14) that

$$(15) \quad \begin{aligned} H_k - B^{-1} + \sigma_k^T \lambda + \lambda^T \sigma_k &= 0, \\ B^{-1} = H_k + \sigma_k^T \lambda + \lambda^T \sigma_k. \end{aligned}$$

It then follows, using the equation  $\sigma_k^T = \sigma_k^T F'(x_{k+1})B^{-1}$  derived from (12), that

$$\sigma_k^T = \sigma_k^T F'(x_{k+1})H_k + \sigma_k^T F'(x_{k+1})\lambda\sigma_k^T + \sigma_k^T F'(x_{k+1})\sigma_k\lambda^T,$$

and hence

$$\begin{aligned} \sigma_k^T F'(x_{k+1})\sigma_k &= \sigma_k^T F'(x_{k+1})H_k F'(x_{k+1})\sigma_k + \sigma_k^T F'(x_{k+1})\lambda\sigma_k^T F'(x_{k+1})\sigma_k \\ &\quad + \sigma_k^T F'(x_{k+1})\sigma_k\lambda^T F'(x_{k+1})\sigma_k. \end{aligned}$$

Rearranging this gives

$$\sigma_k^T F'(x_{k+1})\sigma_k = \frac{1}{2} \left( 1 - \frac{\sigma_k^T F'(x_{k+1})H_k F'(x_{k+1})\sigma_k}{\sigma_k^T F'(x_{k+1})\sigma_k} \right),$$

and so

$$\lambda = \frac{\sigma_k - H_k F'(x_{k+1}) - \frac{1}{2} \left( 1 - \frac{\sigma_k^T F'(x_{k+1})H_k F'(x_{k+1})\sigma_k}{\sigma_k^T F'(x_{k+1})\sigma_k} \right)}{\sigma_k^T F'(x_{k+1})\sigma_k}.$$

Substituting this expression into (6) gives the formula (7).

Using two times of Sherman-Morrison formula, we can get the update (6). Finally, the function  $\psi(H_k^{1/2}BH_k^{1/2})$  is seen to be a strictly convex function on  $B > 0$  by virtue of (13) and Lemma 1.2 in [5], so it follows that the rank two update formula gives the unique solution of the variational problem.  $\square$

### 3. The rank two quasi-Newton method

In this section, we state our quasi-Newton method as follows.

**Algorithm 3. A rank two quasi-Newton method (RTQN)**

Step 0. Choose an initial point  $x_0 \in \mathbb{R}^n$ , an symmetric positive definite matrix  $B_0 \in \mathbb{R}^{n \times n}$ , and constants  $r, \rho \in (0, 1)$ ,  $0 < \sigma_1, \sigma_2 \leq 1$ ,  $k := 0$ .

Step 1. Stop if  $\|F(x_k)\| = 0$ . Otherwise solve the subproblem

$$(16) \quad B_k d + F(x_k) = 0$$

to get  $d_k$ .

Step 2. If

$$(17) \quad \|F(x_k + d_k)\| \leq \rho \|F(x_k)\|,$$

then  $\alpha_k = 1$  and go to Step 4, otherwise go to Step 3.

Step 3. Let  $i_k$  be the smallest nonnegative integer  $i$  such that

$$(18) \quad \|F(x_k + \alpha d_k)\|^2 - \|F(x_k)\|^2 \leq -\sigma_1 \|\alpha F(x_k)\|^2 - \sigma_2 \|\alpha d_k\|^2$$

holds for  $\alpha = r^i$ , let  $\alpha_k = r^{i_k}$ .

Step 4. Get the next iterative  $x_{k+1} = x_k + \alpha_k d_k$ .

Step 5. If  $\sigma_k^T F'(x_{k+1})\sigma_k > 0$ , update  $B_k$  by formula (6), where  $\sigma_k = s_k$ . Otherwise let  $B_{k+1} = B_k$ . Let  $k := k + 1$ , go to Step 1.



*Remark c.* According to the adjoint Broyden tangent condition and the choice of  $\sigma_k$ , we have

$$B_{k+1}s_k = F'(x_{k+1})s_k,$$

which means that  $B_{k+1}$  approximates  $F'(x_{k+1})$  along the direction  $s_k$  exactly.

For the sake of convenience, we make some assumptions as follows.

**Assumption A**

- (i)  $F(x)$  is continuously differentiable on an open convex set  $\Omega_1 \subseteq \Omega$ , where  $\Omega = \{x \mid \|F(x)\| \leq \|F(x_0)\|\}$ .
- (ii) The Jacobian  $F'(x)$  of  $F$  is symmetric and bounded on  $\Omega_1$ , i.e., there exists a positive constant  $M$  such that

$$\|F'(x)\| \leq M, \forall x \in \Omega_1.$$

- (iii)  $F'(x)$  is positive definite on  $\Omega_1$ , i.e., there is a constant  $m > 0$  such that

$$m\|d\|^2 \leq d^T F'(x)d, \forall x \in \Omega_1, d \in \mathbb{R}^n.$$

*Remark d.* 1. Conditions (ii) and (iii) in Assumption A imply that there exist constants  $M \geq m > 0$  such that

$$(19) \quad m\|d\| \leq \|F'(x)d\| \leq M\|d\|, \forall x \in \Omega_1, d \in \mathbb{R}^n,$$

$$(20) \quad \frac{1}{M}\|d\| \leq \|F'(x)^{-1}d\| \leq \frac{1}{m}\|d\|, \forall x \in \Omega_1, d \in \mathbb{R}^n,$$

and

$$(21) \quad m\|x - y\| \leq \|F(x) - F(y)\| \leq M\|x - y\|, \forall x, y \in \Omega_1.$$

In particular, for all  $x \in \Omega_1$ , one has

$$(22) \quad m\|x - x^*\| \leq \|F(x) - F(x^*)\| \leq M\|x - x^*\|,$$

where  $x^*$  is the unique solution of (1) in  $\Omega_1$ .

2. Since  $F'(x)$  is symmetric and positive definite, there exists a positive and bounded matrix  $Q$  such that  $F'(x) = Q^T Q$ . Combining this with Assumption A(ii) and (19), we have

$$(23) \quad \|Q(x)d\| \geq \frac{m}{M_1}\|d\|, \forall x \in \Omega_1, d \in \mathbb{R}^n,$$

where  $M_1$  is the bound of  $Q$ .

3. We define  $p_k = F'(x_{k+1})s_k$ , then by Assumption A, it is easy to get

$$(24) \quad \|p_k\| = \|F'(x_{k+1})s_k\| \leq M\|s_k\|$$

and

$$(25) \quad p_k^T s_k = s_k^T F'(x_{k+1})s_k \geq m\|s_k\|^2.$$

Similar to [16], we also need the following assumption.

**Assumption B**

$B_k$  is a good approximation to  $F'(x_k)$ , i.e.,

$$\|(F'(x_k) - B_k)d_k\| \leq \varepsilon \|F(x_k)\|,$$

where  $\varepsilon \in (0, 1)$ .

We will prove some useful lemmas related to Algorithm 3.

**Lemma 3.1.** *Let Assumption B hold. Then  $d_k$  is a descent direction for  $\theta(x) = \frac{1}{2}\|F(x_k)\|^2$  at  $x_k$ , i.e.,*

$$\nabla\theta(x_k)^T d_k < 0.$$

*Proof.* By (16), we get

$$\begin{aligned} \nabla\theta(x_k)^T d_k &= F(x_k)^T F'(x_k) d_k \\ &= F(x_k)^T ((F'(x_k) - B_k)d_k - F(x_k)) \\ &= F(x_k)^T (F'(x_k) - B_k)d_k - F(x_k)^T F(x_k), \end{aligned}$$

then we have

$$\nabla\theta(x_k)^T d_k + \|F(x_k)\|^2 \leq F(x_k)^T (F'(x_k) - B_k)d_k.$$

Taking the norm on the right-hand side, we can obtain

$$\nabla\theta(x_k)^T d_k \leq \|F(x_k)\| \|(F'(x_k) - B_k)d_k\| - \|F(x_k)\|^2 \leq -(1 - \varepsilon) \|F(x_k)\|^2.$$

We complete the proof.  $\square$

Based on the norm descent of  $F(x_k)$ , we can get the following lemma easily.

**Lemma 3.2.** *Let  $\{x_k\}$  be generated by Algorithm 3. Then  $\{x_k\} \subset \Omega$ . Moreover,  $\{\|F(x_k)\|\}$  converges.*

**Lemma 3.3.** *Let Assumption A hold. Then for any  $r_0 \in (0, 1)$  and  $k \geq 0$ , there are positive constants  $\beta_j$ ,  $j = 1, 2, 3$ , the following inequalities*

$$(26) \quad \beta_2 \|s_i\|^2 \leq s_i^T B_i s_i \leq \beta_3 \|s_i\|^2 \quad \text{and} \quad \|B_i s_i\| \leq \beta_1 \|s_i\|,$$

*hold for at least  $[r_0 k]$  values of  $i \in [0, k]$ .*

The following lemma shows that the Algorithm 3 is well-defined.

**Lemma 3.4.** *Let Assumptions A and B hold and index set  $\bar{K}$  defined by  $\bar{K} = \{k | (26) \text{ holds}\}$ . Then Algorithm 3 will produce the next iteration in a finite number of backtracking steps.*

*Proof.* According to Lemma 3.8 of [1], after finite number of backtracking steps, there must be an  $\alpha_k$  satisfying

$$\|F(x_k + \alpha_k d_k)\|^2 - \|F(x_k)\|^2 \leq \delta \alpha_k F(x_k)^T F'(x_k)^T d_k, \quad \delta \in (0, 1).$$

By the subproblem (16) and  $\alpha_k \leq 1$ , we have

$$\alpha_k g(x_k)^T F'(x_k)^T d_k \leq -\alpha_k (1 - \varepsilon) \|F(x_k)\|^2$$

$$\begin{aligned}
 &= -\frac{(1-\varepsilon)}{2\alpha_k} \|\alpha_k F(x_k)\|^2 - \frac{(1-\varepsilon)}{2\alpha_k} \|\alpha_k F(x_k)\|^2 \\
 &\leq -\frac{(1-\varepsilon)}{2} \|\alpha_k F(x_k)\|^2 - \frac{\beta_2^2(1-\varepsilon)}{2} \|\alpha_k d_k\|^2.
 \end{aligned}$$

Let  $\sigma_1 \in (0, \frac{\delta(1-\varepsilon)}{2})$ ,  $\sigma_2 \in (0, \frac{\delta\beta_2^2(1-\varepsilon)}{0})$ , we can obtain (18). □

#### 4. Global and superlinear convergence

We will show some important lemmas which are important for the global convergence of Algorithm 3. Similar to Corollary 3.4 in [10], it is not difficult to deduce the following results.

**Lemma 4.1.** *Let Assumption A hold. If  $\alpha_k \neq 1$ , then we have*

$$(27) \quad \alpha_k \geq \frac{2M\beta_2 r}{\sigma_1\beta_1^2 + \sigma_2 + M^2}.$$

**Lemma 4.2.** *Let Assumption A hold and (17) holds only for a finite number of  $k$ . Then we have*

$$\sum_{k=0}^{\infty} \|\alpha_k F(x_k)\|^2 < \infty$$

and

$$\sum_{k=0}^{\infty} \|\alpha_k d_k\|^2 = \sum_{k=0}^{\infty} \|s_k\|^2 < \infty.$$

*Proof.* By the conditions of the theorem, there is an index  $\hat{k}$  such that the step length  $\alpha_k$  is determined by (17) for all  $k \geq \hat{k}$ . In other words, the following inequality holds for all  $k \geq \hat{k}$ .

$$(28) \quad \sigma_1 \|\alpha_k F(x_k)\|^2 + \sigma_2 \|s_k\|^2 \leq \|F(x_k)\|^2 - \|F(x_{k+1})\|^2.$$

Since  $\{\|F(x_k)\|\}$  is bounded, we get the results by adding these inequalities. □

Now we give the global convergence of Algorithm 3.

**Theorem 4.3.** *Let Assumption A hold. Then the sequence of  $\{x_k\}$  generated by Algorithm 3 converges to the unique solution  $x^*$  of (1).*

*Proof.* According to Lemma 3.2,  $\{\|F(x_k)\|\}$  converges. Combined with the positive definiteness of  $F'(x)$  and the boundness of  $\Omega$ , it is sufficient to verify that

$$(29) \quad \lim_{k \rightarrow \infty} \|F(x_k)\| = 0.$$

If (17) holds for infinitely many  $k$ , then (29) is trivial. Else, we have from Lemma 4.2 that

$$\sum_{k=0}^{\infty} \|\alpha_k F(x_k)\|^2 < \infty.$$

Combined with (27), we get (29). □

In order to give the superlinear convergence of Algorithm 3, we need the following Assumption.

**Assumption C**

$F'(x)$  is Hölder continuous at  $x^*$ , i.e., there are positive constants  $M_2$  and  $\nu$  such that for every  $x$  in a neighbourhood of  $x^*$

$$(30) \quad \|F'(x) - F'(x^*)\| \leq M_2 \|x - x^*\|^\nu.$$

According to Lemma 3.5 of [10], we only need to verify the Dennis-Moré condition

$$(31) \quad \lim_{k \rightarrow \infty} \frac{\|(B_k - F'(x^*))d_k\|}{\|d_k\|} = 0.$$

In what follows, for the sake of convenience, we denote  $\varphi_k(\nu)$  for

$$\varphi_k(\nu) = \max\{\|x_k - x^*\|^\nu, \|x_{k+1} - x^*\|^\nu\}.$$

**Lemma 4.4.** *Let Assumptions A and B hold. Then for any fixed  $\nu > 0$ , we have*

$$(32) \quad \sum_{k=0}^{\infty} \|x_k - x^*\|^\nu < \infty.$$

Moreover, we have

$$(33) \quad \sum_{k=0}^{\infty} \varphi_k(\nu) < \infty.$$

*Proof.* Firstly, we will prove that there is a constant  $\delta \in (0, 1)$  and an index  $i'$  such that

$$(34) \quad \|F(x_{i+1})\|^2 \leq \delta \|F(x_i)\|^2, \quad \forall i \geq i'.$$

If the step length  $\alpha_k$  is determined by Step 2, then

$$(35) \quad \|F(x_{i+1})\|^2 \leq \lambda \|F(x_i)\|^2.$$

On the other hand, if  $\alpha_i$  is determined by Step 3, then it satisfies (18) with  $k = i$  and hence

$$(36) \quad \|F(x_{i+1})\|^2 \leq (1 - \sigma_1 \alpha_i^2) a \|F(x_i)\|^2$$

$$(37) \quad \leq (1 - \sigma_1 \alpha'^2) a \|F(x_i)\|^2,$$

where the last inequality comes from (27) and the  $\alpha' > 0$  is some constant satisfying  $\alpha_i \geq \alpha'$ .

Then there exists a constant  $\delta_1 \in (0, 1)$  such that  $1 - \sigma_1 \alpha'^2 \leq \delta_1$  holds for all  $i \geq i'$ . Let  $\delta = \min\{\lambda^2, \delta_1\}$ . Then we get (34) from (35) and (37).

Now we prove (32). Let  $K$  denote the set of indices  $i$  for which (34) holds. Also, let  $l_k$  denote the number of indices in  $K$  not exceeding  $k$ . Then we have

$l_k \geq k - i' \geq 0$  for each  $k$ . Multiplying (34) for  $i \in K$  and (36) for  $i \notin K$  from  $i = i'$  to  $i = k$ , we can get

$$\begin{aligned} \|F(x_{k+1})\|^2 &\leq \left( \prod_{i=i', i \notin K}^k (1 - \sigma_1 \alpha_i^2) \right) \delta_k^{l_k} \|F(x_{i'})\|^2 \\ &\leq \left( \prod_{i=0}^k 1 \right) \delta^{k-i'} \|F(x_{i'})\|^2 \\ &\leq \delta^{k-(i'+1)} \|F(x_{i'})\|^2 \\ &= c_1 \delta^k, \end{aligned}$$

where  $c_1 = \delta^{-(i'+1)} \|F(x_{i'})\|^2$ . This together with (22) shows that  $\|x_{k+1} - x^*\|^2 \leq m^{-2} c_1 \delta_k$  holds for all  $k$  large enough. Then we have (32) for any  $\nu$ . Since  $v_k(\nu) \leq \|x_{k+1} - x^*\|^\nu + \|x_k - x^*\|^\nu$ , then (33) is trivial.  $\square$

**Lemma 4.5.** *Let Assumptions A, B and C hold. Then the following inequality*

$$(38) \quad \|F'(x_{k+1})s - F'(x^*)s_k\| \leq M_2 \varphi_k(\nu) \|s_k\|$$

*holds for all  $k$  sufficiently large.*

*Proof.* Since  $x_k \rightarrow x^*$ , (30) holds for all  $k$  large enough. When  $k$  is large enough

$$\begin{aligned} \|F'(x_{k+1})s_k - F'(x^*)s_k\| &\leq \|F'(x_{k+1}) - F'(x^*)\| \cdot \|s_k\| \\ &\leq M_2 \|x_{k+1} - x^*\|^\nu \cdot \|s_k\| \\ &\leq M_2 \varphi_k(\nu) \|s_k\|. \end{aligned} \quad \square$$

Denote  $P = F'(x^*)^{-1/2}$ . For an  $n \times n$  matrix  $A$ , define a matrix norm  $\|A\|_P = \|PAP\|_F$ , where  $\|\cdot\|_F$  denotes the Frobenius norm of a matrix. Then we show a property of the proposed method similar to BFGS method, which can be proved similarly to Lemma 3.8 [10].

**Lemma 4.6.** *Let Assumptions A, B and C hold. Then there exist positive constants  $M_3, M_4, M_5$  and  $w \in (0, 1)$  such that for all  $k$  sufficiently large*

$$(39) \quad \|B_{k+1} - F'(x^*)^{-1}\|_P \leq \|B_k - F'(x^*)\|_P + M_3 \varphi_k(\nu),$$

$$(40) \quad \begin{aligned} &\|H_{k+1} - F'(x^*)^{-1}\|_{P^{-1}} \\ &\leq \left(1 - \frac{1}{2} w \mu_k^2 + M_4 \varphi_k(\nu)\right) \|H_k - F'(x^*)^{-1}\|_{P^{-1}} + M_5 \varphi_k(\nu), \end{aligned}$$

where  $\mu_k$  is given by

$$\mu_k = \frac{\|P^{-1}[H_k - F'(x^*)^{-1}](F'(x_{k+1})s_k)\|}{\|H_k - F'(x^*)^{-1}\|_{P^{-1}} \|P(F'(x_{k+1})s_k)\|}.$$

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

Now, we prove the superlinear convergence of Algorithm 3.

**Lemma 4.7.** *Let Assumptions A, B and C hold. Then*

$$(41) \quad \lim_{k \rightarrow \infty} \frac{\|B_k - F'(x^*)s_k\|}{\|s_k\|} = 0.$$

Moreover,  $\{x_k\}$  converges superlinearly and  $\alpha_k \equiv 1$  for all  $k$  sufficiently large.

*Proof.* (40) can be written as

$$\begin{aligned} & \frac{1}{2}w\mu_k^2\|H_k - F'(x^*)^{-1}\|_{P^{-1}} \\ & \leq \|H_k - F'(x^*)^{-1}\|_{P^{-1}} - \|H_{k+1} - F'(x^*)^{-1}\|_{P^{-1}} \\ & \quad + M_5\|H_k - F'(x^*)^{-1}\|_{P^{-1}}\varphi_k + M_6\varphi_k(\nu). \end{aligned}$$

Notice that  $\|H_k - F'(x^*)^{-1}\|_{P^{-1}}$  is bounded and  $\varphi_k(\nu)$  satisfies (33). Summing the above inequalities, we get

$$\frac{1}{2}w \sum_{k=0}^{\infty} \mu_k^2 \|H_k - F'(x^*)^{-1}\|_{P^{-1}} < \infty.$$

By the definition of  $\mu_k$ , we have

$$\begin{aligned} \lim_{k \rightarrow \infty} \mu_k^2 \|H_k - F'(x^*)^{-1}\|_{P^{-1}} &= \lim_{k \rightarrow \infty} \frac{\|P^{-1}[H_k - F'(x^*)^{-1}](F'(x_{k+1})s_k)\|^2}{\|H_k - F'(x^*)^{-1}\|_{P^{-1}} \|P(F'(x_{k+1})s_k)\|^2} \\ &= 0. \end{aligned}$$

Since  $\|H_k - F'(x^*)^{-1}\|_{P^{-1}}$  is bounded, we have

$$(42) \quad \lim_{k \rightarrow \infty} \frac{\|P^{-1}(H_k - F'(x^*)^{-1})(F'(x_{k+1})s_k)\|}{\|P(F'(x_{k+1})s_k)\|} = 0.$$

According to (24), we can get

$$\|P(F'(x_{k+1})s_k)\| \leq \|P\| \cdot \|(F'(x_{k+1})s_k)\| \leq M\|P\| \cdot \|s_k\|.$$

By (23), we have

$$\begin{aligned} \|P^{-1}[H_k - F'(x^*)^{-1}](F'(x_{k+1})s_k)\| &= \|\nabla^{1/2}(H_k - F'(x^*)^{-1})(F'(x_{k+1})s_k)\| \\ &\geq \frac{m}{M_1} \|(H_k - F'(x^*)^{-1})(F'(x_{k+1})s_k)\|. \end{aligned}$$

Therefore, (42) implies

$$(43) \quad \lim_{k \rightarrow \infty} \frac{\|(H_k - F'(x^*)^{-1})(F'(x_{k+1})s_k)\|}{\|s_k\|} = 0.$$

On the other hand, we have

$$\begin{aligned} & \|(H_k - F'(x^*)^{-1})(F'(x_{k+1})s_k)\| \\ &= \|H_k(F'(x^*) - B_k)F'(x^*)^{-1}(F'(x_{k+1})s_k)\| \\ &\geq \|H_k(F'(x^*) - B_k)s_k\| - \|H_k(F'(x^*) - B_k)(s_k - F'(x^*)^{-1}(F'(x_{k+1})s_k))\| \\ &= \|H_k(F'(x^*) - B_k)s_k\| \end{aligned}$$

$$\begin{aligned}
 & - \|H_k(F'(x^*) - B_k)F'(x^*)^{-1}((F'(x_{k+1})s_k) - F'(x^*)s_k)\| \\
 & \geq \|H_k(F'(x^*) - B_k)s_k\| - M_3v_k\|H_k(F'(x^* - B_k)F'(x^*)^{-1})\| \cdot \|s_k\| \\
 & = \|H_k(F'(x^*) - B_k)s_k\| - o(\|s_k\|).
 \end{aligned}$$

Notice that  $\{\|B_k\|\}$  and  $\{\|H_k\|\}$  are bounded and  $\{H_k\}$  is uniformly nonsingular. Therefore, there exist a constant  $\bar{m} > 0$  such that

$$\|H_k(F'(x^*) - B_k)s_k\| \geq \bar{m}\|(F'(x^*) - B_k)s_k\|$$

for all  $k$ . So we have

$$\|(H_k - F'(x^*)^{-1})y_k\| \geq \bar{m}\|(F'(x^*) - B_k)s_k\| - o(\|s_k\|),$$

and hence (43) yields (41). According to Lemma 4.7, we complete the proof.  $\square$

## 5. Numerical experiments

In this section, we will compare the proposed rank two quasi-Newton method (RTQN) with BFGS method using the same line search, where BFGS method is an effective quasi-Newton method for solving small and medium-scale problems. We test the two methods on two classical test problems with different initial points listed in [10, 16].

**Problem 1.** The discretized two-point boundary-value problem like the problem in [12]

$$F(x) = Ax + \frac{1}{(n+1)^2}g(x) = 0,$$

where  $A$  is a  $n \times n$  tridiagonal matrix given by

$$A = \begin{bmatrix} 8 & -1 & & & & \\ -1 & 8 & -1 & & & \\ & -1 & 8 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & \ddots & -1 \\ & & & & -1 & 8 \end{bmatrix},$$

where  $g(x) = (g_1(x), g_2(x), \dots, g_n(x))^T$  with  $g_i(x) = \sin x_i - 1$ ,  $i = 1, 2, \dots, n$ .

**Problem 2.** Unconstrained optimization problem

$$\min f(x), \quad x \in \mathbb{R}^n,$$

with Engval function [15]  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  defined by

$$f(x) = \sum_{i=2}^n \left( (x_{i-1}^2 + x_i^2)^2 - 4x_{i-1} + 3 \right).$$

The related symmetric nonlinear equations is

$$F(x) = \frac{1}{4}\nabla f(x) = 0,$$

where  $F(x) = (F_1(x), F_2(x), \dots, F_n(x))^T$  with

$$\begin{aligned}
 f_1(x) &= x_1(x_1^2 + x_2^2) - 1, \\
 f_i(x) &= x_i(x_{i-1}^2 + 2x_i^2 + x_{i+1}^2) - 1, \quad i = 2, 3, \dots, n-1, \\
 f_n(x) &= x_n(x_{n-1}^2 + x_n^2).
 \end{aligned}$$

In the numerical experiments, we used the following condition

$$\|F(x_k)\| \leq 10^{-5} \text{ or } k \geq 1000$$

as the termination criterion. The latter condition implies that the method fails to find a solution of the problem. The parameters in the nonmonotone line search are chosen as

$$r = 0.1, \rho = 0.95, \sigma_1 = \sigma_2 = 10^{-5}.$$

The numerical experiments are done by using MATLAB R2012b on a Core (TM) 2 PC with Windows XP. The computation of the rank two quasi-Newton update formula (6) is based on the terms  $F'(x)s$  and  $\sigma^T F'(x)$ , which can be obtained by the forward and reverse modes of automatic differentiation (AD) without the need of the Jacobian.

For the given problems, we compare the two methods on the numbers of iterations “NI” and the number of function evaluations “NF” with different initial points and sizes, which are listed in Tables 1 and 2. We can see from the Tables 1 and 2 that the proposed method converges significantly faster than BFGS method. However, the computation of proposed method is lightly expensive than that of BFGS method.

TABLE 1. Results of Problem 1

$x_0$		(1, ..., 1)	(50, ..., 50)	(100, ..., 100)	(-1, ..., -1)	(-50, ..., -50)	(-100, ..., -100)
Dim	method	NI/NF	NI/NF	NI/NF	NI/NF	NI/NF	NI/NF
$n = 50$	BFGS	49/89	50/90	50/90	49/89	50/90	50/90
	RTQN	49/141	50/143	50/143	49/141	50/143	50/143
$n = 100$	BFGS	81/149	89/161	89/161	81/149	89/161	89/161
	RTQN	76/226	89/262	89/262	76/226	89/262	89/262
$n = 200$	BFGS	82/151	105/194	109/202	82/151	105/194	109/202
	RTQN	65/193	86/256	89/265	65/193	86/256	89/265
$n = 500$	BFGS	82/149	104/190	108/198	82/149	104/190	108/198
	RTQN	65/192	87/256	90/265	65/192	87/256	90/265
$n = 1000$	BFGS	76/141	99/184	104/193	76/141	99/184	104/193
	RTQN	67/196	88/259	91/268	67/196	88/259	91/268
$x_0$		(1, 0, 1, 0, ...)	(50, 0, 50, 0, ...)	(100, 0, 100, 0, ...)	(-1, 0, -1, 0, ...)	(-50, 0, -50, 0, ...)	(-100, 0, -100, 0, ...)
$n = 50$	BFGS	69/131	85/155	87/158	69/131	85/155	87/158
	RTQN	68/202	85/249	87/254	68/202	85/249	87/254
$n = 100$	BFGS	73/134	96/177	100/185	73/134	96/177	100/185
	RTQN	66/196	88/259	92/271	66/196	88/259	92/271
$n = 200$	BFGS	71/132	94/174	98/181	71/132	94/174	98/181
	RTQN	65/193	86/256	89/265	65/193	86/256	89/265
$n = 500$	BFGS	70/129	94/173	98/181	70/129	94/173	98/181
	RTQN	65/192	87/256	90/265	65/192	87/256	90/265
$n = 1000$	BFGS	72/132	96/176	100/183	72/132	96/176	100/183
	RTQN	67/196	88/259	91/268	67/196	88/259	91/268

In order to analyse the efficiency of proposed method more precisely, we adopt a notion of performance profile [4], which is a distribution function for a performance metric to evaluate and compare the performance of the set of solvers  $S$  on a test set  $P$ . Suppose that there exist  $N_s$  solvers and  $N_p$  problems, for each problem  $p$  and solver  $s$ , Dolan and Moré [4] defined  $t_{p,s}$  the number



TABLE 2. Results of Problem 2

$x_0$	(1, ..., 1)	(0.5, ..., 0.5)	(1, 0, 1, 0, ...)	(0.5, 0, 0.5, 0, ...)	$x_0$	(1, ..., 1)	(0.5, ..., 0.5)	(1, 0, 1, 0, ...)	(0.5, 0, 0.5, 0, ...)
Dim	method	NI/NF	NI/NF	NI/NF	Dim	method	NI/NF	NI/NF	NI/NF
$n = 9$	BFGS	17/22	18/23	18/24	$n = 200$	BFGS	43/65	40/60	39/58
	RTQN	17/392	18/39	20/45		RTQN	40/101	37/97	42/103
$n = 50$	BFGS	43/62	41/58	38/55	$n = 500$	RTQN	43/65	42/65	43/64
	RTQN	38/96	38/96	34/86		RTQN	40/101	43/106	38/96
$n = 99$	BFGS	43/65	44/59	44/66	$n = 1000$	BFGS	47/72	46/71	44/65
	RTQN	40/101	39/99	38/97		RTQN	42/106	48/122	38/96

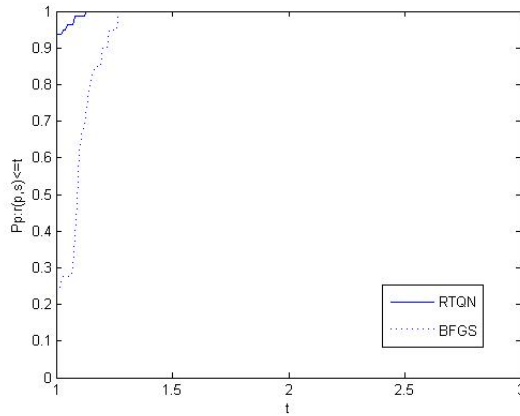


FIGURE 1. Performance profiles based on the number of iterations

of iterations (the number of function evaluations or others) required to solve problem  $p$  by solver  $s$ . Requiring a baseline for comparisons, they compared the performance on problem  $p$  by solver  $s$  with the best performance by any solver on this problem; i.e., using the performance ratio

$$r_{p,s} = \frac{tp,s}{\min\{t_{p,s} : s \in S\}}.$$

Suppose that a parameter  $r_M \geq r_{p,s}$  for all  $p, s$  is chosen, and  $r_{p,s} = r_M$  if and only if solver  $s$  does not solve problem  $p$ . In order to obtain an overall assessment of the performance of the solver, they defined

$$\rho_s(t) = \frac{1}{N_p} \text{size}\{p \in P : r_{p,s} \leq t\},$$

thus  $\rho_s : \mathbb{R} \rightarrow [0, 1]$  was the probability for solver  $s \in S$  that a performance ratio  $r_{p,s}$  was within a factor  $t \in R$  of the best possible ratio.

Figure 1 evaluated the performance of the two methods relative to the number of iteration. Clearly, the top curve corresponds to RTQN, which means that the proposed rank two quasi-Newton method performed better in terms of iteration counts.

## 6. Final remarks

We have presented a new rank two quasi-Newton method based on adjoint Broyden updates for solving symmetric nonlinear equations. The required Jacobian-vector products can be obtained efficiently using automatic differentiation. The new quasi-Newton method possesses some favorable properties, which are shared by BFGS method. Under suitable conditions, we have established the global and superlinear convergence. The numerical results showed that the method is practically effective.

**Acknowledgement.** The author would like to thank the anonymous referees for their suggestions which improved the manuscript greatly.

## References

- [1] P. N. Brown and Y. Saad, *Convergence theory of nonlinear Newton-Krylov algorithms*, SIAM J. Optim. **4** (1994), no. 2, 297–330.
- [2] R. H. Byrd and J. Nocedal, *A tool for the analysis of quasi-Newton methods with application to unconstrained minimization*, SIAM J. Numer. Anal. **26** (1989), no. 3, 727–739.
- [3] Y. H. Dai, *Convergence properties of the BFGS algorithm*, SIAM J. Optim. **13** (2002), no. 3, 693–701.
- [4] E. D. Dolan and J. J. Moré, *Benchmarking optimization software with performance profiles*, Math. Program. **91** (2002), no. 2, 201–213.
- [5] R. Fletcher, *A new variational result for quasi-Newton formulae*, SIAM J. Optim. **1** (1991), no. 1, 18–21.
- [6] ———, *Practical Methods of Optimization*, 2nd ed., John Wiley & Sons, 2013.
- [7] A. Griewank, *The "global" convergence of Broyden-like methods with a suitable line search*, J. Aust. Math. Soc. Ser. B **28** (1986), no. 1, 75–92.
- [8] G. Z. Gu, D. H. Li, L. Qi, et al., *Descent directions of quasi-Newton methods for symmetric nonlinear equations*, SIAM J. Numer. Anal. **40** (2002), no. 5, 1763–1774.
- [9] D. H. Li and W. Y. Cheng, *Recent progress in the global convergence of quasi-Newton methods for nonlinear equations*, Hokkaido Math. J. **36** (2007), no. 4, 729–743.
- [10] D. H. Li and M. Fukushima, *A globally and superlinearly convergent Gauss-Newton-based BFGS method for symmetric nonlinear equations*, SIAM J. Numer. Anal. **37** (1999), no. 1, 152–172.
- [11] D. H. Li and X. L. Wang, *A modified Fletcher-Reeves-type derivative-free method for symmetric nonlinear equations*, Numer. Algebra Control Optim. **1** (2011), no. 1, 71–82.
- [12] J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, SIAM, Academic Press, 1970.
- [13] S. Schlenkrich, A. Griewank, and A. Walther, *On the local convergence of adjoint Broyden methods*, Math. Program. **121** (2010), no. 2, 221–247.
- [14] S. Schlenkrich and A. Walther, *Global convergence of quasi-Newton methods based on adjoint Broyden updates*, Appl. Numer. Math. **59** (2009), no. 5, 1120–1136.
- [15] E. Yamakawa and M. Fukushima, *Testing parallel variable transformation*, Comput. Optim. Appl. **13** (1999), no. 1-3, 253–274.
- [16] G. L. Yuan and S. W. Yao, *A BFGS algorithm for solving symmetric nonlinear equations*, Optimization. **62** (2013), no. 1, 85–99.
- [17] L. Zhang, W. J. Zhou, and D. H. Li, *Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search*, Numer. Math. **104** (2006), no. 4, 561–572.
- [18] W. J. Zhou and D. M. Shen, *An inexact PRP conjugate gradient method for symmetric nonlinear equations*, Numer. Funct. Anal. Optim. **35** (2014), no. 3, 370–388.

HUIPING CAO  
COLLEGE OF MATHEMATICS AND ECONOMETRICS  
HUNAN UNIVERSITY  
CHANGSHA 410082, P. R. CHINA  
*E-mail address:* [huiping\\_cao@hnu.edu.cn](mailto:huiping_cao@hnu.edu.cn)