ACADIMIC World Academic Union

ISSN 1749-3889(print),1749-3897(online)
International Journal of Nonlinear Science
Vol.2(2006) No.3,pp.140-146

A Class of Generalized Quasi-Newton Algorithms with Superlinear Convergence

Honglan Zhu, Shu Wen [1]
Huaiyin Institute of Technology, Huaian, 223003, China
(Received 26 July 2006, accepted 7 September 2006)

Abstract: In this paper, we present a class of new generalized quasi-Newton algorithms for unconstrained optimization. The new algorithms are very extensive, including the algorithms in Jiao's paper and also in Zhangs' even the class of Broyden. The global convergence and the superlinear convergence of the new algorithms are also proved under the weak condition. Numerical experiment indicates that the new algorithms are more feasible and effective.

Keywords: unconstrained optimization; generalized quasi-Newton algorithm; global convergence; superlinear convergence

1 Introduction

Optimization is a very active branch in the computational Mathematics . It is already extensive to apply in many engineering sections. Readers can refer to[$1 \sim 13$] and many other papers.

For unconstrained optimization problem[1]

$$\min f(x), x \in R^n, \tag{1.1}$$

where $f(x)$ is twice continuously differentiable. The standard quasi-Newton equation is

$$B_{k+1}\delta_k = y_k,, \tag{1.2}$$

here $\delta_k = x_{k+1} - x_k$, $g_k = \nabla f(x_k)$, $y_k = g_{k+1} - g_k$ and B_{k+1} is the approximation of Hessian matrix $G(x)$ at x_{k+1} .

Obviously, only two gradients are exploited in (1.2). Techniques using gradients as well as function values (which are also available) have been studied by several authors. Recently, Zhang[2] made use of the quadratic function $q(\tau) = a\tau^2 + b\tau + c$, where $a, b, c \in R^n$ to approximate the gradient and established a new quasi-Newton equation

$$G_{k+1}\delta_k = y_k + \gamma_k\delta_k/\|\delta_k\|^2 \tag{1.3}$$

where $G_{k+1} = \nabla^2 f(x_{k+1}), \gamma_k = 3g_{k+1}^T\delta_k + 3g_k^T\delta_k + 6(f_k - f_{k+1})$. He, furthermore, gave the equivalent form of (1.3) in another article[3]

$$G_{k+1}\delta_k = \hat{y}_k \tag{1.4}$$

where $\hat{y}_k = y_k + \gamma_k\mu_k/\delta_k^T\mu_k$, $\delta_k^T\mu_k \neq 0, \mu_k$ is the vector parameter. Zhang proved that $\hat{y}_k$ in (1.4) can be better approximates $\nabla^2 f(x_{k+1})$ than y_k.

Notice that, if we define $R_k = f_{k+1} - f_k - g_k^T\delta_k$ then we have $R_k \approx \frac{1}{2}\delta_k^T G_{k+1}\delta_k$ for enough small δ_k. Thus

[1]**Corresponding author**. *E-mail address*:wenshu119@163.com

IJNS.2006.12.15/044

$$\delta_k^T G_{k+1}\delta_k \approx 2R_k \tag{1.5}$$

In this paper, we combine (1.4) and (1.5) to the following weighted form

$$\delta_k^T G_{k+1}\delta_k = \theta(\delta_k^T y_k + \gamma_k) + 2(1-\theta)R_k,\ \theta \in [0,1].$$

Then we consider a new extensive equation as follows

$$\delta_k^T B_{k+1}\delta_k = \theta_1\delta_k^T y_k + 2\theta_2 R_k + \theta_3\gamma_k \tag{1.6}$$

where the scalars $\theta_1, \theta_2, \theta_3 \geq 0, \theta_1 + \theta_2 > 0$.

Based on (1.6), we deduced a new class of algorithms to solve the optimization problem. Throughout this paper, we call (1.6) the generalized quasi-Newton equation, and call the algorithms based on (1.6) generalized quasi-Newton algorithms.

The generalized quasi-Newton equation (1.6) makes use of not only gradient of the objective function, but also the value of objective function. Thus compared with the standard quasi-Newton equation, the new algorithms make good use of the information resources. In addition, the generalized quasi-Newton equation is very extensive. It is the expansion of the standard quasi-Newton equation (1.2). They include the algorithms in Zhang's paper[2][3] and also in Jiao'[4] even the class of Broyden. From the above, we can analyze that the algorithms may be better when $\theta_1 + \theta_2 = 1$ and $\theta_3 = \theta_1$.Actually, numerical experiment at Section 4 will prove this point.

This paper is organized as follows: we give the generalized quasi-Newton algorithms which are based on the generalized quasi-Newton equation in Section 2. In Section 3, we prove that the new algorithms have global convergence and the superlinear convergence. Section 4 presents the numerical experiment.

In the remainder of the paper, we use the following notation: $\|\cdot\|$denotes the Euclidean norm; x_* is a minimizer of f and G_* which is the Hessian matrix of f at$x_*$is positive-definite.

2 The Generalized Quasi-Newton Algorithms

In order to obtain the generalized quasi-Newton algorithms, firstly, we consider the rank 1 correction

$$B_{k+1} = B_k + uv^T,$$

where $u, v \in R^n$.

Let $Q_k = \theta_1\delta_k^T y_k + 2\theta_2 R_k + \theta_3\gamma_k$, then from (1.6) we have $Q_k = \delta_k^T B_k\delta_k + \delta_k^T uv^T\delta_k$. Therefore, if $v^T\delta_k \neq 0$, then $\delta_k^T u = \frac{1}{v^T\delta_k}(Q_k - \delta_k^T B_k\delta_k)$. Let $U = \frac{1}{v^T\delta_k}(\frac{Q_k}{y_k^T\delta_k}y_k - B_k\delta_k)$, we obtain the rank 1 correction $B_{k+1} = B_k + \frac{1}{v^T\delta_k}(\frac{Q_k}{y_k^T\delta_k}y_k - B_k\delta_k)v^T$. Since the Hessian matrix $G(x)$ is symmetric, we require that its approximate B_{k+1} be symmetric also. Therefore, we denote $v = \frac{Q_k}{y_k^T\delta_k}y_k - B_k\delta_k$ and obtain the symmetric rank 1 correction formula

$$B_{k+1} = B_k + \frac{1}{Q_k - \delta_k^T B_k\delta_k}(\frac{Q_k}{y_k^T\delta_k}y_k - B_k\delta_k)(\frac{Q_k}{y_k^T\delta_k}y_k - B_k\delta_k)^T.$$

The weaknesses of rank 1 correction formula are: First, it can't guarantee that the matrix $B_{k+1}$is always positive-definite; second, it is possible that the numeral is unsteady.

In order to overcome these weaknesses, we directly consider the rank 2 correction which is different from Jiao[4]. Let

$$B_{k+1} = B_k + auu^T + bvv^T,$$

where the coefficients $a, b \in R$ and the vector $u, v \in R^n$.

Exploiting $u = \frac{1}{v^T\delta_k}(\frac{Q_k}{y_k^T\delta_k}y_k - B_k\delta_k)$, we choose $v = y_k$ and obtain

$$B_{k+1} = B_k + \frac{a}{(y_k^T\delta_k)^2}[(\frac{Q_k}{y_k^T\delta_k})^2 y_k y_k^T - \frac{Q_k}{y_k^T\delta_k}(y_k\delta_k^T B_k + B_k\delta_k y_k^T) + B_k\delta_k\delta_k^T B_k] + by_k y_k^T.$$

In order to simplify the above formula, we let $a = (y_k^T\delta_k)^2$. Because of B_{k+1} satisfies (1.6), then we have $b = \frac{(Q_k-\delta_k^T B_k\delta_k)(1-Q_k+\delta_k^T B_k\delta_k)}{(y_k^T\delta_k)^2}$. Therefore, we have the following rank 2 correction formulae

$$B_{k+1} = B_k - \frac{B_k\delta_k\delta_k^T B_k}{\delta_k^T B_k\delta_k} + \frac{Q_k y_k y_k^T}{(y_k^T\delta_k)^2} + V_k V_k^T$$

where $V_k = (\delta_k^T B_k\delta_k)^{\frac{1}{2}}(\frac{y_k}{y_k^T\delta_k} - \frac{B_k\delta_k}{\delta_k^T B_k\delta_k})$. Since $V_k^T\delta_k = 0$,the formulae still satisfy (1.6) even if the last term multiplies the constant Φ_k.

Thus, we obtain a class of rank 2 correction formulae with four parameters

$$B_{k+1} = B_k - \frac{B_k\delta_k\delta_k^T B_k}{\delta_k^T B_k\delta_k} + \frac{Q_k y_k y_k^T}{(y_k^T\delta_k)^2} + \Phi_k V_k V_k^T \tag{2.1}$$

Obviously, when $\theta_1 = 1, \theta_2 = \theta_3 = 0, \Phi_k = 0$, (2.1)is the famous class of Broyden. Let $H_k = B_k^{-1}$then

$$H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{\delta_k\delta_k^T}{Q_k} + \phi_k z_k z_k^T \tag{2.2}$$

where $z_k = (y_k^T H_k y_k)^{\frac{1}{2}}(\frac{\delta_k}{y_k^T\delta_k} - \frac{H_k y_k}{y_k^T H_k y_k})$ and H_{k+1} is the approximation of inverse of Hessian matrix G_{k+1}. Here the parameters Φ_k and ϕ_k satisfy[1]

$$\phi_k = \frac{\Phi_k - 1}{\Phi_k(1-\mu_k) - 1}, \quad \mu_k = \frac{\delta_k^T B_k\delta_k y_k^T H_k y_k}{(y_k^T\delta_k)^2}.$$

Then we present the generalized quasi-Newton algorithms (**I**)

Step 0: Select an initial point $x_1 \in R^n$ and the positive-definite matrix $B_1 \in R^{n\times n}$ or $H_1 \in R^{n\times n}$, choose a sufficiently small constant $\varepsilon > 0$. Let $k = 1$.

Step 1: If $\|g_k\| = 0$, stop; otherwise, go to Step 2.

Step 2: Solve $B_k d_k = -g_k$or $d_k = -H_k g_k$ to get the search direction d_k.

Step 3: Starts with $\lambda = 1$, find $x_{k+1} = x_k + \delta_k$, where $\delta_k = \lambda_k d_k$, $\lambda_k > 0$ and satisfies the Wolfe search

$$\begin{aligned} &g(x_k + \lambda_k d_k)^T d_k \ge \beta g_k^T d_k, && \alpha < \beta < 1, \\ &f(x_k + \lambda_k d_k) \le f(x_k) + \alpha\lambda_k g_k^T d_k, && 0 \le \alpha < 1/2. \end{aligned}$$

Step 4: If $\|x_{k+1} - x_k\| = 0$, stop; otherwise, go to Step 5.

Step 5: Update B_{k+1} or $H_{k+1}$by formula (2.1) or (2.2), where $Q_k = \theta_1\delta_k^T y_k + 2\theta_2 R_k + \theta_3\gamma_k$, and γ_k is defined by

$$\gamma_k = \begin{cases} 3g_{k+1}^T\delta_k + 3g_k^T\delta_k + 6(f_k - f_{k+1}), & \text{if } Q_k \ge \varepsilon\|\delta_k\|^2; \\ 0, & \text{otherwise.} \end{cases}$$

Step 6: Let $k = k + 1$, and go to Step 1.

3 The Convergence Analysis

As the quasi-Newton algorithms are discussed in [1], we can easily prove that the generalized quasi-Newton algorithms (**I**) also have the properties of symmetric positive-definite. The algorithms can terminate in the *n*th step for quadratic function and not be changed after linear transformation. We state the three properties as follows.

Theorem 3.1 If B_k is a symmetry positive-definite matrix, and then there is a sufficiently small constant $\varepsilon > 0$, such that B_{k+1} which is updated by (2.1) is symmetric positive-definite.

Theorem 3.2 If $f(x)$ is a quadratic function and $G(x)$ is the Hessian matrix, the following equations hold with the exact line search

$$H_{i+1} y_j = \delta_j/(\theta_1 + \theta_2), j = 0, 1, \cdots, i,$$

$$\delta_i^T G\delta_j = 0, \qquad j = 0, 1, \cdots, i-1.$$

The algorithms (**I**) must terminate at the $m+1 \le n$iteration; if $m+1 = n$, then $H_n = G^{-1}$.

Theorem 3.3 If $(H_k)_x = A^{-1}(H_k)_y A^{-T}$ holds for all k, then the algorithms (**I**) are not changed with fixed steplength λ_k after linear transformation $y = Ax + a$, where $A \in R^{n\times n}$,$a \in R^n$, and $(H_k)_x$,$(H_k)_y$ mean that H_k calculates from the vector x and y, respectively. In order to prove that the algorithms (**I**) have the properties of the globe and superlinear convergence which are the main results of this paper, we give the following assumption which is weaker than [5].

Assumption$H$ The function $f(x)$in the problem (1.1) is twice continuously differentiable. It is forcible convex over the level set D.

Theorem3.4[6] For $f(x)$ is a forcible convex function in the Euclidean space R^n,$x_0 \in R^n$. Let

$$D(x_0) = \{x \in R^n \mid f(x) \le f(x_0)\},$$

$$U^* = \{x \in R^n \mid f(x) \le f(z), \forall z \in R^n\},$$

Then the level set $D(x_0)$ is bounded and not empty; both $D(x_0)$ and U^* are closed convex.

Lemma 3.5[6] $f(x)$ is forcible convex over the level set D, if and only if, there is $m > 0$, such that $m\|z\|^2 \le z^T G(x)z$ for all $z \in R^n$ and all $x \in D$.

From the above Assumption H, Theorem 3.4 and Lemma 3.5, we have $G(x)$ are bounded over the level set D for all $x \in D$, namely there is $m, M > 0$, such that $m\|z\|^2 \le z^T G(x)z \le M\|z\|^2$ for all $z \in R^n$.

For the arbitrary matrix $B \in R^{n\times n}$, we define

$$\Psi(B) = tr(B) - \ln[\det(B)] = \sum_{i=1}^{n}(\lambda_i - \ln\lambda_i),$$

where $\lambda_1, \lambda_2, \cdots, \lambda_n$ are the eigenvalue of matrix B.

Similar to [5], we can obtain the following Lemma3.6 easily.

Lemma 3.6 If Assumption H holds, then

1) There are the constants $C_1 > 0$, such that

$$f_{k+1} - f_* \le (1 - \alpha m C_1 \cos^2\omega_k)(f_k - f_*). \tag{3.1}$$

where $-g_k^T\delta_k = \|g_k\| \cdot \|\delta_k\| \cos\omega_k$, $f_* = f(x_*)$.

2) There is the constants $0 \le C_2 < 1$, such that

$$f_{k+1} - f_* \le C_2^k(f_1 - f_*) \tag{3.2}$$

Lemma 3.7 If Assumption H holds, the scalar $\Phi \in [0, 1)$. When $\delta_k \ne 0$, we have

$$\Psi(B_{k+1}) \le \Psi(B_k) + [\overline{M}(\frac{M}{m})^2 - \ln\overline{m} - (1-\Phi)] + (1-\Phi)\ln q_k$$
$$+(1-\Phi)[1 - \frac{(1-\Phi)m^2 - \Phi(M^2\cos^2\omega_k + 2Mm\cos\omega_k + m^2\cos^2\omega_k)}{(1-\Phi)m^2\cos^2\omega_k}q_k], \tag{3.3}$$

where $\overline{M} = (\theta_1+\theta_2)M + \theta_3\varepsilon$, $\overline{m} = (\theta_1+\theta_2)m - \theta_3\varepsilon$ and $q_k = \frac{\delta_k^T B_k\delta_k}{\delta_k^T\delta_k} = \frac{\|B_k\delta_k\|}{\|\delta_k\|}\cos\omega_k$.

According to Lemma3.6 and 3.7, we can now prove the globe and superlinear convergence.

Theorem 3.8 If Assumption H holds, x_1 and B_1 is the initial point and symmetry positive-definite matrix respectively, $\Phi \in [0, 1)$, then the sequence x_k generated by the algorithms (**I**) is convergent at x_*.

Proof: When $\delta_k \ne 0$,B_k can keep positive-definite for k sufficiently large and thus $\Psi(B_k) > 0$.

Suppose

$$\lim_{k\to\infty}\cos^2\omega_k = 0,$$

then there is $k_1 > 0$, we have $\Phi\cos^2\omega_k \le \frac{1}{2}(1-\Phi)$ for all $k > k_1$.

Let $T_k = \frac{(1-\Phi)-\Phi\cos^2\omega_k}{(1-\Phi)\cos^2\omega_k}$, then $T_k > \frac{1}{2\cos^2\omega_k}$, hence $\ln T_k > -\ln 2 - \ln\cos^2\omega_k$. It is easy to understand that $1 - t \le -\ln t$ holds for $t > 0$. Hence from (3.3) we obtain

$$\begin{aligned}\Psi(B_{k+1}) &\leq \Psi(B_k)+(\overline{M}-\ln\overline{m}-1)+(1-\Phi)\ln q_k+(1-\Phi)[1-q_kT_k]\\ &\leq \Psi(B_k)+(\overline{M}-\ln\overline{m}-1)+(1-\Phi)\ln q_k+(1-\Phi)\ln q_kT_k\\ &\leq \Psi(B_k)+(\overline{M}-\ln\overline{m}-1)-(1-\Phi)\ln T_k\\ &< \Psi(B_k)+(\overline{M}-\ln\overline{m}-1)+(1-\Phi)(\ln 2+\ln\cos^2\omega_k)\end{aligned}$$

Thus, according to Assumption H and the k_1, there is $k_2 > k_1 > 0$ such that

$$\ln\cos^2\omega_k \leq -\frac{2}{1-\Phi}[\overline{M}-\ln\overline{m}-1+(1-\Phi)\ln 2]$$

for all $k > k_2$.Thus

$$\begin{aligned}&\Psi(B_{k+1}) < \Psi(B_k)-[\overline{M}-\ln\overline{m}-1+(1-\Phi)\ln 2]\\ &\leq\cdots\leq \Psi(B_{k_2})-[\overline{M}-\ln\overline{M}-1+(1-\Phi)\ln 2](k-k_2+1)\end{aligned}$$

It is easy to find that the term inside the above square brackets is positive. Then the right tends to negative infinite for the sufficient large k. This is contradict with $\Psi(B_k) > 0$, so the above supposition can not hold, namely there is $\overline{k}$, such that $\cos\omega_k > \varepsilon$ for all $k > \overline{k}$, where $0 < \varepsilon < \sqrt{\frac{1}{\alpha m C_1}}$. Substitute it into (3.1), we have $f_{k+1}-f_* \leq (1-\alpha m C_1\varepsilon^2)(f_k-f_*)$. It then follows that the sequence $\{f_k\}$ converges to f_*.

On the other hand, from Assumption H and the Taylor's theorem we have $f_k-f_* \geq \frac{1}{2}m\|x_k-x_*\|^2 \geq 0$. Therefore, the sequence $\{x_k\}$ is convergent at x_*.

Theorem 3.9 If Assumption H holds, $G_* = I$, then the algorithms(**I**) satisfies

$$\lim_{k\to\infty}\frac{\|(B_k-I)\delta_k\|}{\|\delta_k\|}=0 \tag{3.4}$$

and λ_k is taken equal to 1 for sufficient large k, thus the sequence $\{x_k\}$ converges to x_* superlinearly.

Proof: With Assumption H and the Taylor's theorem, we can obtain $f_{k+1}-f_* \geq \frac{1}{2}m\|x_{k+1}-x_*\|^2$.

Using this in (3.2), we have $\|x_{k+1}-x_*\|^2 \leq \frac{2}{m}C_2^k(f_1-f_*)$, then

$$\sum_{k=0}^{\infty}\|x_{k+1}-x_*\| \leq [\frac{2}{m}(f_1-f_*)]^{1/2}\sum_{k=0}^{\infty}C_2^{k/2} < +\infty. \tag{3.5}$$

From the Proposition 4 in [7], we know (3.4) holds.

Using Theorem 2.2 in [8], we know (3.4) is equivalent to the fact that the sequence $x_k$superlinearly converges to x_*.

Next, we will prove $\lambda_k \to 1$ (similar to Theorem 6.3 and 6.4 in [9]). From (3.4), (3.5) and

$$\begin{aligned}0 \leq \frac{\|g_k+G_kd_k\|}{\|d_k\|} &= \frac{\|(B_k-G_k)d_k\|}{\|d_k\|} \leq \frac{\|(B_k-G_*)d_k\|}{\|d_k\|}+\frac{\|(G_k-G_*)d_k\|}{\|d_k\|}\\ &\leq \frac{\|(B_k-G_*)\delta_k\|}{\|\delta_k\|}+M\|x_k-x_*\|\end{aligned}$$

we have

$$\lim_{k\to\infty}\frac{\|g_k+G_kd_k\|}{\|d_k\|}=0, \tag{3.6}$$

and hence $-\frac{g_k^Td_k}{\|d_k\|^2} = \frac{d_k^TG_kd_k}{\|d_k\|^2}-\frac{d_k^T(G_kd_k+g_k)}{\|d_k\|^2} > 0$. Therefore, there is $\eta > 0$ such that

$$-g_k^Td_k \geq \eta\|d_k\|^2 \tag{3.7}$$

From Theorem 6.3 in [9], namely $\lim_{k\to\infty}\frac{g_k^Td_k}{\|d_k\|}=0$, we can obtain $\lim_{k\to\infty}\|d_k\|=0$. Combining (3.6) and (3.7), we know that there is $u_k \in (x_k, x_k+d_k)$ such that the following inequality holds for sufficient large k and the above η

$$f(x_k+d_k)-f(x_k)-\frac{1}{2}g_k^Td_k=\frac{1}{2}d_k^T(G(u_k)d_k+g_k) \leq (\frac{1}{2}-\alpha)\eta\|d_k\|^2 \leq (\alpha-\frac{1}{2})g_k^Td_k.$$

Hence$f(x_k + d_k) - f(x_k) \le \alpha g_k^T d_k$.

On the other hand, similar to the above proof, we know that there is $v_k \in (x_k, x_k + d_k)$ such that the following inequality holds for the sufficient large k and the above η.

$$d_k^T g(x_k + d_k) = d_k^T (g(x_k) + G(v_k) d_k) \le \eta\beta \|d_k\|^2 \le -\beta d_k^T g(x_k).$$

Therefore, $\lambda_k = 1$ satisfies the Wolfe search for k sufficiently large.

4 Numerical Experiment

In this section, we will solve the Rosenbrock function which is the typical unconstrained optimization problem,namely $f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$,where $x = (x_1, x_2)^T$.

We know $x_* = (1,1)^T$ and $f(x_*) = 0$. Let $\alpha = 1/4$, $\beta = 2/3$, $\varepsilon = 10^{-15}$ and $\Phi = 1/2$.

Table 1: The scalars θ_1=0.9, θ_2=0.5, θ_3=0.6

x_1	x_*	x_*^{Jiao}	x_*^{Zhang}
$\begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 1.36839237531073 \\ 1.87390722131381 \end{pmatrix}$	$\begin{pmatrix} 0.86160303144629 \\ 0.739907870396847 \end{pmatrix}$
Itera.	45	8	6
$\begin{pmatrix} 0.5 \\ 1.5 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0.999999999999999 \end{pmatrix}$	$\begin{pmatrix} 1.01747141620734 \\ 1.03532415220092 \end{pmatrix}$	$\begin{pmatrix} 1.00234389438897 \\ 1.00469825826696 \end{pmatrix}$
Itera.	48	8	24

Table 2: The scalars θ_1=0.9, θ_2=0.5, θ_3=0.6

x_1^T	x_*	x_*^T
(0,0)	43	(1,1)
(2,2)	42	(1,0.999999999999999)
(10,10)	54	(1.00000000000001,1.00000000000001)
(100,100)	109	(1,1)
(-10,-1)	65	(0.999999999999999,0.999999999999999)

Table 3: The initial point $x_1 = (-1.2, 1)^T$ and the scalars $\theta_1 + \theta_2 = 1$, $\Phi = 0$ and $\theta_1 = \theta_3$

θ_1	0.1	0.75	0.5	$1(\theta_3 = 0)$
x_*	$\begin{pmatrix} 1.0000109553952 \\ 10000000009543 \end{pmatrix}$	$\begin{pmatrix} 1.0000000053811 \\ 1.0000000008939 \end{pmatrix}$	$\begin{pmatrix} 1.00000000023 \\ 1.00000000446 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
Time	0.046999999999	0.0469999999997	0.047000000025	0.000109799
Itera.	17	6	9	5

In the tables, x_1 is the initial point. x_*, x_*^{Jiao} and x_*^{Zhang} are the results which are solved by the algorithms in this paper, Jiao's[4] and Zhang's[2] paper, respectively. Itera. and Time mean the number and time of iteration. As reported in Table 1, it is easy to find that our new algorithms behave more efficiently than Jiao's and Zhang's. From Table 2, we can see that the new algorithms in our paper can give us a useful globe convergence result. In Table 3: for $\theta_1 = 1$,$\theta_2 = \theta_3 = 0$ we obtain the BFGS method. Obviously, it has better performance. Therefore, it is very interesting that we also validate that BFGS method is indeed one of the best methods up to now.

References

[1] Yaxiang Yuan, Wenyu Sun: Optimization: theory and algorithms. *Beijing: Science Press*.(2001)

[2] Zhang J Z, Deng NY, Chen L H: New quasi-Newton equation and related methods for unconstrained optimization. *JOTA*.102(1),147-167(1999)

[3] Zhang J Z, Xu C X: Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equation. *Journal of Computational Applied Mathematics*.137(2), 269-278(2001)

[4] Baocong Jiao: A class of non-quasi-Newton algorithms with superlinear convergence. *Numerical Mathematics a Journal of Chinese University*.21(6),178-188(1999)

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

[6] Shuzhong Shi: Convex analysis. *Shanghai: Shanghai scientific & Technical Publishers*(1990)

[7] Grienwank, Toint PH L: Local convergence analysis of partitioned quasi-Newton updates. *Numerical Mathematics*.39,429-448(1982)

[8] Dennis J E, Moré J J: A characterization of superlinear convergence and its application to quasi-Newton methods. *Mathematical Computing*.28,549-560(1971)

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

[10] Dongdong Wang , Hongbo Hua: On Stability Number of Upper Irredundance Number in Graphs. *International Joural of Nonlinear Science*. 2(1),39-42(2006)

[11] Yuchun Wang 1, Lixia Wang, Wenbin Zhang: Application of the Adomian Decomposition Method to Fully Nonlinear Sine-Gordon Equation. *International Joural of Nonlinear Science*. 2(1),29-38(2006)

[12] Wenbin Zhang: Attractor of Cauchy Problem in Dissipative KdV Type Equation.*International Joural of Nonlinear Science*.1(3),155-163(2006)

[13] Sevdzhan Hakkaev, Kiril Kirchev: On the well-posedness and stability of Peakons for a generalized Camassa-Holm equation. *International Joural of Nonlinear Science*. 3(1), 139-148(2006)