Journal of Computational and Applied Mathematics 137 (2001) 269–278

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations ☆

Jianzhong Zhang[a,*], Chengxian Xu[b]

[a] *Department of Mathematics, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong*
[b] *Faculty of Sciences, Xian Jiaotong University, People's Republic of China*

Received 27 March 2000; received in revised form 20 October 2000

Abstract

Quasi-Newton (QN) equation plays a core role in contemporary nonlinear optimization. The usual QN equation employs only the gradients, but ignores the available function value information. In this paper, we derive a class of modified QN equations with a vector parameter which use both available gradient and function value information. The modified quasi-Newton methods maintain most properties of the usual quasi-Newton methods, meanwhile they achieve a higher-order accuracy in approximating the second-order curvature of the problem functions than the usual ones do. Numerical experiments are reported which support the theoretical analyses and show the advantages of the modified QN methods over the usual ones. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Quasi-Newton equation; Broyden family of updates; Curvature approximation; Positive-definite update; Superlinear convergence

1. Introduction

This paper is concerned with quasi-Newton methods for unconstrained optimization

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

where $f:\mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable. Starting from a point x_1 and a symmetric (usually positive-definite) matrix B_1, a quasi-Newton method generates sequences $\{x_k\}$ and $\{B_k\}$ by the iteration $x_{k+1} = x_k + \alpha_k s_k$ and an updating formula for B_k, where α_k is a step length and s_k is a search direction that is obtained by solving the equation $B_k s_k = -g_k$ in which $g_k = \nabla f(x_k)$ is

☆ The research is partially supported by City University of Hong Kong under its Strategic Research Grant #7000944 and the National Natural Science Foundation of China.

* Corresponding author. Tel.: +852-2788-8662; fax: +852-2788-8463.
E-mail address: mazhang@cityu.edu.hk (J. Zhang).

PII: S0377-0427(00)00713-5

the gradient of $f(x)$ at x_k and B_k is an approximation to the Hessian $G_k = \nabla^2 f(x_k)$ of $f(x)$. The updating matrix B_k is required to satisfy the *usual quasi-Newton equation*

$$B_{k+1}\delta_k = y_k, \quad \delta_k = x_{k+1} - x_k, \quad y_k = g_{k+1} - g_k, \tag{1}$$

so that B_{k+1} is a reasonable approximation to G_{k+1}.

The most effective quasi-Newton methods are included in the Broyden family [2] in which the updates are defined by

$$\begin{aligned} B_{k+1}(\mu) &= U(B_k, \delta_k, y_k, \mu) = B_{k+1}(0) + \mu w_k w_k^{\mathrm{T}}, \\ B_{k+1}(0) &= B_k - \frac{B_k \delta_k \delta_k^{\mathrm{T}} B_k}{\delta_k^{\mathrm{T}} B_k \delta_k} + \frac{y_k y_k^{\mathrm{T}}}{\delta_k^{\mathrm{T}} y_k}, \quad w_k = (\delta_k^{\mathrm{T}} B_k \delta_k)^{1/2} \left[\frac{y_k}{\delta_k^{\mathrm{T}} y_k} - \frac{B_k \delta_k}{\delta_k^{\mathrm{T}} B_k \delta_k} \right], \end{aligned} \tag{2}$$

where μ is a scale parameter, and $B_{k+1}(0)$ is the BFGS update. The DFP and SR1 updates are obtained by setting $\mu = 1$ and $\mu = 1/(1 - \delta_k^{\mathrm{T}} B_k \delta_k / \delta_k^{\mathrm{T}} y_k)$, respectively.

The quasi-Newton methods possess a number of important theoretical properties (see [3,4,7,8, 10,14]), for example, quadratic termination, invariance under nonsingular affine transformations, heredity of positive-definite updates, and generating identical iterate points with exact line searches (see [9]), locally and superlinearly convergence under mild conditions (see [7,8]).

A drawback of quasi-Newton equation (1) is that it exploits only the gradient information while the available information in function values is neglected. Attempts have been made to modify either the usual quasi-Newton equation [11,16,18] or the local approximation model [5,6] so that more available information can be exploited. Zhang et al. [20] proposed a modification to the usual quasi-Newton equation (1) that exploits both gradient and function value information. However, the resulting quasi-Newton updates do not preserve the *invariance property*. Another attempt is the introduction of the tensor method [15], which also intends to use the function value information to get better local approximations. That type of methods need to employ higher-order derivatives. We shall focus on the quasi-Newton methods which use only first-order derivatives.

The aim of this paper is to derive a class of modified quasi-Newton equations with a vector parameter which use both available function value and gradient information. Some theoretical properties that distinguish the modified quasi-Newton methods from the usual quasi-Newton methods are presented and numerical results about the modified quasi-Newton methods are reported. The modified quasi-Newton equations are derived in Section 2. In Section 3, it is shown that the modified quasi-Newton updates generate more accurate second-order curvature approximations than the usual quasi-Newton updates do, and line search conditions to ensure positive-definite updates are discussed. It is shown in Section 4 that the modified BFGS and DFP methods preserve the local superlinear convergence property of the usual BFGS and DFP methods. Numerical results are presented in Section 5. Throughout the paper, $\|\cdot\|$ denotes the Euclidean vector norm and its induced matrix norm.

2. Deriving new quasi-Newton equations

In this section we derive the modified quasi-Newton equations which exploit not only the gradients but also the function values. We recall the fact that quasi-Newton equation (1) is derived from

$y_k = \int_0^1 G(x_k + t\delta_k)\,\mathrm{d}t\,\delta_k$. Now we assume that the function f is smooth enough. From

$$f(x_k) = f(x_{k+1}) - g_{k+1}^{\mathrm{T}}\delta_k + \frac{1}{2}\,\delta_k^{\mathrm{T}} G_{k+1}\delta_k - \frac{1}{3!}\,\delta_k^{\mathrm{T}}(T_{k+1}\delta_k)\delta_k + \mathrm{O}(\|\delta_k\|^4),$$

$$\delta_k^{\mathrm{T}} g_k = \delta_k^{\mathrm{T}} g_{k+1} - \delta_k^{\mathrm{T}} G_{k+1}\delta_k + \frac{1}{2}\,\delta_k^{\mathrm{T}}(T_{k+1}\delta_k)\delta_k + \mathrm{O}(\|\delta_k\|^4),$$

where $T_{k+1} \in \mathbb{R}^{n\times n\times n}$ is the tensor of f at

$$x_{k+1} \quad \text{and} \quad \delta_k^{\mathrm{T}}(T_{k+1}\delta_k)\delta_k = \sum_{i,j,l=1}^{n} \frac{\partial^3 f(x_{k+1})}{\partial x^i \partial x^j \partial x^l}\,\delta_k^i \delta_k^j \delta_k^l,$$

we obtain, by cancelling the terms which include the tensor,

$$\delta_k^{\mathrm{T}} G_{k+1}\delta_k = (g_{k+1} - g_k)^{\mathrm{T}}\delta_k + 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^{\mathrm{T}}\delta_k + \mathrm{O}(\|\delta_k\|^4).$$

Since $B_{k+1}\delta_k$ is required to approximate $G_{k+1}\delta_k$, it is reasonable to require

$$\delta_k^{\mathrm{T}} B_{k+1}\delta_k = \delta_k^{\mathrm{T}} y_k + \theta_k, \quad \theta_k = 6(f_k - f_{k+1}) + 3(g_k + g_{k+1})^{\mathrm{T}}\delta_k$$

(an equivalent form of this equation was used in [19] to derive some non-quasi-Newton updates). Then one choice for a good approximation to $B_{k+1}\delta_k$ is given by

$$B_{k+1}\delta_k = y_k + \frac{\theta_k}{\delta_k^{\mathrm{T}} u}\,u,$$

where $u \in \mathbb{R}^n$ is any vector such that $\delta_k^{\mathrm{T}} u \neq 0$. Note that the usual quasi-Newton equation (1) is obtained by neglecting the second term on the right-hand side. This gives a class of modified quasi-Newton equations in the form

$$B_{k+1}\delta_k = \hat{y}_k, \quad \hat{y}_k = y_k + \frac{\theta_k}{\delta_k^{\mathrm{T}} u}\,u \quad \text{with } \delta_k^{\mathrm{T}} u \neq 0. \tag{3}$$

Based on this equation, modified quasi-Newton updates in the Broyden family are given in the form $U(B_k, \delta_k, \hat{y}_k, \mu)$, i.e., y_k is replaced by $\hat{y}_k$ in all previous formulae.

Different choices of the vector u in (3) give different quasi-Newton equations. The particular choice $u = \delta_k$ gives the modified quasi-Newton equation (see [20]),

$$B_{k+1}\delta_k = \hat{y}_k^{\mathrm{ZDC}}, \quad \hat{y}_k^{\mathrm{ZDC}} = y_k + \frac{\theta_k}{\delta_k^{\mathrm{T}}\delta_k}\,\delta_k.$$

A disadvantage of this choice is that the resulting update is not invariant under a nonsingular affine transformation on variables (see [17]).

Other choices for the vector u can be $B_k s_k (= -g_k)$ or g_{k+1} if $\delta_k^{\mathrm{T}} g_{k+1} \neq 0$. However, theoretical analyses and numerical experiments (see [17]) show that the choice $u = y_k$ is the one to be strongly recommended. The resulting quasi-Newton equation is denoted by

$$B_{k+1}\delta_k = \hat{y}_k^{\mathrm{HU}}, \quad \hat{y}_k^{\mathrm{HU}} = \left(1 + \frac{\theta_k}{\delta_k^{\mathrm{T}} y_k}\right) y_k.$$

This is Huang's quasi-Newton equation $B_{k+1}\delta_k = \beta_k y_k$ (see [12]) with the particular value $\beta_k = 1 + \theta_k/\delta_k^{\mathrm{T}} y_k$. Note that Biggs [1] gives the inverse BFGS update with $\hat{y}_k^{\mathrm{HU}}$ replacing y_k so that the above equation is satisfied.

3. Properties of quasi-Newton methods satisfying new quasi-Newton equations

In this section we study properties of the modified quasi-Newton methods. These methods possess most of the properties which the usual quasi-Newton methods mentioned in Section 1 own, for example, the finite termination, invariance, and generating identical iterate points with exact line searches. In this section we just discuss the two properties that distinguish the modified quasi-Newton methods from the usual quasi-Newton methods. The detailed discussion for other properties can be found in [17]. It is assumed that the function f is smooth enough.

The following result is proved in [20] for the case $u = \delta_k$, and also holds for all u with $u^T\delta_k \neq 0$.

Theorem 3.1. *Assume that the function $f(x)$ is sufficiently smooth. If $\|\delta_k\|$ is sufficiently small, then for any vector u with $\delta_k^T u \neq 0$ we have*

$$\delta_k^T[G_{k+1}\delta_k - \hat{y}_k] = O(\|\delta_k\|^4), \tag{4}$$

$$\delta_k^T[G_{k+1}\delta_k - y_k] = O(\|\delta_k\|^3). \tag{5}$$

Let $B_{k+1} = U(B_k, \delta_k, y_k, \mu)$ and $\hat{B}_{k+1} = U(B_k, \delta_k, \hat{y}_k, \mu)$. Then from $B_{k+1}\delta_k = y_k$, $\hat{B}_{k+1}\delta_k = \hat{y}_k$, and Eqs. (4) and (5), we have

$$\delta_k^T\hat{B}_{k+1}\delta_k = \delta_k^T G_{k+1}\delta_k + O(\|\delta_k\|^4), \quad \delta_k^T B_{k+1}\delta_k = \delta_k^T G_{k+1}\delta_k + O(\|\delta_k\|^3).$$

These two equations show that if assumptions of Theorem 3.1 are satisfied and if $\|\delta_k\|$ is sufficiently small, then the curvature $\delta_k^T\hat{B}_{k+1}\delta_k$ given by the modified quasi-Newton updating matrix $\hat{B}_{k+1}$ approximates the second-order curvature $\delta_k^T G_{k+1}\delta_k$ with a higher precision than the curvature $\delta_k^T B_{k+1}\delta_k$ does.

If the sequence $\{x_k\}$ converges to x^* with $g(x^*) = 0$, then $\delta_k \to 0$ as $k \to \infty$, and the results of Theorem 3.1 hold for all sufficiently large k. The results of the theorem also hold at a point with small enough $\|\delta_k\|$, for example, with sufficiently small step length in line search methods or with sufficiently small radius in trust region methods.

The following corollary is an immediate consequence of Theorem 3.1 and shows that the modified quasi-Newton updates generate exactly the second-order curvature $\delta_k^T G_{k+1}\delta_k$ for cubic functions.

Corollary 3.2. *If f is a cubic function, then $\delta_k^T\hat{B}_{k+1}\delta_k = \delta_k^T G_{k+1}\delta_k$.*

Another property which we are concerned with is the heredity of positive-definite updates. It is known that the usual quasi-Newton updates in Broyden family (2) preserve positive-definite updates for all values of $\mu > \bar{\mu} = 1/[1 - \delta_k^T B_k\delta_k y_k^T B_k^{-1} y_k/(\delta_k^T y_k)^2]$, if and only if condition

$$\delta_k^T y_k > 0 \tag{6}$$

holds (see [10]). Condition (6) is achieved when either line search is exact or the step length α_k satisfies the Wolfe conditions

$$f(x_k + \alpha_k s_k) \leqslant f(x_k) + \rho\alpha_k g_k^T s_k, \quad \rho \in (0, 1/2), \tag{7}$$

$$g(x_k + \alpha_k s_k)^T s_k \geqslant \sigma g_k^T s_k, \quad \sigma \in (\rho, 1). \tag{8}$$

For the modified quasi-Newton updates, condition (6) is replaced by the condition

$$\delta_k^{\mathrm{T}} \hat{y}_k = \delta_k^{\mathrm{T}} y_k + \theta_k > 0. \tag{9}$$

This shows that the sign of $\delta_k^{\mathrm{T}} \hat{y}_k$ is independent of the choice of the vector u, but relies upon the properties of the function $f(x)$ and the step length α_k.

For general nonlinear smooth functions, it follows from (4) that in the regions where the Hessian $G(x)$ is positive definite, condition (9) will be satisfied for sufficiently small δ_k. As a consequence, when the sequence $\{x_k\}$ generated by a modified quasi-Newton method converges to a strong local minimum point x^* of f, condition (9) will hold for all sufficiently large k.

Proposition 3.3. *Assume that the function f is twice continuously differentiable. If $\{x_k\}$ converges to a point x^* at which $g(x^*) = 0$ and $G(x^*)$ is positive definite, then*

$$\lim_{k\to\infty} \frac{\delta_k^{\mathrm{T}} \hat{y}_k}{\delta_k^{\mathrm{T}} y_k} = 1.$$

Proof. Since $\{x_k\}$ converges to x^*, it follows from Taylor expansions of f and g at x_k that $\theta_k = \mathrm{O}(\|\delta_k\|^3)$ holds for sufficiently large k. Since $G(x^*)$ is positive definite, there exists $c_1 > 0$ such that $\delta_k^{\mathrm{T}} y_k \geqslant c_1 \|\delta_k\|^2$ holds for all sufficiently large k. Then we have $\lim_{k\to\infty} \theta_k/\delta_k^{\mathrm{T}} y_k = 0$, and the conclusion follows from (9). □

At a point x_k outside the above-mentioned regions it is natural to know when we can expect $\delta_k^{\mathrm{T}} \hat{y}_k > 0$ if the step length α_k satisfies the Wolfe conditions. From (7), (8) and the definition of θ_k we have $\delta_k^{\mathrm{T}} \hat{y}_k = 6(f_k - f_{k+1}) + 2g_k^{\mathrm{T}}\delta_k + 4g_{k+1}^{\mathrm{T}}\delta_k \geqslant 2(2\delta_k^{\mathrm{T}} g_{k+1} + (1-3\rho)\delta_k^{\mathrm{T}} g_k) \geqslant 2(2\sigma + 1 - 3\rho)\delta_k^{\mathrm{T}} g_k$. Hence we cannot expect $\delta_k^{\mathrm{T}} \hat{y}_k > 0$ unless we have a large decrease in f, and $\delta_k^{\mathrm{T}} g_{k+1}$ is either positive, or negative but close to zero (if a rather accurate line search is conducted). To achieve this purpose, choose the values of ρ and σ with $1/2 > \rho > 1/3$ and $(2\sigma - 3\rho + 1) < 0$.

When the values σ and ρ $(\sigma > \rho)$ in conditions (7) and (8) are used in practical calculations, the case $\delta_k^{\mathrm{T}} \hat{y}_k \leqslant 0$ may occur (though it happens very rarely). The following strategy

$$\theta_k = (\varepsilon - 1)\delta_k^{\mathrm{T}} y_k, \quad \text{if } \theta_k < (\varepsilon - 1)\delta_k^{\mathrm{T}} y_k, \ \varepsilon \in (0,1) \tag{10}$$

can then be used in practice to give a restriction to the value θ_k so that $\delta_k^{\mathrm{T}} \hat{y}_k \geqslant \varepsilon \delta_k^{\mathrm{T}} y_k$ and positive-definite updates are preserved.

4. Local convergence property

In this section we study the local convergence property of the modified BFGS and DFP methods. The following assumption is required.

(A1) *The function f is twice continuously differentiable in an open convex set D, and $x^* \in D$ is a local minimizer of f at which $g(x^*) = 0$ and $G(x^*)$ is positive definite.*

Since we are only concerned with the local convergence property of the methods, we assume that $x_0 \in D$ and B_0 (or H_0) are sufficiently close to x^* and $G(x^*)$ (or $G(x^*)^{-1}$), respectively. Then according to Proposition 3.3 we can further assume that $\delta_k^{\mathrm{T}}\hat{y}_k > 0$ holds, and hence $\hat{y}_k = y_k + \theta_k u/\delta_k^{\mathrm{T}}u$ with $\delta_k^{\mathrm{T}}u \neq 0$ for all k.

For the usual DFP method (with $\alpha_k = 1$ and $B_k = U(B_{k-1}, \delta_{k-1}, y_{k-1}, 1)$),

$$x_{k+1} = x_k + \delta_k \quad \text{and} \quad B_k\delta_k = -g_k, \tag{11}$$

the local convergence property is given by the following result (see [7,8]).

Theorem 4.1. *Suppose assumption* (A1) *holds. Let z_k be any vector which satisfies the equation*

$$\|z_k - G(x^*)\delta_k\| \leqslant c(\|x_{k+1} - x^*\| + \|x_k - x^*\|)\|\delta_k\| \tag{12}$$

and

$$\frac{1}{\beta}\|\delta_k\| \leqslant \|z_k\| \leqslant \beta\|\delta_k\| \tag{13}$$

for all k, where $c>0$ and $\beta>0$ are constants. Let the sequence $\{x_k\} \subset D$ be generated by the DFP method (11) *with $\alpha_k = 1$ and $B_k = U(B_{k-1}, \delta_{k-1}, z_{k-1}, 1)$. If x_0 and B_0 are sufficiently close to x^* and $G(x^*)$, respectively, then the bounded deterioration condition*

$$\|B_{k+1} - G(x^*)\|_M \leqslant [\sqrt{1-\gamma\xi^2} + \gamma_1\sigma(x_k, x_{k+1})]\|B_k - G(x^*)\|_M + \gamma_2\sigma(x_k, x_{k+1}) \tag{14}$$

holds, the sequences $\{\|B_k\|\}$ and $\{\|B_k^{-1}\|\}$ are bounded, the limit

$$\lim_{k\to\infty} \|[B_k - G(x^*)]\delta_k\|/\|\delta_k\| = 0 \tag{15}$$

holds, and the local superlinear convergence of the sequence $\{x_k\}$ follows, where $\gamma \in [3/8, 1]$, $\xi \in [0, 1]$, γ_1 and γ_2 are positive constants, and $\|A\|_M$ is the weighted matrix norm defined by $\|A\|_M = \|MAM\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm and $M^2 = G(x^*)^{-1}$.

For the normal DFP method, i.e. $z_{k-1} = y_{k-1}$, it is clear that under assumption (A1), (13) holds for sufficiently large k. If the function $f(x)$ also satisfies the following assumption: (A2) *The Hessian $G(x)$ is Lipschitz continuous in D, i.e., there is a $L>0$ such that*

$$\|G(x) - G(y)\| \leqslant L\|x - y\|, \quad \forall x, y \in D, \tag{16}$$

then (12) holds with $c = L/2$ for the sequence $\{x_k\}$, and the conclusions of Theorem 4.1 follow.

For the modified DFP method, that is, $z_{k-1} = \hat{y}_{k-1}$ and the matrix B_k is replaced by $\hat{B}_k = U(\hat{B}_{k-1}, \delta_{k-1}, \hat{y}_{k-1}, 1)$ in method (11), if there exist constants $c_1 > 0$ and $\beta_1 > 0$ such that

$$\|\hat{y}_k - G(x^*)\delta_k\| \leqslant c_1(\|x_{k+1} - x^*\| + \|x_k - x^*\|)\|\delta_k\|, \quad \forall x_k, x_{k+1} \in D \tag{17}$$

and

$$\frac{1}{\beta_1}\|\delta_k\| \leqslant \|\hat{y}_k\| \leqslant \beta_1\|\delta_k\|, \quad \forall x_k, x_{k+1} \in \mathcal{N}(x^*, \varepsilon') \stackrel{\text{def}}{=} \{x \mid \|x - x^*\| \leqslant \varepsilon'\} \tag{18}$$

hold, then the local superlinear convergence of the sequence $\{x_k\}$ follows from Theorem 4.1.

Note that (17) and (18) are extensions of (12) and (13), respectively. In the following, we will prove (17) and (18) for general vector u under the assumption:

(A3) $|u^{\mathrm{T}}\delta_k| \geqslant \varrho\|u\| \cdot \|\delta_k\|$, $\varrho \in (0, 1]$.

This assumption is realistic. If $u=\delta_k$, the condition is satisfied with $\varrho=1$. If $u=y_k$, then $\delta_k^{\mathrm T}u=\delta_k^{\mathrm T}y_k=\delta_k^{\mathrm T}\int_0^1 G(x_k+t\delta_k)\,\mathrm{d}t\,\delta_k$. The positive definiteness of $G(x^*)$ and continuity of $G(x)$ imply that there is a neighbourhood $\mathcal{N}(x^*)$ and two positive constants $\hat{\lambda}$ and $\tilde{\lambda}$ such that $\tilde{\lambda}\|v\|^2\leqslant v^{\mathrm T}G(x)v\leqslant\hat{\lambda}\|v\|^2$ holds for all $x\in\mathcal{N}(x^*)$ and any $v\in\mathbb{R}^n$, and hence it is easy to derive that (A3) is satisfied with $\varrho=\tilde{\lambda}/\hat{\lambda}$ in $\mathcal{N}(x^*)$.

Lemma 4.2. *Under assumptions* (A1)–(A3), *result* (17) *is true.*

Proof. From the definition of $\hat{y}_k$ and assumption (A3), we have

$$\|\hat{y}_k-G(x^*)\delta_k\|\leqslant\|y_k-G(x^*)\delta_k\|+|\theta_k|/(\varrho\|\delta_k\|).$$

Since $f_{k+1}=f_k+g_k^{\mathrm T}\delta_k+\frac{1}{2}\delta_k^{\mathrm T}G(x_k+\zeta\delta_k)\delta_k$, $\zeta\in(0,1)$ and $g_{k+1}^{\mathrm T}\delta_k=g_k^{\mathrm T}\delta_k+\int_0^1\delta_k^{\mathrm T}G(x_k+t\delta_k)\,\mathrm{d}t\,\delta_k$, using (16) we have

$$\begin{aligned}|\theta_k|&=3\left|\delta_k^{\mathrm T}\int_0^1[G(x_k+\zeta\delta_k)-G(x_k+t\delta_k)]\,\mathrm{d}t\,\delta_k\right|\\&\leqslant\frac{3L}{2}(\|x_{k+1}-x^*\|+\|x_k-x^*\|)\|\delta_k\|^2.\end{aligned}\tag{19}$$

Then (17) holds with $c_1=(1+3/\varrho)L/2$. □

Lemma 4.3. *Suppose that assumptions* (A1)–(A3) *hold. Then there exist* $\varepsilon'>0$ *and* $\beta_1>0$ *such that* (18) *holds for all* x_k *and* x_{k+1} *in* $\mathcal{N}(x^*,\varepsilon')$.

Proof. From (13) with y_k replacing z_k and (19) we have

$$\|\hat{y}_k\|\leqslant\|y_k\|+\frac{|\theta_k|}{\varrho\|\delta_k\|}\leqslant\left[\beta+\frac{3L}{\varrho}\sigma(x_k,x_{k+1})\right]\|\delta_k\|,$$
$$\|\hat{y}_k\|\geqslant\|y_k\|-\frac{|\theta_k|}{\varrho\|\delta_k\|}\geqslant\left[\frac{1}{\beta}-\frac{3L}{\varrho}\sigma(x_k,x_{k+1})\right]\|\delta_k\|.$$

Then for $\varepsilon'<\varrho/(3L\beta)$, the lemma is true with $\beta_1=\max\{\beta+3L\varepsilon'/\varrho,(\beta\varrho)/(\varrho-3L\beta\varepsilon')\}$. □

With these two lemmas, the local superlinear convergence of the modified DFP method can be presented as follows.

Theorem 4.4. *Suppose that assumptions* (A1)–(A3) *hold. If* $x_0\in D$ *and* $\hat{B}_0$ *are sufficiently close to* x^* *and* $G(x^*)$, *respectively. Then the sequence* $\{x_k\}$ *generated by the modified DFP method* ($\alpha_k=1$ *for all* k) *superlinearly converges to* x^*.

For the modified BFGS method (with $\alpha_k=1$) we consider the form

$$x_{k+1}=x_k+\delta_k,\quad\delta_k=-\hat{H}_kg_k,\tag{20}$$

where

$$\hat{H}_k = U(\hat{H}_{k-1}, \hat{y}_{k-1}, \delta_{k-1}, 1)$$
$$= \hat{H}_{k-1} + \left(1 + \frac{\hat{y}_{k-1}^{\mathrm{T}} \hat{H}_{k-1} \hat{y}_{k-1}}{\delta_{k-1}^{\mathrm{T}} \hat{y}_{k-1}}\right) \frac{\delta_{k-1}\delta_{k-1}^{\mathrm{T}}}{\delta_{k-1}^{\mathrm{T}} \hat{y}_{k-1}} - \left(\frac{\delta_{k-1}\hat{y}_{k-1}^{\mathrm{T}}\hat{H}_{k-1} + \hat{H}_{k-1}\hat{y}_{k-1}\delta_{k-1}^{\mathrm{T}}}{\delta_{k-1}^{\mathrm{T}}\hat{y}_{k-1}}\right)$$

is the modified inverse BFGS update. The local superlinear convergence of the method can be similarly obtained from Lemmas 4.2, 4.3, and the following additional lemma (see [20]).

Lemma 4.5. *If the sequence* $\{x_k\}$ *generated in* (20) *converges to* x^*, *and* $\{\|\hat{H}_k\|\}$ *and* $\{\|\hat{H}_k^{-1}\|\}$ *are bounded*, *then* $\lim_{k\to\infty} \|[\hat{H}_k - G(x^*)^{-1}]\hat{y}_k\| / \|\hat{y}_k\| = 0$ *ensures that* $\{x_k\}$ *superlinearly converges to* x^*.

Theorem 4.6. *Suppose that assumptions* (A1)–(A3) *hold. If* $x_0(\in D)$ *and* $\hat{H}_0$ *are sufficiently close to* x^* *and* $G(x^*)^{-1}$, *respectively*, *then the sequence* $\{x_k\}$ *generated by the modified BFGS* method ($\alpha_k = 1$ *for all* k) *superlinearly converges to* x^*.

5. Numerical results

This section is devoted to numerical experiments. The purpose is to check whether the modified quasi-Newton methods provide improvements on the corresponding usual quasi-Newton methods. The programs are written in FORTRAN 77 with single precision. The test functions are commonly used unconstrained test problems with standard starting points (see [13]), and a summary of which is given in Table 1.

The methods we implemented are the line search BFGS, SR1 and Hoshino methods with inverse updates. The initial inverse approximations are $H_0 = I$. The line searches determine steplengths α_k satisfying conditions (7) and (8) with $\rho = 0.01$ and $\sigma = 0.9$. The iteration is terminated when one

Table 1
Test problems

No.	Dim.	Problem name	No.	Dim.	Problem name
1	3	Helical valley function	11	4	Brown and Dennis function
2	6	Biggs exp6 function	12	2	Rosenbrock function
3	3	Gaussian function	13	10	Trigonometric function
4	2	Powell badly scaled function	14	10	Extended Rosenbrock function
5	3	Box three-dimensional function	15	4	Extended Powell singular function
6	8	Variably dimensioned function	16	2	Beale function
7	6	Watson function	17	4	Wood function
8	4	Penalty function I	18	7	Chebyquad function
9	4	Penalty function II	19	2	Freudenstein and Roth function
10	2	Brown badly scaled function			

Table 2
Numerical results for the BFGS, SR1 and Hoshino methods

No. of prob.	BFGS		SR1		Hoshino	
	y	$\hat{y}^{HU}$	y	$\hat{y}^{HU}$	y	$\hat{y}^{HU}$
1	28/33/31	25/32/29	25/31/28	25/38/32	32/37/34	31/34/32
2	36/48/43	34/41/38	30/41/35	28/44/35	46/50/49	45/54/52
3	2/5/4	2/5/4	2/5/4	2/5/4	2/5/4	2/5/4
4	63/84/70	60/79/65	71/94/81	49/66/59	69/80/73	58/80/69
5	22/34/28	20/32/27(1)	24/40/33	20/36/29	24/40/35	19/30/27(1)
6	19/20/20	14/15/15	19/20/20	13/14/14	19/21/21	14/15/15
7	33/36/34	29/34/31	28/38/33	25/40/34	37/40/38	34/37/35
8	12/13/13	9/11/11	14/17/17	9/11/11	13/15/15	9/11/11
9	8/9/9	8/9/9	9/12/11	8/9/9	10/12/12	8/9/9
10	10/23/17	12/20/15	12/17/13	11/16/13	13/18/15	12/16/13
11	20/85/22	18/46/20	15/28/17	13/26/15	21/39/24	20/31/22(1)
12	34/47/38	35/45/37	44/70/54	36/61/47	34/47/39	31/39/36
13	24/27/26	22/26/24(1)	23/29/25	25/29/27	27/28/27	27/28/27(1)
14	71/102/88	84/114/103	89/146/119	101/167/138	106/124/115	92/113/100
15	25/31/27	22/29/26	23/28/25	25/34/29	24/29/26	23/31/27
16	14/16/15	12/14/13(1)	14/18/15	11/16/14	14/17/16	14/17/15(1)
17	16/24/16	60/79/70^d	65/102/84^d	15/21/15	92/103/98^d	76/95/87^d
18	11/23/15	14/22/17	14/24/18	14/24/18	14/24/17	13/21/15
19	9/10/9	9/10/9(1)	8/10/8	8/9/8	10/17/10	9/13/9
No. of wins	2	13	5	10	1	16

of the following conditions is satisfied:

$$f_k - f_{k+1} \leqslant 10^{-8} \max\{1.0, |f_k|\} \quad \text{or} \quad \|g_k\| \leqslant 10^{-4}.$$

Since the aim of the experiments is to observe the improvements of the modified quasi-Newton updates over the usual quasi-Newton updates, we here present a comparison between the results for $\hat{y}_k$ and the results for y_k. In the vector $\hat{y}_k$, $u = y_k$ is chosen (i.e. $\hat{y}_k = \hat{y}_k^{\mathrm{HU}}$) and strategy (10) with $\varepsilon = 0.0001$ is used to ensure $\delta_k^{\mathrm{T}} \hat{y}_k > 0$ for the BFGS and Hoshino updates. Since the SR1 update does not preserve positive-definite updates, this strategy is not used. Instead, the commonly used strategy

$$H_{k+1} = H_k \quad \text{if } |(\delta_k - H_k y_k)^{\mathrm{T}} y_k| \leqslant \varepsilon \|\delta_k - H_k y_k\| \, \|y_k\|$$

is used to avoid the singularity of the matrix H_k for the SR1 update, but actually this situation does not occur in our experiments. We also implemented numerical calculations for $\hat{y}_k$ with different choices of the vector u, for example, $u = \delta_k$ and $u = g_k$. They also bring in some improvements, but the performance is not as good as the choice $u = y_k$. The results for $\hat{y}_k^{\mathrm{HU}}$ are given in Table 2.

In the table, we give in the entry *nitr*/*nf*/*ng* the number of iterations and the calls for function and gradient evaluations, respectively. The number in parentheses denotes the number of times that $\theta_k < (\varepsilon - 1)\delta_k^{\mathrm{T}} y_k$ occurs and strategy (10) is used. The superscript d denotes that the iteration terminates at a different local solution. We use $(nf + n \times ng)$ as a measure to compare relative efficiency of the methods. That is, between two results the one with smaller value of $(nf + n \times ng)$ is regarded as better for a particular test problem. This comparison is given in the last row between the usual quasi-Newton method (y_k) and the modified quasi-Newton method ($\hat{y}_k^{\rm HU}$), where “No. of wins” denotes the number of problems for which one method outperforms another one.

It can be seen from the table that the modified quasi-Newton methods with $u = y_k$ in $\hat{y}_k$ give improvements over the usual quasi-Newton methods on most test problems and that the whole comparison favours the modified quasi-Newton methods. The results also show that the analyses in Section 4 is correct, that is, the case $\delta_k^{\rm T} y_k + \theta_k < 0$ rarely occurs, and when it happens, the point is generally far away from the local solution.

References

[1] M.C. Biggs, Minimization algorithms making use of non-quadratic properties of the objective function, J. Inst. Math. Appl. 8 (1971) 315–327.
[2] C.G. Broyden, A class of methods for solving nonlinear simultaneous equations, Math. Comp. 19 (1965) 577–593.
[3] R.H. Byrd, J. Nocedal, Yaxian Yuan, Global convergence of a class of variable metric algorithms, SIAM J. Numer. Anal. 24 (1987) 1171–1190.
[4] A.R. Conn, N. Gould, Ph.L. Toint, Convergence of quasi-Newton matrices generated by the symmetric rank one update, Math. Programming 50 (1991) 177–195.
[5] W.C. Davidon, Conic approximations and collinear scalings for optimizers, SIAM J. Numer. Anal. 17 (1980) 268–281.
[6] J.E. Dennis, D.M. Gay, R.E. Welsch, Algorithm 573 NL2SOL — an adaptive nonlinear least-squares algorithm [e4], ACM Trans. Math. Software 7 (1981) 369–383.
[7] J.E. Dennis, J.J. Moré, A characterization of superlinear convergence and its application to quasi-Newton methods, Math. Comp. 28 (1974) 549–560.
[8] J.E. Dennis, J.J. Moré, Quasi-Newton methods, motivation and theory, SIAM Review 19 (1977) 46–89.
[9] L.C.W. Dixon, Quasi-Newton algorithms generate identical points, Math. Programming 2 (1972) 383–387.
[10] R. Fletcher, Practical Methods of Optimization, Vol. 1, Unconstrained Optimization, Wiley, Chichester, 1980.
[11] J.A. Ford, I.A. Moghrabi, Multi-step quasi-Newton methods for optimization, J. Comput. Appl. Math. 50 (1994) 305–323.
[12] H.Y. Huang, Unified approach to quadratically convergent algorithms for function minimization, J. Optim. Theory Appl. 5 (1970) 405–423.
[13] J.J. Moré, B.S. Garbow, K.E. Hillstrom, Testing unconstrained optimization software, ACM Trans. Math. Software 7 (1981) 17–41.
[14] M.J.D. Powell, Some global convergence properties of a variable metric algorithm for minimization without exact line searches, in: R.W. Cottle, C.E. Lemke (Eds.), Nonlinear Programming, SIAM-AMS Proceedings Vol. IX, SIAM, Philadelphia, 1976.
[15] R.B. Schnabel, T. Chow, Tensor methods for unconstrained optimization using second derivatives, SIAM J. Optim. 1 (1991) 293–315.
[16] E. Spedicato, A class of rank-one positive definite quasi-Newton updates for unconstrained optimization, Math. Oper. Stat. Ser. A, Optim. 14 (1983) 61–70.
[17] C.X. Xu, J.Z. Zhang, Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations, Technical Report, Department of Mathematics, City University of Hong Kong, 1999.
[18] Y. Yuan, A modified BFGS algorithm for unconstrained optimization, IMA J. Numer. Anal. 11 (1991) 325–332.
[19] Y. Yuan, R. Byrd, Non-quasi-Newton updates for unconstrained optimization, J. Comput. Math. 13 (1995) 95–107.
[20] J.Z. Zhang, N.Y. Deng, L.H. Chen, New quasi-Newton equation and related methods for unconstrained optimization, J. Optim. Theory Appl. 102 (1999) 147–167.