Journal of Computational Mathematics, Vol.13, No.2, 1995, 95-107.

NON-QUASI-NEWTON UPDATES FOR UNCONSTRAINED OPTIMIZATION[*1)]

Ya-xiang Yuan
(*Computing Centre, Academia Sinica, Beijing, China*)

Richard H. Byrd
(*Department of Computer Science, University of Colorado at Boulder, USA*)

Abstract

In this report we present some new numerical methods for unconstrained optimization. These methods apply update formulae that do not satisfy the quasi-Newton equation. We derive these new formulae by considering different techniques of approximating the objective function. Theoretical analyses are given to show the advantages of using non-quasi-Newton updates. Under mild conditions we prove that our new update formulae preserve global convergence properties. Numerical results are also presented.

1. Introduction

Unconstrained optimization is to minimize a nonlinear function $f(x)$ in a finite dimensional space, that is

$$\min_{x\in R^n} \quad f(x) \quad . \tag{1.1}$$

Newton's method for problem (1.1) is iterative and at the $k-$th iteration a current approximation solution x_k is available. The Newton step at the $k-$th iteration is

$$d_k = -(\nabla^2 f(x_k))^{-1}\nabla f(x_k) \quad . \tag{1.2}$$

One advantage of Newton's method is that it convergence quadratically. Assume x^* is a stationary point of (1.1) at which $\nabla^2 f(x^*)$ is non-singular. Then for x_k sufficiently close to x^* we have that

$$||x_k + d_k - x^*|| = O(||x_k - x^*||^2) \quad . \tag{1.3}$$

However Newton's method also has some disadvantages. Firstly the Hessian $\nabla^2 f(x_k)$ may be singular, in that case the Newton step (1.2) is not well defined. Secondly when $\nabla^2 f(x_k)$ is not positive definite the Newton step d_k may not necessarily be a descent

[*] Received March 28, 1994.

[1)] This work was supported by the State Key Project "large scale scientific computing" of China.

direction of the objective function. Thirdly the calculation of the Hessian $\nabla^2 f(x_k)$ may be very expensive especially for large scale problems, not to mention that for some problems the Hessian of $f(x)$ is not available.

Quasi-Newton methods are a class of numerical methods that are similar to Newton's method except that the Hessian $(\nabla^2 f(x_k))^{-1}$ is replaced by an $n \times n$ symmetric matrix H_k which satisfies the "quasi-Newton" equation

$$H_k y_{k-1} = s_{k-1} \tag{1.4}$$

where

$$s_{k-1} = x_k - x_{k-1} = \alpha_{k-1} d_{k-1} \tag{1.5}$$

$$y_{k-1} = \nabla f(x_k) - \nabla f(x_{k-1}) \quad , \tag{1.6}$$

and $\alpha_{k-1} > 0$ is a step-length which satisfies some line search conditions. Assume H_k is nonsingular, we define $B_k = (H_k)^{-1}$. It is easy to see that the "quasi-Newton step"

$$d_k = -H_k \nabla f(x_k) \tag{1.7}$$

is a stationary point of the following problem:

$$\min_{d \in R^n} \quad \phi_k(d) = f(x_k) + d^T \nabla f(x_k) + \frac{1}{2} d^T B_k d \tag{1.8}$$

which is an approximation to problem (1.1) near the current iterate x_k, since $\phi_k(d) \simeq f(x_k + d)$ for small d. In fact, the definition of $\phi_k(.)$ in (1.8) implies that

$$\phi_k(0) = f(x_k), \tag{1.9}$$

$$\nabla \phi_k(0) = \nabla f(x_k), \tag{1.10}$$

and the quasi-Newton condition (1.4) is equivalent to

$$\nabla \phi_k(x_{k-1} - x_k) = \nabla f(x_{k-1}) \ . \tag{1.11}$$

Thus, $\phi_k(x - x_k)$ is a quadratic interpolation of $f(x)$ at x_k and x_{k-1}, satisfying conditions (1.9)-(1.11). The matrix B_k (or H_k) can be updated so that the quasi-Newton equation is satisfied. One well known update formula is the BFGS formula which updates B_{k+1} from B_k, s_k and y_k in the following way:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} \quad . \tag{1.12}$$

In Yuan (1991), approximate function $\phi_k(d)$ in (1.8) is required to satisfy the interpolation condition

$$\phi_k(x_{k-1} - x_k) = f(x_{k-1}) \ , \tag{1.13}$$

instead of (1.11). This change was inspired from the fact that for one dimension problem, using (1.13) gives a slightly faster local convergence if we assume $\alpha_k = 1$ for all k. Equation (1.13) can be rewritten as

$$s_{k-1}^T B_k s_{k-1} = 2[f(x_{k-1}) - f(x_k) + s_{k-1}^T \nabla f(x_k)] \quad . \tag{1.14}$$

In order to satisfy (1.14), the BFGS formula is modified as follows:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + t_k \frac{y_k y_k^T}{s_k^T y_k} \quad , \tag{1.15}$$

where

$$t_k = \frac{2}{s_k^T y_k}[f(x_k) - f(x_{k+1}) + s_k^T \nabla f(x_{k+1})] \ . \tag{1.16}$$

Assume that B_k is positive definite and that $s_k^T y_k > 0$, B_{k+1} defined by (1.15) is positive definite if and only if $t_k > 0$. The inequality $t_k > 0$ is trivial if $f(x)$ is strictly convex, and it is also true if the step-length α_k is chosen by an exact line search which requires $s_k^T \nabla f(x_{k+1}) = 0$. For an uniformly convex function, it can be easily shown that there exists a constant $\delta > 0$ such that $t_k \in [\delta, 2]$ for all k, and consequently global convergence of the modified method can be proved by slightly modifying the global convergence proof of the BFGS method for convex functions with inexact line searches, which was given by Powell (1976). However, for a general nonlinear function $f(x)$, inexact line searches do not imply the positivity of t_k, hence Yuan (1991) truncated t_k to the interval $[0.01, 100]$, and showed that the global convergence of the BFGS algorithm is preserved for convex functions.

If the objective function $f(x)$ is cubic along the line segment between x_{k-1} and x_k then we have the following relation

$$s_{k-1}^T \nabla^2 f(x_k) s_{k-1} = 4 s_{k-1}^T \nabla f(x_k) + 2 s_{k-1}^T \nabla f(x_{k-1}) - 6(f(x_k) - f(x_{k-1})) \ , \tag{1.17}$$

by considering the Hermit interpolation on the line between x_k and x_{k-1}. Hence it is reasonable to require that the new approximate Hessian satisfy condition:

$$s_{k-1}^T B_k s_{k-1} = 4 s_{k-1}^T \nabla f(x_k) + 2 s_{k-1}^T \nabla f(x_{k-1}) - 6(f(x_k) - f(x_{k-1})) \tag{1.18}$$

instead of (1.14). Biggs (1971, 1973) gives the inverse update of (1.15) with the value t_k so chosen that (1.18) holds. For one-dimensional problems, Wang and Yuan (1992) showed that (1.18) without line searches (that is, $\alpha_k = 1$ for all k) implies R-quadratic convergence, and except some special cases (1.18) also gives Q-quadratic convergence. It is well known that the convergence rate of the secant method is only $(1 + \sqrt{5})/2$ which is approximately $1.618 < 2$.

There are other methods that use information of function values of the objective function to construct the subproblem. The conic model method requires the approximate function $\phi_k(d)$ to satisfy conditions (1.9)-(1.11) and (1.13), and $\phi_k(d)$ is a conic function, instead of a quadratic function. More details on conic methods can be found in Davidon (1980) and Schnabel (1983). The quasi-Newton method without derivatives of Greenstadt (1972) uses a quadratic subproblem, where the approximate Hessian is updated based on function values only. Our approach is to consider quadratic subproblem (1.8) where the approximate Hessian B_k is updated based on some approximation to the second order curvature along the last search direction. For example, update the approximate Hessian satisfying relation (1.18) implies that we use the cubic approximation to get an approximation to the second order curvature, which in turn is used to construct the quadratic subproblem.

In this report, we study update formulae that satisfy the condition

$$s_k B_{k+1} s_k = \rho_k \tag{1.19}$$

where we assume that $\rho_k > 0$ is some kind approximate value of the second curvature of $f(x)$ between x_k and x_{k+1}. For example, ρ_{k-1} may be the value of the right hand side of (1.14) or (1.18). Except for $\rho_k = s_k^T y_k$, B_{k+1} can not satisfy the quasi-Newton equation if condition (1.19) holds. However, as in the Broyden family update formulae, we also consider update formulae for B_{k+1} that depend on B_k, $B_k s_k$, and y_k. In the next section, we derive a class of update formulae by minimizing certain norm of the residual of the quasi-Newton equation. In section 3, we study a specific method of the family. We show the global convergence of the method under certain conditions.

2. A Class of Update Formulae

We consider update formulae that have the following form:

$$\begin{aligned} B_{k+1}(\beta,\gamma,\tau) =& B_k + \beta\frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \gamma\frac{y_k y_k^T}{s_k^T y_k} \\ &+ \tau\left(\frac{B_k s_k}{s_k^T B_k s_k} - \frac{y_k}{s_k^T y_k}\right)\left(\frac{B_k s_k}{s_k^T B_k s_k} - \frac{y_k}{s_k^T y_k}\right)^T \quad . \end{aligned} \tag{2.1}$$

where β, γ and τ are parameters. For simplicity, we shall use the notations

$$u_k = y_k / s_k^T y_k \quad , \qquad v_k = -B_k s_k / s_k^T B_k s_k \quad . \tag{2.2}$$

Let $W \in \Re^{n\times n}$ be a positive definite matrix, we require that matrix B_{k+1} solves the least norm problem:

$$\min ||B_{k+1} s_k - y_k||_W \tag{2.3}$$

subject to

$$s_k^T B_{k+1} s_k = \rho_k \tag{2.4}$$

where $\rho_k > 0$ being some approximation of the second order curvature $s_k^T \nabla f(x_{k+1}) s_k$, and $||v||_W = \sqrt{v^T W v}$. Replacing B_{k+1} in (2.3)–(2.4) by formula (2.1), we have that

$$\min ||B_k s_k - y_k - \beta s_k^T B_k s_k v_k + \gamma s_k^T y_k u_k||_W^2 \tag{2.5}$$

subject to

$$s_k^T B_k s_k + \beta s_k^T B_k s_k + \gamma s_k^T y_k = \rho_k \ . \tag{2.6}$$

If $u_k + v_k = 0$, it is straightforward that solutions of problem (2.5)-(2.6) are any β, γ pairs that satisfy $\beta s_k^T B_k s_k + \gamma s_k^T y_k = \rho_k - s_k^T B_k s_k$, all the solutions will give the same update formula which can be written as

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \rho_k \frac{y_k y_k^T}{(s_k^T y_k)^2} \ . \tag{2.7}$$

Assume that $u_k + v_k \neq 0$, One can easily find that the unique solution of (2.5)-(2.6) is given by

$$\beta_k = (v_k + u_k)^T W[(\rho_k - s_k^T B_k s_k - s_k^T y_k)u_k - s_k^T B_k s_k v_k]/(s_k^T B_k s_k ||v_k + u_k||_W^2), \tag{2.8}$$

$$\gamma_k = (v_k + u_k)^T W(\rho_k v_k + s_k^T y_k u_k)/(s_k^T y_k ||v_k + u_k||_W^2). \tag{2.9}$$

Thus, we have obtained a class of update formulae

$$\begin{aligned} B_{k+1}(\tau) = & B_k + \frac{(v_k + u_k)^T W[(\rho_k - s_k^T B_k s_k - s_k^T y_k)u_k - s_k^T B_k s_k v_k]}{||v_k + u_k||_W^2} v_k v_k^T \\ & + \frac{(v_k + u_k)^T W(\rho_k v_k + s_k^T y_k u_k)}{||v_k + u_k||_W^2} u_k u_k^T + \tau(u_k + v_k)(u_k + v_k)^T. \end{aligned} \tag{2.10}$$

We can rewrite (2.10) in the following form:

$$\begin{aligned} B_{k+1}(\tau) = & B_k - s_k^T B_k s_k v_k v_k^T + \rho_k u_k u_k^T \\ & + \sigma_k [v_k v_k^T - u_k u_k^T] + \tau(u_k + v_k)(u_k + v_k)^T \end{aligned} \tag{2.11}$$

where $\sigma_k = (\rho_k - s_k^T y_k)(v_k + u_k)^T W u_k / ||v_k + u_k||_W^2$. It is easy to see that update formulae (2.11) are the Broyden family if $\rho_k = s_k^T y_k$.

The following result is obvious.

Lemma 2.1. *Assume that $s_k^T y_k > 0$, $\rho_k > 0$ and B_k is positive definite, then $B_{k+1}(0)$ has at least $n - 1$ positive eigenvalues.*

Proof. We can write $B_{k+1}(0)$ into two parts, namely

$$B_{k+1}(0) = \bar{B}_{k+1} + \sigma_k [v_k v_k^T - u_k u_k^T] \tag{2.12}$$

where $\bar{B}_{k+1}$ is the same matrix as given in the right hand side of (2.7). It is easy to see that $\bar{B}_{k+1}$ is positive definite and that the matrix $\sigma_k [v_k v_k^T - u_k u_k^T]$ has at most one negative eigenvalue. Therefore $B_{k+1}(0)$ has at least $n - 1$ positive eigenvalues.

From the above lemma, it can be shown that

Corollary 2.2. *Under the conditions of Lemma 2.1, the matrix $B_{k+1}(\tau)$ is positive definite if and only if*

$$\det[B_{k+1}(\eta)] > 0 \tag{2.13}$$

for all $\eta \geq \tau$.

Proof. If $B_{k+1}(\tau)$ is positive definite, then $B_{k+1}(\eta)$ is also positive definite for all $\eta \geq \tau$. Hence inequality (2.13) holds.

Assume inequality (2.13) holds for all $\eta \geq \tau$. Let $\xi = \max[0, \tau]$, it is easy to see that $\det[B_{k+1}(\xi)] > 0$. Because $\xi \geq 0$, due to Lemma 2.1 $B_{k+1}(\xi)$ has at least $n-1$ positive eigenvalues. Therefore $B_{k+1}(\xi)$ must be positive definite. Consequently $B_{k+1}(\tau)$ is also positive definite because $\det[B_{k+1}(t)] > 0$ for all $t \in [\tau, \xi]$ and because the eigenvalues of $B_{k+1}(t)$ are continuous functions of t.

As we are interested in update formulae that give a positive definite B_{k+1}, we now calculate the determinate of $B_{k+1}(\tau)$. First, in general we have

Lemma 2.3. *For any symmetric matrix* $B \in \Re^{n\times n}$, *and vectors* $v, u \in \Re^n$, *if* B^{-1} *exists and* B_+ *is defined by*

$$B_+ = B + \beta vv^T + \gamma uu^T + \tau(v+u)(v+u)^T \tag{2.14}$$

where β, γ *and* τ *are any real numbers, then the following relation*

$$\frac{\det(B_+)}{\det(B)} = [1+\beta v^T B^{-1} v + \gamma u^T B^{-1} u + \beta\gamma\chi + \tau((v+u)^T B^{-1}(v+u) + (\beta+\gamma)\chi)] \tag{2.15}$$

holds, where $\chi = v^T B^{-1} v u^T B^{-1} u - (v^T B^{-1} u)^2$. *Furthermore, if* $\det(B_+) \neq 0$ *we have that*

$$B_+^{-1} = B^{-1} + \bar\beta B^{-1} vv^T B^{-1} + \bar\gamma B^{-1} uu^T B^{-1} + \bar\tau B^{-1}(v+u)(v+u)^T B^{-1} \tag{2.16}$$

where

$$\bar\beta = [-\beta - (\beta\gamma + (\beta+\gamma)\tau) u^T B^{-1}(u+v)] \det(B)/\det(B_+) \tag{2.17}$$

$$\bar\gamma = [-\gamma - (\beta\gamma + (\beta+\gamma)\tau) v^T B^{-1}(u+v)] \det(B)/\det(B_+) \tag{2.18}$$

$$\bar\tau = [-\tau + (\beta\gamma + (\beta+\gamma)\tau) v^T B^{-1} u] \det(B)/\det(B_+). \tag{2.19}$$

Proof. Because B^{-1} exists and because the nonzero eigenvalues of UV^T and $V^T U$ are the same for any $U, V \in \Re^{n\times m}$, it follows that

$$\begin{aligned}
\det(B_+) &= \det(B)\det(\mathrm{I} + \beta B^{-1/2} vv^T B^{-1/2} + \gamma B^{-1/2} uu^T B^{-1/2} \\
&\qquad + \tau B^{-1/2}(v+u)(v+u)^T B^{-1/2}) \\
&= \det(B)\det(\mathrm{I} + [B^{-1/2}v \quad B^{-1/2}u] \begin{bmatrix} \beta+\tau & \tau \\ \tau & \gamma+\tau \end{bmatrix} \begin{bmatrix} v^T B^{-1/2} \\ u^T B^{-1/2} \end{bmatrix}) \\
&= \det(B)\det\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} \beta+\tau & \tau \\ \tau & \gamma+\tau \end{bmatrix} \begin{bmatrix} v^T B^{-1} v & v^T B^{-1} u \\ v^T B^{-1} u & u^T B^{-1} u \end{bmatrix}\right) \\
&= \det(B)[(1 + (\beta+\tau) v^T B^{-1} v + \tau v^T B^{-1} u)(1 + (\gamma+\tau) u^T B^{-1} u + \tau v^T B^{-1} u) \\
&\qquad - (\tau v^T B^{-1} v + (\gamma+\tau) v^T B^{-1} u)(\tau u^T B^{-1} u + (\beta+\tau) v^T B^{-1} u)] \quad ,
\end{aligned} \tag{2.20}$$

which gives equation (2.15). When $\det(B_+) \neq 0$, the relation (2.16) can be proved by direct calculations.

Using equations (2.15) and (2.11), we have that

$$\frac{\det[B_{k+1}(\tau)]}{\det(B_k)} = [\rho_k + (v_k + u_k)^T B_k^{-1}(v_k + u_k)[\tau\rho_k + \sigma_k(\rho_k - \sigma_k)]]/s_k^T B_k s_k \quad . \tag{2.21}$$

Hence it is easy to see that $\det[B_{k+1}(\tau)] > 0$ if and only if $\tau > \tau_0$, where

$$\tau_0 = -\sigma_k(1 - \sigma_k/\rho_k) - 1/(v_k + u_k)^T B_k^{-1}(v_k + u_k) \quad . \tag{2.22}$$

Consequently, from Corollary 2.2, $B_{k+1}(\tau)$ is positive definite if and only if $\tau > \tau_0$.

The following function

$$\Psi_1(\tau) = \mathrm{Tr}[B_{k+1}(\tau)] - \log[\det(B_{k+1}(\tau))] \tag{2.23}$$

is given by Byrd and Nocedal (1989), where Tr(.) denotes the trace of a matrix. They use this function to "measure" the distance of B_{k+1} from the identity matrix. In order to force the update formula having certain smallest change property, it is desirable to have the matrix $B_k^{-1/2} B_{k+1} B_k^{1/2}$ being as close to the identity matrix as possible. Therefore we consider the following scaled function:

$$\Psi_2(\tau) = \mathrm{Tr}[B_k^{-1/2} B_{k+1}(\tau) B_k^{-1/2}] - \log[\det(B_k^{-1/2} B_{k+1}(\tau) B_k^{-1/2})] \tag{2.24}$$

which is similar to (2.23) except that B_{k+1} is replaced by $B_k^{-1/2} B_{k+1} B_k^{-1/2}$. In the case when $\rho_k = s_k^T y_k$, function (2.24) was used by Fletcher(1990) to show that B_{k+1} updated by the BFGS update formula minimizes this function. It is straightforward to calculate

$$\begin{aligned}\Psi_2(\tau) = n + \sigma_k/s_k^T B_k s_k - 1 + (\rho_k - \sigma_k) y_k^T B_k^{-1} y_k/(s_k^T y_k)^2 + \tau(v_k + u_k)^T B_k^{-1}(v_k + u_k) \\ - \log[\rho_k + (v_k + u_k)^T B_k^{-1}(v_k + u_k)(\tau\rho_k + (\rho_k - \sigma_k)\sigma_k)] + \log[s_k^T B_k s_k]\end{aligned} \tag{2.25}$$

Let $\tau_1 \in [\tau_0, +\infty]$ minimize $\Psi_2(\tau)$, we can easily find that

$$\tau_1 = -\sigma_k(1 - \sigma_k/\rho_k) = \tau_0 + 1/(v_k + u_k)^T B_k^{-1}(v_k + u_k) > \tau_o \quad , \tag{2.26}$$

which yields the following update formula

$$\begin{aligned}B_{k+1}(\tau_1) = &B_k - (s_k^T B_k s_k - \sigma_k^2/\rho_k) v_k v_k^T \\ &+ \rho_k(1 - \sigma_k/\rho_k)^2 u_k u_k^T - \sigma_k(1 - \sigma_k/\rho_k)[v_k u_k^T + u_k v_k^T] \quad .\end{aligned} \tag{2.27}$$

In the next section, under certain conditions we prove the global convergence of the method if update formula (2.27) is used.

3. Global Convergence Analyses

In this section, we consider the update formula (2.27). We study two cases when the matrix $W = I$ and $W = B_k^{-1}$.

From (2.27), it is easy to see that

$$\mathrm{Tr}(B_{k+1}(\tau_1)) = \mathrm{Tr}(B_k) - s_k^T B_k s_k ||v_k||_2^2 + \rho_k ||u_k||_2^2 + \sigma_k^2 ||v_k + u_k||_2^2/\rho_k - 2\sigma_k (v_k + u_k)^T u_k \tag{3.1}$$

First we study the case when $W = I$, which gives

$$\sigma_k = (\rho_k - s_k^T y_k)(v_k + u_k)^T u_k / ||v_k + u_k||_2^2 \tag{3.2}$$

and (2.27) can be rewritten as

$$\begin{aligned} B_{k+1} = B_k - s_k^T B_k s_k v_k v_k^T + \rho_k u_k u_k^T \\ + [(v_k + u_k) w_k^T + w_k (v_k + u_k)^T](\rho_k - s_k^T y_k)(v_k + u_k)^T u_k / ||v_k + u_k||_2^2 \end{aligned} \tag{3.3}$$

where

$$w_k = -u_k + (v_k + u_k)(\rho_k - s_k^T y_k)(v_k + u_k)^T u_k / (2\rho_k ||v_k + u_k||_2^2). \tag{3.4}$$

It follows from (3.1) and (3.2) that

$$\begin{aligned} \mathrm{Tr}(B_{k+1}) &= \mathrm{Tr}(B_k) - s_k^T B_k s_k ||v_k||_2^2 + \rho_k ||u_k||_2^2 - \sigma_k(\rho_k + s_k^T y_k)(v_k + u_k)^T u_k / \rho_k \\ &= \mathrm{Tr}(B_k) - s_k^T B_k s_k ||v_k||_2^2 + \rho_k ||u_k||_2^2 \\ &+ [(s_k^T y_k)^2 - \rho_k^2][(v_k + u_k)^T u_k]^2 / \rho_k ||v_k + u_k||_2^2 \ . \end{aligned} \tag{3.5}$$

Now the last term of equation (3.5) is non-positive if $s_k^T y_k < \rho_k$ and it is bounded above by $[(s_k^T y_k)^2 - \rho_k^2]||u_k||_2^2$ if $s_k^T y_k \geq \rho_k$. Thus relation (3.5) implies the inequality

$$\mathrm{Tr}(B_{k+1}) \leq \mathrm{Tr}(B_k) - s_k^T B_k s_k ||v_k||^2 + \rho_k \max[1, (s_k^T y_k)^2/\rho_k^2] ||u_k||_2^2. \tag{3.6}$$

We assume that there exist positive constants ω_1 and ω_2 such that

$$\omega_1 s_k^T y_k \leq \rho_k \leq \omega_2 s_k^T y_k \tag{3.7}$$

which means that our approximation of second order curvature of $f(x)$ between x_k and x_{k+1} is not far away from the approximation based on finite difference of first order directional derivatives. From (3.6) and (3.7), it follows that there exists a positive constant $\bar{c}$ such that

$$Tr(B_{k+1}) \leq Tr(B_k) - s_k^T B_k s_k ||v_k||^2 + \bar{c} s_k^T y_k ||u_k||_2^2 \ . \tag{3.8}$$

From (2.26) and (2.21), we have that $\det(B_{k+1}) = \det(B_k)\rho_k / s_k^T B_k s_k$ which implies that

$$\omega_1 s_k^T y_k / s_k^T B_k s_k \leq \det(B_{k+1})/\det(B_k) \leq \omega_2 s_k^T y_k / s_k^T B_k s_k \ . \tag{3.9}$$

Consequently, using relations (3.8) and (3.9), we can show the global convergence of the method by slightly modifying the global convergence of the BFGS method given by Powell (1976), or by applying the techniques given by Byrd, Nocedal and Yuan (1987). Hence we can state our convergence result as follows.

Theorem 3.1. *Assume that $f(x)$ is convex and twice continuously differential and that at the k-th iteration the search direction is $d_k = -B_k^{-1}\nabla f(x_k)$ and the stepsize α_k is chosen by inexact line search that satisfies*

$$f(x_k + \alpha_k d_k) \le f(x_k) + c_1 \alpha_k d_k^T \nabla f(x_k), \tag{3.10}$$

$$d_k^T \nabla f(x_k + \alpha_k d_k) \ge c_2 d_k^T \nabla f(x_k) \tag{3.11}$$

where $c_1 < c_2 < 1$ are two positive constants, if B_1 is symmetric and positive definite and B_{k+1} is updated by (3.3), if (3.7) holds for all k, and if the sequence $\{x_k, k = 1, 2, 3...\}$ is bounded, then x_k converges in the sense that

$$\lim_{k\to\infty} \quad ||\nabla f(x_k)||_2 = 0 \quad . \tag{3.12}$$

Now we consider the case that $W = B_k^{-1}$. It follows that

$$\sigma_k = \rho_k - s_k^T y_k \quad . \tag{3.13}$$

Hence update formula (2.27) becomes

$$\begin{aligned} B_{k+1} =& B_k - (s_k^T B_k s_k - (\rho_k - s_k^T y_k)^2/\rho_k) v_k v_k^T \\ &+ \rho_k^{-1}(s_k^T y_k)^2 u_k u_k^T - \rho_k^{-1}(\rho_k - s_k^T y_k) s_k^T y_k [v_k u_k^T + u_k v_k^T] \quad . \end{aligned} \tag{3.14}$$

For update formula (3.14), direct application of Powell's techniques to prove the global convergence requires (3.7) and that

$$(\rho_k - s_k^T y_k)^2/\rho_k \le \omega_3 s_k^T B_k s_k \tag{3.15}$$

holds for some constant $\omega_3 < 1$, which is equivalent to

$$s_k^T y_k/\omega_4 \le \rho_k \le s_k^T y_k \omega_4 \tag{3.16}$$

where

$$\omega_4 = 1 + 0.5\omega_3 s_k^T B_k s_k / s_k^T y_k + \sqrt{\omega_3 s_k^T B_k s_k (1 + 0.25\omega_3 s_k^T B_k s_k / s_k^T y_k)/s_k^T y_k} \ . \tag{3.17}$$

Theorem 3.2. *Under the conditions of Theorem* 3.1*, if update formula* (3.14) *is used instead of* (3.3)*, and if* (3.15) *holds for all k, then* (3.12) *is true.*

Update formulae (3.3) and (3.14) can be rewritten as

$$\begin{aligned} B_{k+1} =& B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} \\ &+ (\rho_k - s_k^T y_k)[u_k u_k^T + ((u_k + v_k) w_k^T + w_k (u_k + v_k)^T) \frac{(u_k + v_k)^T u_k}{||u_k + v_k||_2^2} \end{aligned} \tag{3.18}$$

and

$$\begin{aligned} B_{k+1} = & B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} \\ & + (\rho_k - s_k^T y_k)[v_k v_k^T - \frac{s_k^T y_k}{\rho_k}(u_k + v_k)(u_k + v_k)^T] \end{aligned} \tag{3.19}$$

respectively, where w_k is defined by (3.4). Assume that x_k converges to a solution x_* at which $\nabla^2 f(x_*)$ is positive definite and that B_k and B_k^{-1} are bounded, either of (3.18) and (3.19) ensures that

$$B_{k+1} = B_{k+1}^{BFGS} + O(|\rho_k - s_k^T y_k|/||s_k||_2^2) = B_{k+1}^{BFGS} + O(||s_k||_2) \tag{3.20}$$

if the objective function $f(x)$ is three times differentiable near x_*. B_{k+1}^{BFGS} is the BFGS update matrix which is defined by

$$B_{k+1}^{BFGS} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} \quad . \tag{3.21}$$

From (3.20), we can apply techniques of Byrd and Nocedal (1989), Dennis and Moré (1977) and Powell (1976) to prove the local Q-superlinear convergence of Algorithm 4.1 when update formula (3.3) or (3.14) is used. A more detailed study on the local superlinear convergence of (3.3) and (3.14), and that of a modified BFGS method given by Yuan (1991) will be given in a forthcoming report.

4. Algorithms and Numerical Results

In this section, we state an algorithm that applies update formulae (3.3) and (3.14) given in section 3. For either update formula, we run a collection of standard test problems, and present numerical results.

First, our algorithm is given as follows.

Algorithm 4.1.

Step 1. Given $x_1 \in \Re^n$, $B_1 \in \Re^{n\times n}$ symmetric and positive definite,
$0 < c_1 < c_2 < 1, c_1 < 1/2$, $\epsilon > 0$ very small and $k = 1$;
Step 2. Calculate $g_k = \nabla f(x_k)$, if $||g_k||_\infty < \epsilon$ then stop;
Step 3. $d_k = -B_k^{-1} g_k$, calculate $\alpha_k > 0$ such that (3.10)-(3.11) hold;
Step 4. $x_{k+1} = x_k + \alpha_k d_k$, $s_k = x_{k+1} - x_k$, $y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$;
Step 5. Compute ρ_k, and update B_{k+1};
Step 6. k:=k+1 and go to Step 2.

Our algorithm is exactly the same as a quasi-Newton method except the updating of the approximate Hessian B_k. We have not specified the update formula in the algorithm, which can be either (3.3) or (3.14). The step-length α_k in Step 3 of the algorithm is calculated by quadratic and cubic interpolations and extrapolations with bracketing techniques (more details can be found in Fletcher, 1987). For the calculation of ρ_k, we let

$$\rho_k = 4s_k^T \nabla f(x_{k+1}) + 2s_k^T \nabla f(x_k) - 6(f(x_{k+1}) - f(x_k)) \tag{4.1}$$

which is (1.18) with k replaced by $k+1$, and it is an approximate second order curvature based on the cubic interpolation. As B_{k+1} is updated by (3.3) or (3.14), we require the positiveness of ρ_k to ensure B_{k+1} positive definite because of relation (2.4). However, it is possible that ρ_k defined by (4.1) is negative. This can happen even for strictly convex objective functions. For example, if we let $f(x)$ to be the one dimensional function x^4, and if we let $x_k = -1$ and $x_{k+1} = 0$, formula (4.1) gives $\rho_k = -2 < 0$. Therefore, in our algorithm if ρ_k computed by (4.1) does not satisfy inequalities (3.7), we truncate ρ_k in the following way:

$$\rho_k = \begin{cases} \omega_1 s_k^T y_k & \text{if } \rho_k < \omega_1 s_k^T y_k, \\ \omega_2 s_k^T y_k & \text{if } \rho_k > \omega_2 s_k^T y_k. \end{cases} \tag{4.2}$$

For convex objective functions $f(x)$, the second inequality of (3.7) is always true if $\omega_2 \geq \frac{4}{3}$, as it is easy to show that

$$\begin{aligned}
\rho_k &= 4s_k^T \nabla f(x_{k+1}) + 2s_k^T \nabla f(x_k) - 6(f(x_{k+1}) - f(x_k)) \\
&= 4s_k^T \nabla f(x_{k+1}) + 2s_k^T \nabla f(x_k) - 6\int_0^1 s_k^T \nabla f(x_k + ts_k)dt \\
&= 4\int_0^1 s_k^T[\nabla f(x_{k+1}) - \nabla f(x_k + ts_k)]dt - 2\int_0^1 s_k^T[\nabla f(x_k + ts_k) - \nabla f(x_k)]dt \\
&= \int_0^1 \left(\int_t^1 4s_k^T \nabla^2 f(x_k + us_k)s_k du\right) dt - \int_0^1 \left(\int_0^t 2s_k^T \nabla^2 f(x_k + us_k)s_k du\right) dt \\
&= \int_0^1 \left(\int_0^u 4s_k^T \nabla^2 f(x_k + us_k)s_k dt\right) du - \int_0^1 \left(\int_u^1 2s_k^T \nabla^2 f(x_k + us_k)s_k dt\right) du \\
&= \int_0^1 (6u - 2)s_k^T \nabla^2 f(x_k + us_k)s_k du \leq \frac{4}{3} s_k^T y_k \ .
\end{aligned} \tag{4.3}$$

And one can also easily see from (4.3) that $\rho_k = s_k^T \nabla^2 f(x_{k+1})s_k$ if $f(x)$ is cubic on the line segment between x_k and x_{k+1}, and $\rho_k = s_k^T y_k$ if $f(x)$ is quadratic.

The test problems we run are the 18 standard unconstrained optimization test problems suggested by Moré, Garbow and Hillstrom (1981). Our stopping criterion is $||\nabla f(x_k)||_\infty \leq 10^{-6}$. We let $c_1 = 0.01$, $c_2 = 0.9$. ρ_k calculated by (4.1) is truncated so that (3.7) is satisfied with $\omega_1 = 1/4$ and $\omega_2 = 4$. When update formula (3.14) is used, we also truncated ρ_k if needed so that (3.15) holds for $\omega_3 = 0.8$. The numerical results of Algorithm 4.1 with B_k updated by (3.3) and (3.14) are given in Table 1, where results of the BFGS algorithm are also presented. For each algorithm, the numbers in

Table 1. Numerical Results of Algorithm 4.1 and BFGS algorithm

	BFGS			(3.3)			(3.14)		
Problem	NI	NF	NG	NI	NF	NG	NI	NF	NG
1	28	40	30	25	36	26	30	42	32
2	36	45	40	37	42	39	35	45	40
3	3	5	4	3	5	4	3	5	4
4	159∗	221	170	148∗	197	169	158∗	215	180
5	19	32	27	13	23	19	17	27	24
6	18	23	19	12	16	13	12	16	13
7	68	76	69	70	79	71	70	79	71
8	27	43	39	45	65	56	51	75	64
9	8	12	10	8	12	10	7	11	9
10	10	23	15	12	24	17	11	24	16
11	25	42	29	21∗	39	25	21∗	40	26
12	30	44	35	34	51	39	24	37	29
13	46	51	49	44	53	50	44	47	46
14	128	179	133	125	172	130	122	171	128
15	102	123	105	54	73	55	71	95	72
16	13	17	14	14	20	15	13	18	15
17	81	114	87	70	97	77	78	109	87
18	21	35	23	22	32	24	22	35	23

columns NI, NF, and NG are numbers of iterations, function evaluations and gradient evaluations respectively. A star "*" indicates an usual stop due to very small reduction in the objective function, that is, $[f(x_k) - f(x_{k+1})]/(1 + |f(x_{k+1})|) < 10^{-16}$. In all 5 such cases, we found that the infinity norm of the gradient at the final point is less then $1.1 * 10^{-5}$. The numerical results indicate that both (3.3) and (3.14) give a slight improvement over the original BFGS method.

Acknowledgement. We would like to thank Professor John E. Dennis of Rice University who drew our attention to the work of Greenstadt (1972).

References

[1] M.C. Biggs, Minimization algorithms making use of non-quadratic properties of the objective function, *Journal of the Institute of Mathematics and Its Applications*, 8 (1971), 315–327.

[2] M.C. Biggs, A note on minimization algorithms which make use of non-quadratic properties of the objective function, *Journal of the Institute of Mathematics and Its Applications*, 12 (1973), 337–338.

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

[4] R. Byrd, J. Nocedal and Y. Yuan, Global convergence of a class of variable metric algorithms, *SIAM J. Numerical Analysis*, 4 (1987), 1171–1190.

[5] W.C. Davidon, Conic approximations and collinear scalings for optimizers, *SIAM J. Numer. Anal.,* 17 (1980), 268–281.

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

[7] R. Fletcher, Practical Methods of Optimizations, 2nd Edition, John Wiley and Sons, New York, 1987.

[8] J. Grenstadt, A quasi-Newton method with no derivatives, *Mathematics of Computations*, 26 (1972), 145–166.

[9] J.J. Moré, B.S. Garbow and K.E. Hillstrom, Testing unconstrained optimization software, *ACM Transactions on Mathematical Software*, 7 (1981), 17–41

[10] M.J.D. Powell, Some global convergence properties of a variable metric algorithm for minimization without exact line searches, in: R.W. Cottle and C. E. Lemke, eds., Nonlinear Programming SIAM-AMS Proceeding, Vol 9, SIAM Publications, 1976, 53–72.

[11] R.B. Schnabel, Conic methods for unconstrained optimization and tensor methods for nonlinear equations, in: A. Bachem, M. Grotschel and B. Korte eds., Mathematical Programming, The State of the Art, Springer-Verlag, Berlin, 1983, 417–438.

[12] H.J. Wang and Y. Yuan, A quadratic convergence method for one dimensional optimization, *Chinese J. of Operations Research*, 11 (1992) 1–10.

[13] Y. Yuan, A modified BFGS algorithm for unconstrained optimization, *IMA J. Numerical Analysis*, 11 (1991) 325–332.