SIAM J. NUMER. ANAL.
Vol 24 No 5 October 1987

015

GLOBAL CONVERGENCE OF A CLASS OF QUASI-NEWTON METHODS ON CONVEX PROBLEMS*

RICHARD H. BYRD†, JORGE NOCEDAL‡ AND YA-XIANG YUAN§

Abstract. We study the global convergence properties of the restricted Broyden class of quasi-Newton methods, when applied to a convex objective function. We assume that the line search satisfies a standard sufficient decrease condition and that the initial Hessian approximation is any positive definite matrix. We show global and superlinear convergence for this class of methods, except for DFP. This generalizes Powell's well-known result for the BFGS method. The analysis gives us insight into the properties of these algorithms; in particular it shows that DFP lacks a very desirable self-correcting property possessed by BFGS.

Key words. nonlinear optimization, quasi-Newton methods, minimization

AMS(MOS) subject classifications. 65, 49

1. Introduction. The convergence of quasi-Newton methods for unconstrained optimization has been the subject of much analysis. However there are some important gaps in what is understood about the behavior of these methods as implemented in practice. Powell (1976) has proved a global convergence result for the BFGS method using the line search commonly found in computer implementations. This result however has not been extended to the DFP method, nor to any other algorithm in Broyden's class. In this paper we show how to extend Powell's result to the restricted Broyden class, excluding DFP, using the same line search strategy and the same assumptions on the objective function. The analysis gives us considerable insight into the properties of quasi-Newton methods.

We will be considering the behavior of quasi-Newton methods from Broyden's class for the unconstrained optimization problem

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

This class consists of iterations of the form

$$x_{k+1} = x_k + \lambda_k d_k, \tag{1.2}$$

where

$$d_k = -B_k^{-1} g_k. \tag{1.3}$$

Here g_k is the gradient of f at x_k, the scalar λ_k is a steplength parameter, and the

* Received by the editors July 8, 1986; accepted for publication March 18, 1987.

† Department of Computer Science, University of Colorado, Boulder, Colorado 80309. This author was supported, in part, by National Science Foundation Grant No. DCR-84-03483.

‡ Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, Illinois 60201. This author was supported, in part, by National Science Foundation Grant No. DCR-86-02071

§ Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England CB3 9EW. This author was supported by a research fellowship from Fitzwilliam College, Cambridge.

Hessian approximation B_k is updated by the formula of Broyden (1967)

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} + \phi(s_k^T B_k s_k) v_k v_k^T \tag{1.4}$$

where ϕ is a scalar,

$$\begin{aligned} y_k &= g_{k+1} - g_k, \\ s_k &= x_{k+1} - x_k, \end{aligned}$$

and

$$v_k = \left[\frac{y_k}{y_k^T s_k} - \frac{B_k s_k}{s_k^T B_k s_k} \right].$$

This class contains the two most popular update formulas: for $\phi = 0$ we obtain the BFGS method and for $\phi = 1$ the DFP method; see for example Dennis and Moré (1977) Although Dixon (1972) shows that with exact line searches all members of this class produce the same iterates, their performance varies markedly when inexact lines searches are performed. We will assume that the steplength parameter λ_k is chosen by an inexact line search satisfying the two conditions

$$f(x_k + \lambda_k d_k) \le f(x_k) + \alpha \lambda_k g_k^T d_k, \tag{1.5}$$

$$g(x_k + \lambda_k d_k)^T d_k \ge \beta g_k^T d_k, \tag{1.6}$$

where $0 < \alpha < \frac{1}{2}$ and $\alpha < \beta < 1$.

There are several important results about this class of methods which do not require an exact line search. Let x_* denote a minimizer of f and assume that the Hessian matrix of f at x_*, G_*, is positive definite. Dennis and Moré (1977) proved that if λ_k is always taken equal to 1 in the DFP and BFGS methods, and if

$$\sum_{k=0}^{\infty} \|x_{k+1} - x_*\| < \infty, \tag{1.7}$$

then $\{x_k\}$ converges to x_* Q-superlinearly. They also show that this sum will be finite if $\|x_1 - x_*\|$ and $\|B_1 - G_*\|$ are sufficiently small, and that the stepsize $\lambda_k = 1$ will eventually satisfy the two line search conditions (1.5) and (1.6). These results have been extended to the restricted Broyden class ($\phi \in [0, 1]$) by Stachurski (1981), Ritter (1979) and Griewank and Toint (1982). Removing the assumption about the initial Hessian approximation has been difficult. Stoer (1975), analyzing the restricted Broyden class with an inexact line search, proved that for any positive definite starting matrix B_1 there exists $\epsilon > 0$, dependent on the condition number of B_1, such that if $\|x_1 - x_*\| < \epsilon$, then linear convergence occurs. Using these arguments, superlinear convergence can be shown if steplengths of 1 are used when possible. The weakness of this result is the dependence of ϵ on the initial Hessian approximation.

An analysis that corresponds closely to quasi-Newton methods as they are implemented is that of Powell (1976). He showed that if the objective function f is convex, then for the sequence $\{x_k\}$ generated by the BFGS method, with line search satisfying (1.5) and (1.6), we have lim inf $\{g_k\} = 0$. If in addition the sequence $\{x_k\}$ converges to a point x_*, with $G(x_*)$ positive definite, then the convergence rate is R-linear, which implies (1.7). If the line search procedure always tries $\lambda_k = 1$ in trying to satisfy (1.5) and (1.6), then the results of Dennis and Moré (1974) apply and superlinear convergence is obtained. It is not known if the assumption that f is convex can be weakened. Werner (1978) extended Powell's result to several other practical line search strategies under the assumption of uniform convexity of f. Ritter (1981) proved a global convergence result, related to Powell's, for the restricted Broyden class. However his result is true only under certain restrictive assumptions on the behavior of the line search, and it is not known if these assumptions are satisfied by any practical line search.

In this paper we show that if the objective function is convex then the sequence generated by algorithm (1.2)-(1.4) with $\phi \in [0, 1)$ satisfies $\liminf\{g_k\} = 0$. Moreover, if we assume that the objective function is uniformly convex the iterates converge R-linearly to the solution. Then if we always try $\lambda = 1$ the results of Dennis and Moré (1974) and of Griewank and Toint (1982) imply superlinear convergence. Thus we extend Powell's result to the restricted Broyden class, except for DFP. Note that this result implies the following local convergence theorem for a general function: if an iterate falls sufficiently close to a strong minimizer, and if the iterates remain in the neighborhood of this minimizer, then the sequence converges superlinearly. Our results also apply if ϕ takes on a different value on [0,1) at each step, as long as it remains bounded away from 1. It has come to our attention that Xie Yuan-Fu (1986) has independently proven a result similar to the one of §3, but with additional restrictions on the choice of ϕ. The convergence analysis given in this paper shows that BFGS has a property that enables it to rapidly correct large eigenvalues, and also shows that this property is diminished as ϕ is increased in $[0, 1)$.

The paper is organized as follows. First we assume that the objective function is uniformly convex, and in §3 and §4 prove global and superlinear convergence. This analysis gives us much insight into the properties of the Broyden class of algorithms, and is one of the main contributions of this paper. In §5 we assume only that f is convex and bounded below, and show that convergence to the solution is obtained. Finally, in §6 we discuss the behavior of the DFP method, again under the assumption of a uniformly convex objective function.

2. Preliminaries. Throughout the paper we will assume that $f(x)$ is twice continuously differentiable and denote its matrix of second derivatives by $G(x)$. We will use $\|\cdot\|$ to denote the Euclidean norm or its induced matrix norm. The starting point for the algorithm (1.2)-(1.3) is x_1, and we define the level set $D = \{x \in \mathbf{R}^n : f(x) \leq f(x_1)\}$. In this section, as well as in §3,4 and 6, we will assume that f is uniformly convex on D, which implies that f has a unique minimizer x_* in D.

ASSUMPTION 2.1. *The level set D is convex and there exist positive constants m and M such that*

$$m\|z\|^2 \leq z^T G(x) z \leq M\|z\|^2 \tag{2.1}$$

for all $z \in \mathbf{R}^n$ and all $x \in D$.

An immediate consequence of Assumption 2.1 is that, if we define

$$\overline{G} = \int_0^1 G(x_k + \tau s_k) d\tau, \tag{2.2}$$

then we have

$$y_k = \overline{G} s_k, \tag{2.3}$$

which implies

$$m\|s_k\|^2 \leq y_k^T s_k \leq M\|s_k\|^2. \tag{2.4}$$

We will denote by θ_k the angle between the steepest descent direction $-g_k$ and the displacement s_k. Hence

$$-g_k^T s_k = \|g_k\| \|s_k\| \cos\theta_k \tag{2.5}$$

As a consequence of the line search conditions (1.5) and (1.6), the angle θ_k will determine important properties about the length of the displacement and the decrease in the function per step. Many of these conditions have been proved elsewhere; see for example Wolfe (1969), (1971), Stoer (1975), Powell (1976) or Warth and Werner (1977). For clarity and completeness we will derive and discuss these conditions here, using arguments similar to those of Powell. We note that all the results of this section apply to any quasi-Newton method with positive definite Hessian approximation, regardless of how B_k is updated.

From (1.6) it follows that

$$y_k^T s_k = g_{k+1}^T s_k - g_k^T s_k \geq -(1-\beta) g_k^T s_k, \tag{2.6}$$

and hence from (2.4) and (2.5) we obtain

$$\|s_k\| \geq c_1 \|g_k\| \cos\theta_k, \tag{2.7}$$

where $c_1 = (1-\beta)/M$.

We can use this lower bound on the displacement to bound the amount of objective function decrease per step. From the first line search condition (1.5) and (2.7) we have that

$$\begin{aligned} f_{k+1} - f_k &\leq \alpha g_k^T s_k \\ &\leq -\alpha c_1 \|g_k\|^2 \cos^2\theta_k. \end{aligned} \tag{2.8}$$

Thus, if $\cos\theta_k$ is not too small, we see from (2.7) and (2.8) that the displacement is proportional to the gradient, and the objective function decrease to the squared gradient.

Since f is convex on D,

$$\begin{aligned} f_k - f_* &\leq g_k^T (x_k - x_*) \\ &\leq \|g_k\| \|x_k - x_*\| \end{aligned} \tag{2.9}$$

Let $\widetilde{G}$ be defined by

$$\widetilde{G} = \int_0^1 G\left(x_k + \tau(x_* - x_k)\right) d\tau,$$

so that $g_k = \widetilde{G}(x_k - x_*)$. Then, since $\widetilde{G}$ satisfies (2.1), we have

$$m\|x_k - x_*\|^2 \leq (x_k - x_*)^T g_k$$

and thus

$$\|x_k - x_*\| \leq \frac{1}{m}\|g_k\|.$$

Using (2.9) we have

$$\|g_k\|^2 \geq m(f_k - f_*) \tag{2.10}$$

Substituting (2.10) into (2.8) gives

$$f_{k+1} - f_* \leq [1 - \alpha m c_1 \cos^2\theta_k](f_k - f_*).$$

It then follows that, if there is a subsequence of the iterates for which the $\cos\theta_k$ are bounded away from zero, then the sequence $\{x_k\}$ converges to x_*. In the next section we will establish the existence of such a subsequence.

The line search condition (1.5) also gives a useful upper bound on the length of the displacement. By Taylor's theorem

$$f_{k+1} - f_k = g_k^T s_k + \frac{1}{2} s_k^T G(\xi_k) s_k$$

for some ξ_k between x_{k+1} and x_k. Therefore by (1.5)

$$\alpha g_k^T s_k \geq g_k^T s_k + \frac{1}{2} s_k^T G(\xi_k) s_k, \tag{2.11}$$

and using (2.1)

$$(1 - \alpha)\|g_k\|\|s_k\|\cos\theta_k \geq \frac{1}{2} m\|s_k\|^2.$$

Thus

$$\|s_k\| \leq c_2\|g_k\|\cos\theta_k,$$

where $c_2 = 2(1 - \alpha)/m$. We have thus proved the following result.

LEMMA 2.1. *Consider the iteration* (1.2), *where* λ_k *satisfies* (1.5) *and* (1.6). *If Assumption* 2.1 *holds, then*

$$c_1\|g_k\|\cos\theta_k \leq \|s_k\| \leq c_2\|g_k\|\cos\theta_k \tag{2.12}$$

and

$$f_{k+1} - f_* \le [1 - \alpha m c_1 \cos^2\theta_k](f_k - f_*), \tag{2.13}$$

where $c_1 = (1-\beta)/M$ *and* $c_2 = 2(1-\alpha)/m$

Similar arguments give us bounds on the steplength λ_k. Since $\lambda_k g_k = -B_k s_k$, equations (2.3) and (2.6) imply

$$\begin{aligned}(1-\beta)s_k^T B_k s_k &= -(1-\beta)\lambda_k s_k^T g_k \\ &\le \lambda_k s_k^T \overline{G} s_k.\end{aligned}$$

Therefore

$$\lambda_k \ge (1-\beta)\frac{s_k^T B_k s_k}{s_k^T \overline{G} s_k}. \tag{2.14}$$

Likewise (2.11) implies that

$$\begin{aligned}\frac{1}{2}s_k^T G(\xi_k)s_k &\le -(1-\alpha)g_k^T s_k \\ &= (1-\alpha)\frac{s_k^T B_k s_k}{\lambda_k}.\end{aligned}$$

Thus

$$\lambda_k \le 2(1-\alpha)\frac{s_k^T B_k s_k}{s_k^T G(\xi_k)s_k} \tag{2.15}$$

Using (2.1) in (2.14) and (2.15), and recalling the definitions of c_1 and c_2, we obtain the following.

LEMMA 2.2 *The steplength parameter* λ_k *satisfies*

$$c_1\frac{s_k^T B_k s_k}{\|s_k\|^2} \le \lambda_k \le c_2\frac{s_k^T B_k s_k}{\|s_k\|^2}. \tag{2.16}$$

Relations (2.14) and (2.15) show that the steplength parameter λ_k is a measure of the ratio between our estimate $s_k^T B_k s_k$ of the second derivative of the objective function along s_k and its true value If B_k is a good approximation $\lambda_k = 1$ should satisfy (1.5) and (1.6). In the next section we will use the quantities $\cos\theta_k$ and λ_k to measure how bad the matrix B_k is, along the current step.

3. Global convergence for uniformly convex problems. One can show that if B_k is positive definite, $\phi \in [0,1]$ and $y^T s > 0$ then the new matrix B_{k+1} generated by Broyden's formula (1.4) is also positive definite We will assume that these three conditions hold at every step and thus $\cos\theta_k > 0$ for all k. Our convergence analysis borrows heavily from the techniques developed by Powell (1976).

Since the progress per step is dependent on $\cos\theta$ we are interested in bounding $\cos\theta$ from below Note that

$$\cos\theta_k = \frac{\|s_k\|}{\|B_k s_k\|}\frac{s_k^T B_k s_k}{\|s_k\|^2}. \tag{3.1}$$

The factor $s_k^T B_k s_k/\|s_k\|^2$ will be bounded below in mean by using the determinant of B_k. For the other term we note that $\|B_k s_k\|/\|s_k\| \leq \|B_k\| < \mathrm{Tr}\,(B_k)$, where $\mathrm{Tr}\,(B_k)$ denotes the trace of B_k. We thus begin by estimating, from (1.4), the trace of the Hessian approximation:

$$\mathrm{Tr}\,(B_{k+1}) = \mathrm{Tr}\,(B_k) - \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + \frac{\|y_k\|^2}{y_k^T s_k} + \phi(s_k^T B_k s_k)\|v_k\|^2.$$

Since

$$\|v_k\|^2 = \frac{\|y_k\|^2}{(y_k^T s_k)^2} - 2\frac{y_k^T B_k s_k}{(y_k^T s_k)(s_k^T B_k s_k)} + \frac{\|B_k s_k\|^2}{(sk^T B_k s_k)^2},$$

we have

$$\begin{aligned} \mathrm{Tr}\,(B_{k+1}) = \mathrm{Tr}\,(B_k) \quad &+ \quad \frac{\|y_k\|^2}{y_k^T s_k} + \phi\frac{\|y_k\|^2}{y_k^T s_k}\frac{s_k^T B_k s_k}{y_k^T s_k} \\ &- \quad (1-\phi)\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} - 2\phi\frac{y_k^T B_k s_k}{y_k^T s_k}. \end{aligned} \tag{3.2}$$

We will now bound all the terms in (3.2).

LEMMA 3.1. *The following four inequalities hold:*

$$\frac{\|y_k\|^2}{y_k^T s_k} \leq M, \tag{3.3}$$

$$\frac{s_k^T B_k s_k}{y_k^T s_k} \leq \frac{\lambda_k}{1-\beta}, \tag{3.4}$$

$$\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} \geq \frac{\lambda_k}{c_2 \cos^2\theta_k}, \tag{3.5}$$

$$\frac{|y_k^T B_k s_k|}{y_k^T s_k} \leq \frac{\lambda_k M}{m c_1 \cos\theta_k} \tag{3.6}$$

Proof. The proof of the first inequality is due to Goldfarb and is given in Powell (1972) as follows. Let us define $z_k = \overline{G}^{\frac{1}{2}} s_k$, where $\overline{G}^{\frac{1}{2}}\overline{G}^{\frac{1}{2}} = \overline{G}$. Then from (2.1) and (2.3)

$$\begin{aligned} \frac{y_k^T y_k}{y_k^T s_k} &= \frac{s_k^T \overline{G}^2 s_k}{s_k^T \overline{G} s_k} \\ &= \frac{z_k^T \overline{G} z_k}{z_k^T z_k} \\ &\leq M. \end{aligned}$$

Inequality (3.4) follows from (2.6)

$$\frac{s_k^T B_k s_k}{y_k^T s_k} \le \frac{s_k^T B_k s_k}{(1-\beta)(-g_k^T s_k)} = \frac{\lambda_k}{1-\beta}.$$

Using (2.12)

$$\begin{aligned} \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} &= \frac{\lambda_k^2 \|g_k\|^2}{\lambda_k \|s_k\| \|g_k\| \cos\theta_k} \\ &= \frac{\lambda_k \|g_k\|}{\|s_k\| \cos\theta_k} \\ &\ge \frac{\lambda_k}{c_2 \cos^2\theta_k}. \end{aligned}$$

Finally, from (2.4), (2.3) and (2.7)

$$\begin{aligned} \frac{|y_k^T B_k s_k|}{y_k^T s_k} &\le \frac{\lambda_k \|y_k\| \|g_k\|}{m \|s_k\|^2} \\ &\le \frac{\lambda_k M \|g_k\|}{m c_1 \|g_k\| \cos\theta_k}. \end{aligned}$$

□

Substituting these bounds in (3.2) we have,

$$\text{Tr}\ (B_{k+1}) \le \text{Tr}\ (B_k) + M + \frac{\phi M \lambda_k}{1-\beta} - \frac{(1-\phi)\lambda_k}{c_2 \cos^2\theta_k} + \frac{2\phi M \lambda_k}{m c_1 \cos\theta_k}. \tag{3.7}$$

These bounds, together with (3.2), give us some insight into the properties of the Broyden class of updates. Note that the second and third terms on the right-hand side of (3.2) produce an average *shift to the right* in the eigenvalues of B_{k+1}, in the sense that they increase the trace. The fourth term on the right-hand side of (3.2) is crucial to the convergence analysis given below. It produces a *shift to the left* in the eigenvalues which is proportional to $\lambda_k / \cos^2\theta_k$. The last term on the right-hand side of (3.2) can produce a *shift in either direction,* which is proportional to $\lambda_k / \cos\theta_k$; in (3.7) we have substituted an upper bound for this term.

We can now reason as follows. If the algorithm produces steps for which $\cos\theta$ is not very small, it will advance towards the solution, but some of the eigenvalues of B_{k+1} could become large. On the other hand, if steps with very small $\cos\theta$ are produced little progress may be achieved, but a self correcting mechanism takes place. To see this note that from (2.16) and the fact that $\|B_k s_k\| / \|s_k\| \le \text{Tr}\ (B_k)$, (3.1) gives

$$\cos\theta_k \ge \frac{\lambda_k}{c_2 \text{Tr}\ (B_k)}. \tag{3.8}$$

Thus $\cos\theta_k$ can be small only if λ_k is small or Tr (B_k) is large. Let us consider the following two cases, assuming $\phi < 1$. (1) Suppose that the steplengths λ_k are bounded away from zero. Then if $\cos\theta_k$ becomes arbitrarily small, the fourth term in (3.7) will dominate all the others, thus reducing the trace. Therefore, we see from (3.8) that the tendency of $\cos\theta_k$ to go to zero is self limiting due to the shift to the left in the trace equation. This argument is made rigorous in Theorem 3.1. (2) Suppose that the steplengths λ_k tend to zero. From Lemma 2.2 we see that this implies that $s_k^T B_k s_k/\|s_k\|^2$ becomes very small. This situation is due to very small eigenvalues of B_k and thus cannot be monitored by the trace. However, if λ_k is near zero this indicates that the determinant is very small. Fortunately, it turns out that all the updates in the restricted Broyden class have a strong self correcting property with respect to the determinant. We now derive this property and use it to show that, in fact, λ_k is bounded away from zero in mean.

Since $\phi(s_k^T B_k s_k) \geq 0$ the last term in (1.4) increases the eigenvalues, and hence

$$\det\,(B_{k+1}) \geq \det\,\left(B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}\right).$$

It is not difficult to show (see for example Pearson (1969)) that

$$\det\left(B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}\right) = \det(B_k)\left(\frac{y_k^T s_k}{s_k^T B_k s_k}\right),$$

and thus

$$\det\,(B_{k+1}) \geq \det\,(B_k)\frac{y_k^T s_k}{s_k^T B_k s_k}. \tag{3.9}$$

Note that when $s_k^T B_k s_k$ is small relative to $y_k^T s_k = s_k^T \overline{G} s_k$ the determinant increases, reflecting the fact that the small curvature of our model is corrected, thus increasing some eigenvalues.

LEMMA 3.2 *If $\phi \in [0,1]$, there is a constant $c_4 > 0$ such that for all $k \geq 1$*

$$\prod_{j=1}^{k} \lambda_j \geq c_4^k. \tag{3.10}$$

Proof. Using (2.16) and (3.1)

$$\frac{\lambda_k}{\cos\theta_k} \leq c_2 \frac{\|B_k s_k\|}{\|s_k\|}.$$

We can use this to bound the fifth term on the right side of (3.7), and since $\cos\theta_k \leq 1$, also the third term. Thus deleting the always negative fourth term of (3.7) we have

$$\begin{aligned}\mathrm{Tr}\,(B_{k+1}) &\leq \mathrm{Tr}\,(B_k) + M + \left(\frac{1}{1-\beta} + \frac{2}{mc_1}\right)\phi M c_2 \frac{\|B_k s_k\|}{\|s_k\|} \\ &\leq M + \left[1 + \left(\frac{1}{1-\beta} + \frac{2}{mc_1}\right)\phi M c_2\right]\mathrm{Tr}\,(B_k).\end{aligned}$$

This inequality implies that there is a constant c_3 such that

$$\mathrm{Tr}\,(B_{k+1}) \leq c_3^k. \tag{3.11}$$

From (3.9) and (3.4)

$$\begin{aligned}\det(B_{k+1}) &\geq \det(B_k)\frac{y_k^T s_k}{s_k^T B_k s_k}\\ &\geq \det(B_k)\frac{1-\beta}{\lambda_k}\\ &\geq \det(B_1)\prod_{j=1}^{k}\frac{1-\beta}{\lambda_j} \qquad (3.12)\end{aligned}$$

Using the geometric/arithmetic mean inequality we have

$$\det(B_{k+1}) \leq \left[\frac{\mathrm{Tr}(B_{k+1})}{n}\right]^n,$$

and thus from (3.11) and (3.12)

$$\begin{aligned}\prod_{j=1}^{k}\frac{1-\beta}{\lambda_j} &\leq \frac{1}{\det(B_1)}\left[\frac{\mathrm{Tr}(B_{k+1})}{n}\right]^n\\ &\leq \frac{1}{\det(B_1)n^n}(c_3^n)^k.\end{aligned}$$

Thus there exists a constant c_4 such that

$$\prod_{j=1}^{k}\lambda_j \geq c_4^k$$

for all $k \geq 1$. □

Now that we have bounded the steplengths below in mean we can use the trace equation to prove global convergence for the restricted Broyden methods, *except* for DFP. The basic idea of the proof is as follows. The algorithm cannot produce too many steps with $\cos\theta \approx 0$ for otherwise the shift to the left in equation (3.7) would give negative eigenvalues for the matrices B_k, which is not possible. We therefore show in Theorem 3.1 that there is a subsequence of the iterates $\{x_k\}$ for which $\cos\theta_k$ is bounded away from zero. The convergence result then follows from (2.13).

Note that the crucial term responsible for the shift to the left, namely the fourth term on the right-hand side of (3.7), is not present in the DFP method ($\phi = 1$). This is quite significant, as we will discuss in §6.

THEOREM 3.1. *Let x_1 be a starting point for which f satisfies Assumption 2.1. Then for any positive definite B_1, Broyden's class of algorithms,* (1.2)-(1.4), *with $\phi \in [0,1)$ and line search satisfying* (1.5)-(1.6), *generates iterates which converge to x_*.*

Proof. Let us write (3.7) as

$$\mathrm{Tr}(B_{k+1}) \leq \mathrm{Tr}(B_k) + M + \eta_k\lambda_k, \qquad (3.13)$$

where

$$\eta_k = \frac{\phi M}{1-\beta} - \frac{(1-\phi)}{c_2\cos^2\theta_k} + \frac{2\phi M}{mc_1\cos\theta_k} \qquad (3.14)$$

We seek a contradiction: assume $\{\cos\theta_k\} \to 0$. Then it is clear from (3.14) that $\{\eta_j\} \to -\infty$; thus there is an index K_0 such that $\eta_j < -2M/c_4$ for all $j \geq K_0$. Now using (3.13) and the fact that B_{k+1} is positive definite we have, for all $k \geq K_0$,

$$\begin{aligned} 0 < \operatorname{Tr}(B_{k+1}) &\leq \operatorname{Tr}(B_{K_0}) + M(k+1-K_0) + \sum_{j=K_0}^{k} \eta_j \lambda_j \\ &< \operatorname{Tr}(B_{K_0}) + M(k+1-K_0) - \frac{2M}{c_4} \sum_{j=K_0}^{k} \lambda_j . \end{aligned} \tag{3.15}$$

Applying the geometric/arithmetic mean inequality to (3.10) we have

$$\sum_{j=1}^{k} \lambda_j \geq k c_4,$$

and hence

$$\sum_{j=K_0}^{k} \lambda_j \geq k c_4 - \sum_{j=1}^{K_0-1} \lambda_j .$$

Substituting this into (3.15) we have

$$0 < \operatorname{Tr}(B_{K_0}) + M(k+1-K_0) - \frac{2M}{c_4} k c_4 + \frac{2M}{c_4} \sum_{j=1}^{K_0-1} \lambda_j ;$$

therefore

$$0 < \operatorname{Tr}(B_{K_0}) + M(1-k) - MK_0 + \frac{2M}{c_4} \sum_{j=1}^{K_0-1} \lambda_j .$$

For sufficiently large k the right-hand side is negative, which is a contradiction. Therefore there is a subsequence for which $\cos\theta_k$ is bounded away from zero. Using (2.13) we conclude that the iterates converge to the solution. □

4. Superlinear Convergence. To analyze the rate of convergence of these algorithms we must use a more precise measure than the trace of B_k. As is well known, the Broyden class of methods is invariant under a linear change of variables. Therefore in this section we will assume without loss of generality that $G(x_*) = I$, as this results when we make the change of variables from x to $x_* + G(x_*)^{\frac{1}{2}}(x - x_*)$. (This is equivalent to using the original variables and studying $\operatorname{Tr}(G_*^{-\frac{1}{2}} B_k G_*^{-\frac{1}{2}})$.) We begin by combining the third and fifth terms on the right-hand side of (3.2).

LEMMA 4.1. *For any $0 < \epsilon \leq 1$ there is a neighborhood $N(x_*)$ of x_* such that if x_{k+1} and $x_k \in N(x_*)$ then*

$$\frac{\|y_k\|^2}{y_k^T s_k} \frac{s_k^T B_k s_k}{y_k^T s_k} - 2\frac{y_k^T B_k s_k}{y_k^T s_k} \leq \frac{2\lambda_k \epsilon}{m c_1 \cos\theta_k} . \tag{4.1}$$

Proof. Since $G(x_*) = I$ we have

$$\begin{aligned} y_k &= \left[\int_0^1 G(x_k + \tau s_k) d\tau\right] s_k \\ &= \left[\int_0^1 (G(x_k + \tau s_k) - I)\, d\tau\right] s_k + s_k \\ &\equiv E_k s_k + s_k . \end{aligned} \tag{4.2}$$

Since G is continuous, for $N(x_*)$ sufficiently small we have

$$\|E_k\| \leq \epsilon \tag{4.3}$$

From (4.2) and (4.3)

$$\begin{aligned} \frac{y_k^T y_k}{y_k^T s_k} &= \frac{s_k^T (I+E_k)^2 s_k}{s_k^T (I+E_k) s_k} \\ &\leq 1+\epsilon \end{aligned}$$

Using this with (4.2), (4.3), (2.1) and (2.12) we have

$$\begin{aligned} \frac{y_k^T y_k}{y_k^T s_k}\frac{s_k^T B_k s_k}{y_k^T s_k} - 2\frac{y_k^T B_k s_k}{y_k^T s_k} &= \left(\frac{y_k^T y_k}{y_k^T s_k} - 2\right)\frac{s_k^T B_k s_k}{y_k^T s_k} - 2\frac{s_k^T E_k B_k s_k}{y_k^T s_k} \\ &\leq (-1+\epsilon)\frac{s_k^T B_k s_k}{y_k^T s_k} + 2\epsilon\frac{\|s_k\|\|B_k s_k\|}{m\|s_k\|^2} \\ &\leq \frac{2\epsilon\lambda_k\|g_k\|}{m\|s_k\|} \\ &\leq \frac{2\epsilon\lambda_k}{mc_1\cos\theta_k}, \end{aligned}$$

since $\epsilon \leq 1$ □

We can now prove R-linear convergence.

LEMMA 4 2. *Assume that $\phi \in [0,1)$ Then there is a constant $0 \leq c_8 < 1$ such that*

$$f_{k+1} - f_* \leq c_8^k [f_1 - f_*] \tag{4.4}$$

holds for all sufficiently large k.

Proof. We know from Theorem 3.1 that the iterates converge to the solution. Therefore by Lemma 4.1, for any $0 < \epsilon \leq 1$, substituting (3.3), (3.5) and (4.1) into (3.2), for k sufficiently large, gives

$$\begin{aligned} 0 < \mathrm{Tr}\,(B_{k+1}) &\leq \mathrm{Tr}\,(B_k) + M + \frac{2\lambda_k\phi\epsilon}{mc_1\cos\theta_k} - (1-\phi)\frac{\lambda_k}{c_2\cos^2\theta_k} \\ &\leq \mathrm{Tr}\,(B_k) + M + \frac{1}{\cos^2\theta_k}\left[\frac{2\phi\epsilon\cos\theta_k}{mc_1} - \frac{(1-\phi)}{c_2}\right]\lambda_k. \end{aligned} \tag{4.5}$$

We may take ϵ small enough that the term inside the square brackets in (4.5) is less than $-c_5$, where c_5 is some positive constant. Thus from (4.5), there is a constant $c_6 > 0$ such that

$$0 < \mathrm{Tr}\,(B_{k+1}) < \mathrm{Tr}\,(B_1) + Mk + c_6 - c_5\sum_{j=1}^{k}\frac{\lambda_j}{\cos^2\theta_j},$$

and hence

$$\sum_{j=1}^{k} \frac{\lambda_j}{\cos^2\theta_j} \le c_7 k, \tag{4.6}$$

for some constant c_7. Applying the geometric/arithmetic mean inequality to (4.6) gives

$$\prod_{j=1}^{k} \frac{\lambda_j}{\cos^2\theta_j} \le c_7^k. \tag{4.7}$$

From Lemma 3.2

$$\prod_{j=1}^{k} \lambda_j \ge c_4^k,$$

and hence

$$\prod_{j=1}^{k} \cos^2\theta_j \ge \left(\frac{c_4}{c_7}\right)^k. \tag{4.8}$$

By (2.13) the objective function decrease is given by

$$f_{k+1} - f_* \le \prod_{j=1}^{k} (1 - \alpha m c_1 \cos^2\theta_j)\,[f_1 - f_*].$$

Using the geometric/arithmetic mean inequality twice we have

$$\begin{aligned} f_{k+1} - f_* &\le \left[\frac{1}{k}\sum_{j=1}^{k}(1 - \alpha m c_1 \cos^2\theta_j)\right]^k [f_1 - f_*] \\ &\le \left[1 - \alpha m c_1 \left(\prod_{j=1}^{k}\cos^2\theta_j\right)^{\frac{1}{k}}\right]^k [f_1 - f_*] \end{aligned}$$

Then by (4.8)

$$f_{k+1} - f_* \le c_8^k [f_1 - f_*], \tag{4.9}$$

where

$$c_8 = \left[1 - \alpha m c_1 \left(\frac{c_4}{c_7}\right)\right]. \qquad \square$$

To prove superlinear convergence we will use the well-known results of Dennis and Moré together with a result of Griewank and Toint. However to apply them it is necessary that the algorithm tries steplengths of 1 in the line search. We also require that the Hessian matrix G be Hölder continuous at x_*, i.e. that there exist positive constants p, L such that

$$\|G(x) - G(x_*)\| \leq L\|x - x_*\|^p \tag{4.10}$$

for all x in a neighborhood of x_*.

THEOREM 4.1. *Assume that the algorithm* (1.2)-(1.4), *with* $\phi \in [0, 1)$, *is implemented so that* λ_k *satisfies* (1.5)-(1.6) *and* $\lambda_k = 1$ *whenever this satisfies* (1.5)-(1.6). *Then if Assumption* 2.1 *and* (4.10) *hold, and if* B_1 *is any positive definite matrix, the sequence* $\{x_k\}$ *converges to* x_* *Q-superlinearly.*

Proof. From (2.1)

$$f_{k+1} - f_* \geq \frac{1}{2}m\|x_{k+1} - x_*\|^2,$$

which combined with (4.9) gives

$$\|x_{k+1} - x_*\|^2 \leq \frac{2}{m}c_8^k[f_1 - f_*]$$

Therefore, for p given by (4.10),

$$\begin{aligned}\sum_{k=0}^{\infty}\|x_{k+1} - x_*\|^p &\leq \left(\frac{2}{m}[f_1 - f_*]\right)^{\frac{p}{2}}\sum_{k=0}^{\infty}c_8^{pk/2} \\ &< \infty\end{aligned}$$

In analyzing the restricted Broyden class, Griewank and Toint (1982) prove in their Proposition 4 that, given (4.10), this summability implies

$$\lim_{k\to\infty}\frac{\|(B_k - G(x_*))\,s_k\|}{\|s_k\|} = 0 \tag{4.11}$$

This fact alone implies Q-linear convergence since $\cos\theta_k = s_k^T B_k s_k/\|B_k s_k\|\|s_k\|$, and by (4.11) B_k converges to $G(x_*)$ along s_k. However from Theorem 6.4 of Dennis and Moré (1977) we conclude that the steplength $\lambda_k = 1$ is admissible for all sufficiently large k, and that the rate of convergence is superlinear. □

5. Relaxing the Uniform Convexity Assumption. In this section we show that the global convergence result of §3 can be established under milder assumptions on the objective function. We no longer relate the shifts in the eigenvalues to the angle θ, as was done in §3. However the result is of much interest, because it establishes global convergence for very general convex functions. We do not extend the rates of convergence results of §4 because these require the nonsingularity of $G(x_*)$, and thus the uniform convexity of f in a neighborhood of x_*.

Let us replace Assumption 2.1 by the following assumption.

ASSUMPTION 5.1. *The function* f *is twice continuously differentiable, convex and bounded below. Moreover the Hessian matrix is bounded*

$$\|G(x)\| \leq M \tag{5.1}$$

for all x in the level set D.

Let us define f_* to be the infimum of f. From (1 5) we have that $s_k^T g_k$ tends to zero, since

$$\begin{aligned} \sum_{k=1}^{\infty} -s_k^T g_k &= -\sum_{k=1}^{\infty} \lambda_k d_k^T g_k \\ &\leq \frac{1}{\alpha} \sum_{k=1}^{\infty} [f(x_k) - f(x_{k+1})] \\ &\leq \frac{1}{\alpha} [f(x_1) - f_*] \\ &< \infty. \end{aligned} \tag{5.2}$$

Our approach is to compare the third and fifth terms of (3.2) with the fourth term. To this end, let us define the number ψ_k by the equation

$$\frac{\|y_k\|^2}{y_k^T s_k} \frac{s_k^T B_k s_k}{y_k^T s_k} - 2\frac{y_k^T B_k s_k}{y_k^T s_k} = \psi_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}, \tag{5.3}$$

so that the trace equation (3 2) becomes

$$\text{Tr}\,(B_{k+1}) = \text{Tr}\,(B_k) + \frac{\|y_k\|^2}{y_k^T s_k} - (1 - \phi - \psi_k \phi) \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}. \tag{5.4}$$

The proof of the following theorem is based on showing that, if the algorithm fails, then $\psi_k \to 0$, so that (5.4) becomes essentially the same as the trace equation for BFGS. The argument makes use only of (2.6), (3.2) and (3.3), which do not require uniform convexity of f. We conclude the proof by quoting some of the arguments given is §3

THEOREM 5.1. *If the algorithm* (1 2)-(1.4), *with $\phi \in [0, 1)$, and with line search satisfying* (1 5)-(1 6) *is applied to a function that satisfies Assumption* 5 1, *and if x_1 is any starting point and B_1 is any positive definite matrix, then*

$$\liminf_{k \to \infty} \|g_k\| = 0 \tag{5.5}$$

Proof. Let us assume that $\|g_k\|$ is bounded away from zero, say

$$\|g_k\| \geq \gamma > 0, \tag{5.6}$$

in order to deduce a contradiction Note that (5.2) and (5.5) imply that

$$\frac{s_k^T g_k}{\|g_k\|^2} \to 0 \tag{5.7}$$

as $k \to \infty$. We will first show that $\psi_k \to 0$ To treat the first term of (5.3), we deduce the following bound, using (2.6) and (3.3)

$$\begin{aligned}
\frac{\|y_k\|^2}{y_k^T s_k}\frac{s_k^T B_k s_k}{y_k^T s_k} \div \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} &\leq M\frac{\left(s_k^T B_k s_k\right)^2}{y_k^T s_k \|B_k s_k\|^2} \\
&= \frac{M(s_k^T g_k)^2}{y_k^T s_k \|g_k\|^2} \\
&\leq \frac{M(-s_k^T g_k)}{(1-\beta)\|g_k\|^2}.
\end{aligned}$$

Condition (5.7) implies that this expression tends to zero as $k \to \infty$. We now consider the second term of (5.3). From (5.1), (3.3) and (2.6)

$$\begin{aligned}
\frac{|y_k^T B_k s_k|}{y_k^T s_k} \div \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} &\leq \frac{\|y_k\| s_k^T B_k s_k}{y_k^T s_k \|B_k s_k\|} \\
&\leq \frac{\sqrt{M} s_k^T B_k s_k}{\sqrt{y_k^T s_k}\, \|B_k s_k\|} \\
&= \frac{\sqrt{M}(-s_k^T g_k)}{\sqrt{y_k^T s_k}\, \|g_k\|} \\
&\leq \frac{\sqrt{M(-s_k^T g_k)}}{\sqrt{1-\beta}\, \|g_k\|}.
\end{aligned}$$

Once more (5.7) implies that this expression tends to zero. Hence $\psi_k \to 0$ as $k \to \infty$.

The trace equation (5.4) is therefore essentially the same as for BFGS, since for large k the term $(1 - \phi - \psi_k \phi) \in (0, 1]$. From (5.4) and (3.3) we see that the trace grows at most linearly, and therefore (3.11) is satisfied for some constant c_3. Then, using the same arguments as those appearing after (3.11) in the proof of Lemma 3.2, we conclude that (3.10) holds for some constant c_4,

$$\prod_{j=1}^{k} \lambda_j \geq c_4^k.$$

Now

$$-(1 - \phi - \psi_k \phi)\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} = \lambda_k \tilde{\eta}_k,$$

where

$$\tilde{\eta}_k = -(1 - \phi - \psi_k \phi)\frac{\|g_k\|^2}{(-s_k^T g_k)}.$$

Therefore the trace increase (5.4) can be bounded as in (3.13), with $\tilde{\eta}_k$ substituted for η_k. From (5.7) we see that $\{\tilde{\eta}_k\} \to -\infty$. Therefore there is a K_0 such that (3.15) holds. We now follow the same steps as in Theorem 3.1 to arrive at a contradiction. □

Although this theorem proves only that $\liminf_{k\to\infty} \|g_k\| = 0$ we can say more with slightly stronger assumptions. If we assume in addition to Assumption 5.1 that the sequence of iterates remains in a bounded set, then there is a subsequence $\{x_{k_i}\}$ such that $\{x_{k_i}\} \to z_*$, with $g(z_*) = 0$. Therefore z_* is a global minimizer and, since

$\{f(x_k)\}$ converges, it converges to the infimum of f. Moreover, $\lim_{k\to\infty}\|g_k\| = 0$, as the following argument shows. Suppose that there is a subsequence $\{x_{l_i}\}$ such that

$$\|g(x_{l_i})\| > \delta > 0.$$

Passing to a thinner subsequence if necessary we may assume that $\{x_{l_i}\} \to y_*$. Therefore, since $\{f(x_k)\}$ converges, y_* is a global minimizer and thus $g(y_*) = 0$. This contradiction shows that $\lim_{k\to\infty}\|g_k\| = 0$.

In order to prove superlinear convergence of the sequence $\{x_k\}$ one must assume that $G(x_*)$ is nonsingular. This implies that f is uniformly convex in some neighborhood of x_*, and since the sequence eventually is contained in that neighborhood the analysis of §4 implies that convergence is superlinear.

6. The DFP method. We now give some consideration to the behavior of the DFP method, which is not covered by the convergence theorems of the previous sections. Throughout this section we will assume that f is uniformly convex, i.e that Assumption 2.1 holds. The main difficulty with the DFP method seems to be with the possibility of unrestricted growth in the large eigenvalues of B_k. That small eigenvalues do not cause problems can be seen from Lemma 3.2 which, for all ϕ in [0,1], gives a lower bound for $\prod_{j=1}^{k}\lambda_j^{1/k}$, and thus by Lemma 2.2 on the geometric mean of $s_k^T B_k s_k/\|s_k\|^2$. Since $\cos\theta_k \geq \lambda_k/[c_2 \text{ Tr } (B_k)]$ (see (3.8)), an upper bound on the eigenvalues would ensure convergence. Our essential tool for bounding the eigenvalues from above is the trace equation (3.2) which for $\phi < 1$ provides a shift to the left when $\cos\theta_k$ is small. However, with the DFP method it is not clear that this shift will occur. Indeed, when $\phi = 1$ (3.2) becomes

$$\text{Tr } (B_{k+1}) = \text{Tr } (B_k) + \frac{\|y_k\|^2}{y_k^T s_k}\left(1 + \frac{s_k^T B_k s_k}{y_k^T s_k}\right) - 2\frac{y_k^T B_k s_k}{y_k^T s_k}. \tag{6.1}$$

The second term on the right shifts the eigenvalues to the right, but there is no term guaranteed to shift them to the left. The last term is bounded by $2\lambda_k M/[mc_1 \cos\theta_k]$ (see (3.6)) and so, in addition to being of uncertain sign, in the case of small cosine it is smaller in magnitude than the guaranteed shift of $(1-\phi)\lambda_k/[c_2\cos^2\theta_k]$ provided by other members of the restricted Broyden class. Looking at this uncertain term close to the solution, and if we scale so that $G(x_*) = I$, we have

$$-2\frac{y_k^T B_k s_k}{y_k^T s_k} = -2\left[\frac{s_k^T B_k s_k}{y_k^T s_k} + \frac{s_k^T E_k B_k s_k}{y_k^T s_k}\right],$$

where E_k is as in (4.2). Thus there is a term providing a guaranteed shift to the left proportional to λ_k, but still a term of uncertain sign with magnitude at most proportional to $\lambda_k/\cos\theta_k$ times the error in x. Thus even close to the solution it is not clear whether a shift to the left will occur, and even if it does occur we expect progress to be slower than for methods with $\phi < 1$.

In the case of a quadratic objective function the term containing E_k does not occur and we have a relatively small term guaranteed to give a shift to the left proportional to λ_k. These observations seem to correspond to the recent study by Powell (1986) of BFGS and DFP applied to a two dimensional quadratic, although he used steplengths of one at all iterates. For this case he showed that the DFP will take much longer than the BFGS both to improve an initial Hessian approximation with very large eigenvalues, and to converge. He also showed that neither method has much difficulty with eigenvalues of B_1 that are too small, as is indicated by Lemma 3.2 in the general case

We have made several numerical experiments on nonquadratic problems using steplengths satisfying (1.5) and (1.6), and the results confirm the predictions of the analysis The following example is representative of what we have observed.

Consider the function

$$f(x) = \frac{1}{2}x^T x + \sigma\left(\frac{1}{2}x^T Ax\right)^2, \tag{6.2}$$

where $n = 2, \sigma = 0.1$ and

$$A = \begin{pmatrix} 5 & 1 \\ 1 & 3 \end{pmatrix}$$

This function satisfies Assumption 2.1 for any starting point x_1. We follow Powell (1986) and choose $x_1 = (\cos 70°, \sin 70°)$, and define the starting matrix so that it has one very large eigenvalue:

$$B_1 = \begin{pmatrix} 1 & 0 \\ 0 & 10^4 \end{pmatrix}.$$

We used a Vax 11/780 and double precision arithmetic For the line search we used a routine written by J Moré, which satisfies (1.5) and (1.6). The line search parameters were chosen to be $\alpha = 10^{-4}, \beta = 0.9$ The table below shows the number of iterations needed to obtain $\|x_k\| \leq 10^{-4}\|x_1\|$, for various values of ϕ.

	BFGS								DFP
ϕ	0	.2	.4	.6	.8	.9	.99	.999	1
no. iter.	15	21	26	32	66	115	630	2233	4041

It is clear, in this example, that the performance of the algorithms deteriorates dramatically as ϕ approaches 1. For the BFGS method the trace was reduced from 10^4 to 3 in 10 iterations; whereas after 3000 iterations of the DFP method the trace was still 1100 All methods used steplengths of one at all iterates, except for the first few. We have made tests with other objective functions and found that it is not uncommon to observe this type of result. We have also observed that when small eigenvalues are given, neither method has difficulties.

7. Final remarks. It is important to consider whether the results of this paper give us a useful local convergence result for nonconvex functions. Note that if x_* is a strong local minimizer, then there is a neighborhood $N(x_*)$, contained in a connected component of the level set of f, in which (2.1) holds. Let us now suppose that an iterate x_k falls in $N(x_*)$ and that the line search forces the rest of the sequence to stay within this neighborhood. Then our results imply that the iterates converge to x_* superlinearly. However there is no guarantee that the line search will prevent the sequence from leaving this connected component of the level set, and thus our results may not be applicable. Of course, if x_* is a strict global minimizer, convergence would be guaranteed.

We have not been able to either prove global convergence for DFP or to exhibit a counterexample. It is known that global convergence for convex functions can be shown if the line searches are exact; see Powell (1971),(1972). We are interested, however, in a practical line search, such as one based on (1.5) and (1.6).

Nevertheless, we have seen in this paper that the DFP method can be much less efficient than the BFGS method, and that the reason for this is that it lacks the self-correcting property given by the fourth term on the right-hand side of equation (3.2). Finally we should note that the results of this paper also apply if ϕ takes on a different value on [0,1) at every step, as long as it remains bounded away from 1.

Acknowledgements. We would like to thank Jorge Moré and Michael Powell for their many valuable comments and suggestions.

REFERENCES

C. G. BROYDEN (1967), *Quasi-Newton methods and their application to function minimization*, Math. Comp., 21, pp. 368-381.

J. E. DENNIS AND J. J. MORÉ (1974), *A characterization of superlinear convergence and its application to quasi-Newton methods*, Math. Comp., 28, pp. 549-560.

J. E. DENNIS AND J. J. MORÉ (1977), *Quasi-Newton methods, motivation and theory*, SIAM Rev., 19, pp. 46-89.

L. C. W. DIXON (1972), *Variable metric algorithms: Necessary and sufficient conditions for identical behavior on nonquadratic functions*, J. Optim. Theory Appl., 10, pp. 34-40.

A. GRIEWANK AND PH. L. TOINT (1982), *Local convergence analysis of partitioned quasi-Newton updates*, Numer. Math., 39, pp. 429-448.

J. D. PEARSON (1969), *Variable metric methods of minimization*, Comput. J. 12, pp. 171-178.

M. J. D. POWELL (1971), *On the convergence of the variable metric algorithm*, J. Inst. Math. Appl., 7, pp. 21-36.

——— (1972), *Some properties of the variable metric algorithm*, in Numerical Methods for Nonlinear Optimization, F.A. Lootsma, ed., Academic Press, New York.

——— (1976), *Some global convergence properties of a variable metric algorithm for minimization without exact line searches*, in Nonlinear Programming, SIAM-AMS Proceedings, Vol. IX, R.W. Cottle and C.E. Lemke, eds., Society for Industrial and Applied Mathematics, Philadelphia.

——— (1986), *How bad are the BFGS and DFP methods when the objective function is quadratic?*, Math. Programming, 34, pp. 34-47.

K. RITTER (1979), *Local and superlinear convergence of a class of variable metric methods*, Computing, 23, pp. 287-297.

——— (1981), *Global and superlinear convergence of a class of variable metric methods*, Math. Programming Stud., vol. 14, pp. 178-205.

A. STACHURSKI (1981), *Superlinear convergence of Broyden's class of variable metric methods*, Math. Programming Stud., 20, pp. 196-212.

J. STOER (1975), *On the convergence rate of imperfect minimization algorithms in Broyden's beta-class*, Math. Programming Stud., 9, pp. 313-335.

W. WARTH AND J. WERNER (1977), *Effiziente Schrittweitenfunktionen bei unrestringierten Optimierungsaufgaben*, Computing, 19, 1, pp.59-72.

J. WERNER (1978), *Über die globale konvergenz von Variable-Metric Verfahren mit nichtexakter Schrittweitenbestimmung*, Numer. Math., 31, pp. 321-334.

P. WOLFE (1969), *Convergence conditions for ascent methods*, SIAM Rev., 11, pp. 226-235.

——— (1971), *Convergence conditions for ascent methods. II: Some corrections*, SIAM Rev. 13, pp. 185-188

Y. F. XIE (1986), *The global convergence of a set of variable metric formulas without exact line searches*, manuscript, Department of Operations Research, Institute of Applied Mathematics, Academia Sinica, Beijing, China.