Computational Optimization and Applications, 19, 297–317, 2001

A Study of the Dennis-Wolkowicz Method on Convex Functions

GUANGHUI LIU guanghui.liu@ual.com
WHQKB, Research and Development, United Airlines, Elk Grove Township, IL 60007, USA

LIXING HAN* lih95001@uconnvm.uconn.edu
Department of Mathematics, U-3009, University of Connecticut, Storrs, CT 06269, USA

Received April 22, 1999; Accepted March 30, 2000

Abstract. In this paper, we analyze the global convergence of the least-change secant method proposed by Dennis and Wolkowicz, when applied to convex objective functions. One of the most distinguished features of this method is that the Dennis-Wolkowicz update doesn't necessarily belong to the Broyden convex family and can be close to the DFP update, but it still has the self-correcting property. We prove that, for convex objective functions, this method with the commonly used Wolfe line search is globally convergent. We also provide some numerical results.

Keywords: unconstrained optimization, Dennis-Wolkowicz method, Wolfe line search, global convergence

1. Introduction

We consider the quasi-Newton methods from the Broyden family for the unconstrained optimization problem

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

These methods are iterative with the form

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

where λ_k is the stepsize computed by some one dimensional line search techniques, and d_k is the search direction determined by

$$d_k = -H_k g_k = -B_k^{-1} g_k,$$

where g_k denotes the gradient of f at the point x_k, B_k is an approximation of the true Hessian $\nabla^2 f(x)$ at x_k. The matrix B_k (or its inverse matrix H_k) is updated by the formula

---

*Present address: Department of Mathematics, University of Michigan-Flint, Flint, MI 48502, email: lxhan@umflint.edu.

of Broyden:

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

where

$$s_k = x_{k+1} - x_k = \lambda_k d_k, \quad y_k = g_{k+1} - g_k,$$
$$v_k = \frac{y_k}{y_k^T s_k} - \frac{B_k s_k}{s_k^T B_k s_k},$$

and ϕ_k is a parameter.

Two most famous update formulas within the Broyden family are the DFP update and the BFGS update, with the DFP update associated with $\phi_k = 0$ and the BFGS update associated with $\phi_k = 1$. The Broyden convex family is defined as all the updates associated with $\phi_k \in [0, 1]$ in the Broyden family. If $\phi_k \in [\delta, 1]$, $\delta \in (0, 1]$, then the corresponding Broyden convex family is called the restricted Broyden convex family, excluding the DFP update.

Although there are many quasi-Newton methods, the BFGS method surpasses the others in popularity. It has been believed that this method is more robust than its relatives. Many attempts have been made to answer the question whether the BFGS method is the best quasi-Newton method, at least within the Broyden family. In fact, the search for a quasi-Newton method that is more efficient than the BFGS method began in the 1970s and has not ceased (see, for example, [1, 6–8, 16, 18, 29]).

In particular, Dennis and Wolkowicz [11] derived a new and promising update in the Broyden family by optimizing the measure

$$\omega(B) = \frac{Tr(B)}{n \det(B)^{1/n}}.$$

We call this new update the DW update and the corresponding method the DW method. In the DW update, the parameter ϕ_k in (1) is defined as

$$\phi_k = \frac{1}{\big[b_k/h_k + 1 - b_k^2/(a_k h_k)\big]}, \tag{2}$$

with the notations

$$a_k = y_k^T H_k y_k, \quad b_k = y_k^T s_k, \quad h_k = s_k^T B_k s_k.$$

The corresponding DW update for the inverse Hessian approximation H_k is (see [11]) given by

$$H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{y_k^T s_k} + \big(y_k^T s_k\big) u_k u_k^T,$$

where $u_k = s_k / y_k^T s_k - H_k y_k / y_k^T H_k y_k$.

The numerical results of Dennis and Wolkowicz [11] show that the DW method performs better than the BFGS method. Our tests on some problems with moderate dimensions (see Section 5) confirm their conclusion. This motivates us to study its convergence properties. In Han and Liu [15], we proved that the DW method possesses a self-correcting property, and it is globally convergent when applied to uniformly convex objective functions. According to the Cauchy-Schwarz inequality

$$(b_k)^2 \leq a_k h_k, \tag{3}$$

we have that $\phi_k > 0$. Since it is possible that $\phi_k > 1$, the DW method doesn't necessarily belong to the Broyden convex family. Moreover, if $h_k/b_k \ll 1$, then $\phi_k \approx 0$. Thus the DW update can be very close to the DFP update. These properties of ϕ_k make the convergence analysis very tricky since it is still an open question whether the DFP update is globally convergent with an inexact line search for strongly convex functions (Nocedal [19]). In [29], the convergence of the restricted Broyden family

$$1 - \phi_k \in [0, 1 - \delta], \;\; 1 - \delta \in [0, 1)$$

was analyzed, while the convergence behavior when $1 - \phi_k$ approaches 1 (i.e., near the DFP update) is unknown.

Since most convergence analyses for the quasi-Newton methods in the literature assume that the objective functions are only convex, a natural question to ask is that whether the convergence results in [15] can be extended to cover general convex objective functions. In this paper, we will show that the DW method with the Wolfe line search is globally convergent on convex functions. We will also provide some numerical results to illustrate the superiority of the DW method over the BFGS method.

2. Preliminaries

All of our convergence analysis will be based on the following assumption:

Assumption H. The function f is twice continuously differentiable, convex and bounded below. Moreover the Hessian matrix is bounded

$$\|G(x)\| \leq M$$

for all x in the level set $D = \{x \in R^n : f(x) \leq f(x_0)\}$.

This assumption is mild because for general non-convex objective functions little has been known regarding the convergence of the original quasi-Newton methods due to the difficulty arising from estimating the growth of the magnitude of the eigenvalues of B_k. Some researchers (see for example, [17] and [26]) have considered modifying the quasi-Newton methods and proved global convergence for the modified quasi-Newton methods on non-convex objective functions.

We analyze the DW algorithm with the following Wolfe line search

$$\begin{aligned} f(x_k+\lambda_k d_k) &\le f(x_k)+c_1\lambda_k g_k^T d_k \\ g(x_k+\lambda_k d_k)^T d_k &\ge c_2 g_k^T d_k, \end{aligned}$$

where $c_1 \in (0,1)$ and $c_2 \in [c_1, 1)$. From the second inequality, we obtain

$$y_k^T s_k \ge (1-c_2)\lambda_k\left(-g_k^T d_k\right) = (1-c_2)\left(-g_k^T s_k\right). \tag{4}$$

We will often use this relation in the convergence analysis. Also, we will often use

$$B_k s_k = -\lambda_k g_k, \; s_k^T B_k s_k = \lambda_k\left(-s_k^T g_k\right).$$

The first Lemma is an important property of convex functions. It was proved and used by Powell to analyze the global convergence of the DFP method with the exact line search [24] and of the BFGS method with the Wolfe line search [25] on convex objective functions.

Lemma 1 (*see* Powell [24] *or* [25])**.** *Under the Assumption* (*H*),

$$\frac{\|y_k\|^2}{y_k^T s_k} \le M.$$

In the literature, the convergence of quasi-Newton methods has been investigated by means of estimating the trace and determinant of the Hessian approximation matrix B_k. It's easy to derive the iterative relations between the traces $\mathrm{Tr}(B_{k+1})$ and $\mathrm{Tr}(B_k)$ and between the determinants $\det(B_{k+1})$ and $\det(B_k)$ from (1).

Lemma 2 (*see* Byrd et al. [4] *or* [2])**.**

$$\begin{aligned} \mathrm{Tr}(B_{k+1}) = \mathrm{Tr}(B_k) &- \phi_k\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + \frac{\|y_k\|^2}{y_k^T s_k} + (1-\phi_k)\frac{\|y_k\|^2}{y_k^T s_k}\frac{s_k^T B_k s_k}{y_k^T s_k} \\ &-(1-\phi_k)\frac{2y_k^T B_k s_k}{y_k^T s_k}. \end{aligned} \tag{5}$$

Lemma 3 (*see* Byrd et al. [2] *and* Pearson [22])**.**

$$\det(B_{k+1}) = \det(B_k)\left[\phi_k\frac{y_k^T s_k}{s_k^T B_k s_k} + (1-\phi_k)\frac{y_k^T B_k^{-1} y_k}{y_k^T s_k}\right].$$

The DW algorithm proceeds from any starting point and any positive initial matrix B_1. Since the Wolfe line search always ensures

$$y_k^T s_k > 0,$$

it is easy to see that all the matrices $B_k (k = 2, 3, \ldots)$ generated from the DW algorithm are positive definite (Dennis and Wolkowicz [11]). Thus all the following quantities

$$a_k,\ b_k,\ h_k, \mathrm{Tr}(B_{k+1})$$

are positive.

Since some of our analysis in the later sections will be developed by discussing two cases depending on whether $b_k > h_k$ or not, we define the following:

$$p(j) = \begin{cases} 1, & \text{if } b_j > h_j, \\ 2, & \text{if } b_j \leq h_j. \end{cases}$$

Throughout this paper, we always assume that $\{x_k\}$ is generated by the DW algorithm, where the steplength λ_k satisfies the Wolfe line search.

In the rest of this section, we present some basic lemmas which will be used in the convergence analysis. Some of them can be found in Han and Liu [15], while the proofs of the others are straightforward.

Lemma 4. *Assume that f is bounded below. Then,*

$$\sum_{k=1}^{+\infty} \left(-g_k^T s_k\right) < +\infty.$$

This lemma is a direct result from the first condition of the Wolfe line search.

Lemma 5 (*See* [15])**.**

(*a*) *If* $\phi_k \geq 1$, *then* $|1 - \phi_k| \leq \phi_k$.

(*b*) *If* $0 < \phi_k < 1$, *then* $|1 - \phi_k| \leq \frac{b_k}{h_k}\phi_k$.

(*c*) $\frac{1}{\phi_k} \leq \frac{b_k}{h_k} + 1$.

Lemma 6.

$$\phi_k \geq \frac{1}{2}\left(\frac{h_k}{b_k}\right)^{2-p(k)}.$$

Proof: Consider two cases:

Case (1) $h_k < b_k$.

In this case, $p(k) = 1$, and from (3), we have

$$\phi_k \geq \frac{1}{1 + b_k/h_k} \geq \frac{1}{b_k/h_k + b_k/h_k} = \frac{h_k}{2b_k} = \frac{1}{2}\left(\frac{h_k}{b_k}\right)^{2-p(k)}.$$

Case (2) $h_k \geq b_k$.

In this case, $p(k) = 2$, and from (3), we have

$$\phi_k \geq \frac{1}{1 + b_k/h_k} \geq 1/2 = \frac{1}{2}\left(\frac{h_k}{b_k}\right)^{2-p(k)}.$$

Therefore, this Lemma is true. □

Lemma 7. *For any* $x > 0$,

$$x - 2\ln x \geq 2 - 2\ln 2 > 0,$$
$$x - 1/2\ln x \geq 1/2 + 1/2\ln 2 > 0.$$

Proof: Define

$$q(x) = x - 2\ln x.$$

Note that

$$q'(x) = 1 - \frac{2}{x},\ q''(x) = \frac{2}{x^2} > 0.$$

The first string of inequalities hold. Similarly, the second string of inequalities hold by defining $q(x) = x - 1/2\ln x$. □

Lemma 8 (*See* [15])**.**

$$\det(B_{k+1}) = \det(B_k)\frac{\phi_k a_k}{h_k}. \tag{6}$$

Lemma 9.

$$\det(B_{k+1}) \geq \frac{1}{2}\left(\frac{b_k}{h_k}\right)^{p(k)}\det(B_k).$$

Proof: By Lemma 8 we have

$$\det(B_{k+1}) = \det(B_k)\left(\frac{b_k}{h_k}\right)\left(\frac{a_k}{b_k}\right)\phi_k.$$

From (3), we have

$$\frac{a_k}{b_k} \geq \frac{b_k}{h_k}.$$

Thus

$$\det(B_{k+1}) \geq \left(\frac{b_k}{h_k}\right)^2 \phi_k \det(B_k). \tag{7}$$

Using Lemma 6, we obtain

$$(b_k/h_k)^2\phi_k \geq (b_k/h_k)^2 \times 1/2(h_k/b_k)^{2-p(k)} = \frac{1}{2}(b_k/h_k)^{p(k)}.$$

Substituting this into (7), we see that this lemma is true. □

3. Results under an assumption

In this section, we assume that there exists a constant $\epsilon > 0$ such that

$$\|g_k\| \geq \epsilon. \tag{8}$$

All of the results obtained in this section are derived by assuming that (8) and Assumption (H) hold.

We are going to derive an upper bound for $\mathrm{Tr}(B_{k+1})$. The next two theorems are devoted to estimating $|(1-\phi_k)\frac{s_k^T B_k s_k}{y_k^T s_k}|$ and $|(1-\phi_k)\frac{y_k^T B_k s_k}{y_k^T s_k}|$. They appear in the right hand side of $\mathrm{Tr}(B_{k+1})$ formula in Lemma 2.

Theorem 1. *If*

$$0 < \phi_k < 1,$$

then we have

$$\left|(1-\phi_k)\frac{s_k^T B_k s_k}{y_k^T s_k}\right| \leq c_3.$$

and

$$\left|(1-\phi_k)\frac{y_k^T B_k s_k}{y_k^T s_k}\right| \leq c_4 + \frac{1}{8}\phi_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}.$$

for some positive constants c_3 *and* c_4.

Proof: Since $0 < \phi_k < 1$, from Lemma 5 and the definitions of b_k and h_k we have

$$\left|(1-\phi_k)\frac{s_k^T B_k s_k}{y_k^T s_k}\right| \leq \phi_k \frac{b_k}{h_k}\left|\frac{s_k^T B_k s_k}{y_k^T s_k}\right| = \phi_k < 1. \tag{9}$$

To estimate $|(1-\phi_k)\frac{y_k^T B_k s_k}{y_k^T s_k}|$, we divide our analysis into two cases:

Case (1) $|y_k^T g_k| \leq \lambda_k \|g_k\|^2/8$.

By Lemma 5 we have

$$\begin{aligned}
\left|(1-\phi_k)\frac{y_k^T B_k s_k}{y_k^T s_k}\right| \Bigg/ \left(\phi_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}\right) &\le \phi_k \frac{b_k}{h_k}\frac{\left|y_k^T B_k s_k\right|}{y_k^T s_k} \Bigg/ \left(\phi_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}\right)\\
&= \frac{\left(y_k^T s_k\right)\left|y_k^T B_k s_k\right|\left(s_k^T B_k s_k\right)}{\left(s_k^T B_k s_k\right)\left(y_k^T s_k\right)\|B_k s_k\|^2}\\
&= \frac{\left|y_k^T B_k s_k\right|}{\|B_k s_k\|^2}\\
&= \frac{\left|y_k^T g_k\right|}{\lambda_k \|g_k\|^2}\\
&\le \frac{1}{8}.
\end{aligned}$$

Case (2) $|y_k^T g_k| > \lambda_k \|g_k\|^2/8$.
In this case, from $0 < \phi_k < 1$, the Cauchy-Schwarz inequality and Lemma 1 we have

$$\begin{aligned}
\left|(1-\phi_k)\frac{y_k^T B_k s_k}{y_k^T s_k}\right| &\le \left|\frac{y_k^T B_k s_k}{y_k^T s_k}\right|\\
&= \frac{\left|y_k^T B_k s_k\right|^2}{y_k^T s_k \left|y_k^T B_k s_k\right|}\\
&\le \frac{\|y_k\|^2 \|B_k s_k\|^2}{y_k^T s_k \left|y_k^T B_k s_k\right|}\\
&= \left(\frac{\|y_k\|^2}{y_k^T s_k}\right)\left(\frac{\lambda_k \|g_k\|^2}{\left|y_k^T g_k\right|}\right)\\
&\le M\left(\frac{\lambda_k \|g_k\|^2}{\left|y_k^T g_k\right|}\right) \le 8M.
\end{aligned}$$

Summarizing Case (1) and Case (2), we have

$$\left|(1-\phi_k)\frac{y_k^T B_k s_k}{y_k^T s_k}\right| \le 8M + \frac{1}{8}\phi_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}.$$

Taking $c_4 = 8M$ in the above inequality and $c_3 = 1$ in (9) completes the proof. □

Theorem 2. *If*

$$\phi_k \ge 1,$$

then there exist positive constants c_5 and c_6 such that

$$\left|(1-\phi_k)\frac{s_k^T B_k s_k}{y_k^T s_k}\right| \le c_5\left(-g_k^T s_k\right)\phi_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}$$

and

$$\left|(1-\phi_k)\frac{y_k^T B_k s_k}{y_k^T s_k}\right| \le c_6\sqrt{\left(-g_k^T s_k\right)}\phi_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}.$$

Proof: First we estimate $|(1-\phi_k)\frac{s_k^T B_k s_k}{y_k^T s_k}|$.

From Lemma 5, we have

$$\begin{aligned}\left|(1-\phi_k)\frac{s_k^T B_k s_k}{y_k^T s_k}\right| \Big/ \left(\phi_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}\right) &\le \phi_k \frac{s_k^T B_k s_k}{y_k^T s_k} \Big/ \left(\phi_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}\right)\\ &= \frac{\left(s_k^T B_k s_k\right)\left(s_k^T B_k s_k\right)}{\|B_k s_k\|^2 y_k^T s_k}\\ &= \frac{\left(-s_k^T g_k\right)^2}{\|g_k\|^2 y_k^T s_k}\\ &\le \frac{\left(-s_k^T g_k\right)^2}{\epsilon^2(1-c_2)\left(-s_k^T g_k\right)} = c_5\left(-s_k^T g_k\right),\end{aligned}$$

where $c_5 = 1/[\epsilon^2(1-c_2)]$.

Therefore,

$$\left|(1-\phi_k)\frac{s_k^T B_k s_k}{y_k^T s_k}\right| \le c_5\phi_k\left(-s_k^T g_k\right)\frac{\|B_k s_k\|^2}{s_k^T B_k s_k}. \tag{10}$$

We then estimate $|(1-\phi_k)\frac{y_k^T B_k s_k}{y_k^T s_k}|$.

By Lemma 5, the Cauchy-Schwarz inequality and Lemma 1 we deduce

$$\begin{aligned}\left|(1-\phi_k)\frac{y_k^T B_k s_k}{y_k^T s_k}\right| \Big/ \left(\phi_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}\right) &\le \phi_k \frac{\left|y_k^T B_k s_k\right|}{y_k^T s_k} \Big/ \left(\phi_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}\right)\\ &= \frac{\left|y_k^T B_k s_k\right|\left(s_k^T B_k s_k\right)}{y_k^T s_k\|B_k s_k\|^2}\\ &\le \frac{\|y_k\|\left(s_k^T B_k s_k\right)}{y_k^T s_k\|B_k s_k\|}\\ &\le \frac{\sqrt{M y_k^T s_k}\left(s_k^T B_k s_k\right)}{y_k^T s_k\|B_k s_k\|}\end{aligned}$$

$$
\begin{aligned}
&= \frac{\left(-s_k^T g_k\right)}{\|g_k\|}\sqrt{\frac{M}{y_k^T s_k}} \\
&\leq \frac{\left(-s_k^T g_k\right)}{\epsilon}\sqrt{\frac{M}{(1-c_2)\left(-s_k^T g_k\right)}} \\
&= c_6\sqrt{\left(-s_k^T g_k\right)},
\end{aligned}
$$

where $c_6 = \frac{1}{\epsilon}\sqrt{\frac{M}{1-c_2}}$.

Therefore,

$$
\left|(1-\phi_k)\frac{y_k^T B_k s_k}{y_k^T s_k}\right| \leq c_6\phi_k\sqrt{\left(-s_k^T g_k\right)}\frac{\|B_k s_k\|^2}{s_k^T B_k s_k}. \tag{11}
$$

□

Theorem 3. *There exists a constant* $c_7 > 0$ *and an index* $K > 0$ *such that for all* $k \geq K$,

$$
\operatorname{Tr}(B_{k+1}) \leq \operatorname{Tr}(B_k) - \frac{\phi_k}{2}\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + c_7.
$$

Proof: By Lemmas 1 and 2 and Theorems 1 and 2, we have

$$
\begin{aligned}
\operatorname{Tr}(B_{k+1}) \leq\ & \operatorname{Tr}(B_k) - \phi_k\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + \frac{\|y_k\|^2}{y_k^T s_k} + \left|(1-\phi_k)\frac{\|y_k\|^2}{y_k^T s_k}\frac{s_k^T B_k s_k}{y_k^T s_k}\right| \\
& + \left|(1-\phi_k)\frac{2y_k^T B_k s_k}{y_k^T s_k}\right| \\
\leq\ & \operatorname{Tr}(B_k) - \phi_k\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + M + M\left|(1-\phi_k)\frac{s_k^T B_k s_k}{y_k^T s_k}\right| + \left|(1-\phi_k)\frac{2y_k^T B_k s_k}{y_k^T s_k}\right| \\
\leq\ & \operatorname{Tr}(B_k) - \phi_k\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + M + M\left[c_3 + c_5\left(-g_k^T s_k\right)\phi_k\frac{\|B_k s_k\|^2}{s_k^T B_k s_k}\right] \\
& + \left[2c_4 + \frac{1}{4}\phi_k\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + 2c_6\sqrt{-g_k^T s_k}\phi_k\frac{\|B_k s_k\|^2}{s_k^T B_k s_k}\right] \\
=\ & \operatorname{Tr}(B_k) - \left[3/4 - c_8\left(-s_k^T g_k + \sqrt{-s_k^T g_k}\right)\right]\phi_k\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + c_7.
\end{aligned} \tag{12}
$$

where $c_7 = M + Mc_3 + 2c_4$ and $c_8 = Mc_5 + 2c_6$.

By Lemma 4, we know

$$
\lim_{k\to\infty} c_8\left[-s_k^T g_k + \sqrt{-s_k^T g_k}\right] = 0,
$$

i.e., there exists an index $K > 0$ such that $k \geq K$,

$$c_8\left(- s_k^T g_k + \sqrt{-s_k^T g_k}\right) < 1/4,$$

and then from (12), for all $k \geq K$,

$$\mathrm{Tr}(B_{k+1}) \leq \mathrm{Tr}(B_k) - \frac{1}{2}\phi_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + c_7. \qquad \Box$$

From this theorem, there exists a positive constant c_9 such that for all $k \geq 1$,

$$\begin{aligned} \mathrm{Tr}(B_{k+1}) &\leq \mathrm{Tr}(B_k) + c_9, \\ \frac{\phi_k}{2}\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} &\leq [\mathrm{Tr}(B_k) - \mathrm{Tr}(B_{k+1})] + c_9, \end{aligned}$$

which give two immediate corollaries:

Corollary 1. *There exists a constant* $c_{10} > 0$ *such that*

$$\mathrm{Tr}(B_{k+1}) \leq c_{10} k.$$

Corollary 2. *There exists a constant* $c_{11} > 0$ *such that*

$$\sum_{j=1}^{k} \phi_j \frac{\|B_j s_j\|^2}{s_j^T B_j s_j} \leq c_{11} k.$$

4. Convergence analysis

Our aim is to prove the global convergence

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \tag{13}$$

To prove this, we first assume that (13) doesn't hold, i.e., we can assume that (8) is true. Thus all the results in the preceding sections apply. Then we derive a contradiction, which implies that (13) must hold.

Lemma 10. *Assume that* (8) *holds. Then, there exists a positive constant* c_{12} *such that*

$$\sum_{j=1}^{k} \frac{1}{\left(-g_j^T s_j\right)} \left(\frac{h_j}{b_j}\right)^{3-p(j)} \leq c_{12} k.$$

Proof:

$$\begin{aligned}\frac{\|B_k s_k\|^2}{s_k^T B_k s_k} \Big/ \frac{h_k}{b_k} &= \frac{\left(y_k^T s_k\right)\|B_k s_k\|^2}{\left(s_k^T B_k s_k\right)^2} \\ &\geq \frac{(1-c_2)\left(-g_k^T s_k\right)\|B_k s_k\|^2}{\left(s_k^T B_k s_k\right)^2} \\ &= \frac{(1-c_2)\|g_k\|^2}{\left(-g_k^T s_k\right)} \\ &\geq \frac{(1-c_2)\epsilon^2}{\left(-g_k^T s_k\right)} = \frac{1}{\left(-g_k^T s_k\right)c_5}.\end{aligned}$$

This and Lemma 6 imply

$$\begin{aligned}\phi_k \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} &\geq \phi_k \frac{h_k}{b_k} \frac{1}{\left(-g_k^T s_k\right)c_5} \\ &\geq 1/2(h_k/b_k)^{2-p(k)} \times \frac{h_k}{b_k} \frac{1}{\left(-g_k^T s_k\right)c_5} \\ &= c_{13} \frac{1}{\left(-g_k^T s_k\right)} \left(\frac{h_k}{b_k}\right)^{3-p(k)}.\end{aligned}$$

By using this and Corollary 2, we have

$$\sum_{j=1}^{k} \frac{1}{\left(-g_j^T s_j\right)} \left(\frac{h_j}{b_j}\right)^{3-p(j)} \leq 1/c_{13} \sum_{j=1}^{k} \phi_j \frac{\|B_j s_j\|^2}{s_j^T B_j s_j} \leq c_{12} k, \tag{14}$$

where $c_{12} = c_{11}/c_{13}$. □

Lemma 11. *Assume that* (8) *holds. Then there exists a positive constant* c_{14} *such that*

$$\sum_{j=1}^{k} \ln\left(\frac{h_j}{b_j}\right)^{p(j)} \geq -c_{14} k.$$

Proof: From Lemma 9, we deduce

$$\begin{aligned}\det\,(B_{k+1}) &\geq \frac{1}{2}\left(\frac{b_k}{h_k}\right)^{p(k)} \det(B_k) \\ &\geq \cdots\cdots \\ &\geq \frac{1}{2^k} \prod_{j=1}^{k} \left(\frac{b_j}{h_j}\right)^{p(j)} \det(B_1),\end{aligned}$$

i.e.,

$$\prod_{j=1}^{k}\left(\frac{b_j}{h_j}\right)^{p(j)} \le 2^k\,\frac{\det(B_{k+1})}{\det(B_1)}. \tag{15}$$

Using the geometric-arithmetic inequality and Corollary 1, we have

$$\det(B_{k+1}) \le [\mathrm{Tr}(B_{k+1})/n]^n \le (c_{10}k/n)^n.$$

Substituting this into (15), we have

$$\prod_{j=1}^{k}\left(\frac{b_j}{h_j}\right)^{p(j)} \le \frac{2^k(c_{10}k/n)^n}{\det(B_1)} = c_{15}k^n2^k.$$

Taking the natural logarithm, we have

$$\sum_{j=1}^{k}\ln\left(\frac{b_j}{h_j}\right)^{p(j)} \le \ln\bigl(c_{15}k^n2^k\bigr) \le c_{14}k.$$

for some positive constant c_{14}. Multiplying the above string of inequalities by -1, we obtain

$$\sum_{j=1}^{k}\ln\left(\frac{h_j}{b_j}\right)^{p(j)} \ge -c_{14}k.$$ □

Theorem 4. *Suppose that the assumption* (H) *holds. Let x_1 be any starting point, and B_1 be any symmetric positive definite matrix. Assume that $\{x_k\}$ is generated by the DW algorithm with the Wolfe line search. Then*

$$\lim_{k\to\infty}\inf\|g_k\| = 0.$$

Proof: Assume that by contradiction,

$$\lim_{k\to\infty}\inf\|g_k\| > 0,$$

i.e., we can assume that (8) holds. Thus Lemma 10 and Lemma 11 hold.

Using Lemma 10 and Lemma 11, we have

$$\sum_{j=1}^{k}\left[\frac{1}{\bigl(-g_j^Ts_j\bigr)}\left(\frac{h_j}{b_j}\right)^{3-p(j)} - \ln\left(\frac{h_j}{b_j}\right)^{p(j)}\right] \le (c_{12}+c_{14})k = c_{16}k.$$

Denote

$$\xi_j = \frac{1}{\bigl(-g_j^Ts_j\bigr)}\left(\frac{h_j}{b_j}\right)^{3-p(j)}.$$

Then

$$\begin{aligned}
&\frac{1}{\left(-g_j^T s_j\right)}\left(\frac{h_j}{b_j}\right)^{3-p(j)} - \ln\left(\frac{h_j}{b_j}\right)^{p(j)} \\
&= \xi_j - \ln\left(\frac{h_j}{b_j}\right)^{p(j)} - \frac{p(j)}{3-p(j)}\ln\frac{1}{\left(-s_j^T g_j\right)} + \frac{p(j)}{3-p(j)}\ln\frac{1}{\left(-s_j^T g_j\right)} \\
&= \xi_j - \left(\frac{p(j)}{3-p(j)}\right)\ln\left[\left(\frac{h_j}{b_j}\right)^{3-p(j)}\frac{1}{\left(-s_j^T g_j\right)}\right] + \frac{p(j)}{3-p(j)}\ln\frac{1}{\left(-s_j^T g_j\right)} \\
&= \xi_j - \left(\frac{p(j)}{3-p(j)}\right)\ln \xi_j + \frac{p(j)}{3-p(j)}\ln\frac{1}{\left(-s_j^T g_j\right)}.
\end{aligned}$$

Therefore

$$\sum_{j=1}^{k}\left[\xi_j - \left(\frac{p(j)}{3-p(j)}\right)\ln \xi_j + \frac{p(j)}{3-p(j)}\ln\frac{1}{\left(-s_j^T g_j\right)}\right] \le c_{16}k. \tag{16}$$

Since $\frac{p(j)}{3-p(j)} = 1/2$ or 2, by Lemma 7, we have

$$\xi_j - \left(\frac{p(j)}{3-p(j)}\right)\ln \xi_j \ge \min\{2 - 2\ln 2, 1/2 + 1/2\ln 2\} > 0.$$

Substituting this into (16), we have

$$\sum_{j=1}^{k}\frac{p(j)}{3-p(j)}\ln\frac{1}{\left(-s_j^T g_j\right)} \le c_{16}k.$$

Since $\frac{p(j)}{3-p(j)} \ge 1/2$, we have

$$\sum_{j=1}^{k}\ln\frac{1}{\left(-s_j^T g_j\right)} \le 2c_{16}k,$$

i.e., for all k,

$$\prod_{j=1}^{k}\frac{1}{\left(-s_j^T g_j\right)} \le \exp(2c_{16}k). \tag{17}$$

By Lemma 4, we know that there exists a constant $c_{17} > 0$ such that for all k,

$$\sum_{j=1}^{k}\left(-s_j^T g_j\right) \le c_{17}.$$

Using the geometric-arithmetic inequality gives

$$\prod_{j=1}^{k}\left(-s_j^T g_j\right) \leq \left[\sum_{j=1}^{k}\left(-s_j^T g_j\right) \Big/ k\right]^k \leq (c_{17}/k)^k,$$

i.e.,

$$\prod_{j=1}^{k}\frac{1}{\left(-s_j^T g_j\right)} \geq (k/c_{17})^k.$$

Combining this with (17), we obtain that for all k,

$$(k/c_{17})^k \leq \exp(c_{16}k),$$

i.e.,

$$k \leq c_{17}\ \exp(c_{16}),$$

which is a contradiction. Therefore

$$\lim_{k\to\infty}\inf \|g_k\| = 0. \qquad \square$$

5. Numerical experiments

In this section we report some numerical experiments on the DW method.

The first experiment is to test the DW method against the BFGS method on some unconstrained optimization problems with moderate dimensions. This can be viewed as a supplement to the numerical results of Dennis and Wolkowicz [11], since their tests were focused on problems of low dimensions. The sketch of our implementation of the DW method is as follows.

```
Step 1. Choose constants $c_1 = 10^{-4}$, $c_2 = 10^{-1}$.
        Select any starting point $x_1$, and any
        symmetric and positive definite matrix $H_1$.
        Compute $g_1 = g(x_1)$. Set $k := 1$.
Step 2. Compute the search direction $d_k = -H_k g_k$.
Step 3. Compute the steplength $\lambda_k$ satisfing the following strong Wolfe line search:
        $f(x_k + \lambda_k d_k) \leq f(x_{k-1}) + c_1 \lambda_k g_k^T d_k$,
        $|g_{k+1}^T d_k| \geq c_2 |g_k^T d_k|$.
Step 4. Iterate $x_k$ by
        $x_{k+1} = x_k + \lambda_k d_k$.
        Compute the gradient $g_{k+1} = g(x_{k+1})$ and the function value $f(x_{k+1})$.
```

Table 1. List of the problems.

Problem	Name	Dimension
1	Calculus variation 1	100
2	Calculus variation 2	100
3	Calculus variation 3	100
6	Generalized Rosenbrock	100
28	Extended Powell singular	100
42	Extended ENGVL1	100
43	Extended F & R	100
48	Extended Rosenbrock	100
49	Extended Powell	100
50	Tridiagonal 2	100
51	Modified Trigonometric	100
52	Modified Penalty 1	100

Compute the vectors
$s_k = x_{k+1} - x_k, \quad y_k = g_{k+1} - g_k,$
and $v_k = \frac{y_k}{y_k^T s_k} - \frac{B_k s_k}{s_k^T B_k s_k}$.

Step 5. Test the termination condition. If $\|g_{k+1}\|$ satisfies
$\|g_{k+1}\| \le \epsilon[1 + |f(x_{k+1})|], \quad \epsilon = 10^{-5}$,
then **STOP** ; otherwise, go to Step 6.

Step 6. Update the inverse Hessian approximation H_k by the DW formula. Set $k := k + 1$, go to Step 2.

Table 1 lists our test problems which are from [14]. The experiments were done on a SPARC 5 Sun Workstation in double precision with f 77 compiler. In our tests, we chose $H_1 = \frac{y_1^T s_1}{y_1^T y_1} I$ and $10^3 x_1$ and $10^5 x_1$ as the initial points, where x_1 is the initial point in the collection [14] of test problems.

We summarize our numerical results in Table 2 (corresponding to initial point $10^3 x_1$) and Table 3 (corresponding to starting point $10^5 x_1$). In these tables, *it* denotes the number of iterations and $f - g$ denotes the number of function-gradient evaluations. * indicates that line search couldn't find an appropriate stepsize after 20 function evaluations. For the DW method, we also record the smallest and largest observed values of the parameter $\theta_k = 1 - \phi_k$ during the computation. We should mention that if $\theta_k = 1$ then it corresponds to the DFP update and $\theta_k = 0$ corresponds to the BFGS update.

We can conclude from Tables 2 and 3 that the DW method outperforms the BFGS method on the test problems in Table 1. The range of θ_k tells us that the DW update almost surely goes beyond the Broyden convex family, and can be close to the DFP update. We believe this contributes to its superiority over the BFGS method.

Table 2. Comparison of the DW method with BFGS method (starting point $10^3 x_1$).

	BFGS	DW	
Problem	$it/f-g$	$it/f-g$	[min θ, max θ]
1	482/536	481/515	[−0.2430D+00 0.9630D+00]
2	173/174	165/166	[−0.2135D+00 0.7925D−01]
3	71/86	39/48	[−0.2454D+02 0.8685D+00]
6	779/795	740/757	[−0.8829D+00 0.9240D+00]
28	76/84	65/75	[−0.1840D+01 0.8464D+00]
42	105/109	100/104	[−0.8454D+00 0.4797D+00]
43	94/96	85/90	[−0.1035D+01 0.8973D−02]
48	62/79	63/80	[−0.4379D+01 0.9490D+00]
49	76/84	65/75	[−0.1840D+01 0.8464D+00]
50	54/61	53/55	[−0.3979D−12 0.4849D+00]
51	155/162	147/150	[−0.1969D+01 0.8600D+00]
52	14/16	13/15	[−0.8457D+00 0.4619D−13]

Our next experiment tested the DW method against the BFGS method on the following quartic function with the dimension $n = 100$ (see [5])

$$\min\ f(x) = \frac{1}{2}(x-1)^T D(x-1) + \frac{\sigma}{4}[(x-1)^T B(x-1)]^2 + 1,$$

where D is a positive definite diagonal matrix, σ is a parameter that controls the deviation from quartic, and

$$B = U^T U,\ U = \begin{pmatrix} 1 & \cdot & \cdot & \cdot & 1 \\ & \cdot & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ & & & & 1 \end{pmatrix}.$$

The starting point was chosen as $(-1)^i \times 50 (i = 1, \ldots, 100)$. The solution is $x = (1, \ldots, 1)$. As iterates approach the solution, D will dominate the Hessian of the objective function. We chose three different D's with the following structure:

$$D = \mathrm{diag}[(1+\epsilon)^{-50}, (1+\epsilon)^{-49}, \ldots, (1+\epsilon)^{49}],$$

where ϵ was chosen from

$$0,\ 0.1,\ 0.2.$$

Table 3. Comparison of the DW method with BFGS method (starting point $10^5 x_1$).

Problem	BFGS method $it/f-g$	DW method $it/f-g$	DW method $[\min\theta, \max\theta]$
1	343/533	340/470	[−0.1957D+00 0.9838D+00]
2	713/720	682/691	[−0.1072D+01 0.2705D+00]
3	*	*	*
6	2695/2873	2393/2460	[−0.1188D+01 0.9832D+00]
28	198/223	162/185	[−0.1884D+01 0.7950D+00]
42	270/279	260/268	[−0.1709D+01 0.5067D+00]
43	*	*	*
48	265/312	271/327	[−0.4166D+01 0.9701D+00]
49	196/216	169/186	[−0.1884D+01 0.7950D+00]
50	38/47	35/42	[−0.1137D−12 0.5568D+00]
51	80/81	86/87	[−0.1143D+01 0.2167D+00]
52	39/45	38/44	[−0.8978D+00 −0.1164D−09]

Table 4. BF GS method.

	ϵ		
σ	0.0 $it/f-g$	0.1 $it/f-g$	0.2 $it/f-g$
0.0	2/6	504/507	1084/1087
0.01	497/498	1792/1794	1732/1733
0.02	516/517	1842/1845	1782/1783

They give the matrix D the condition numbers of 1, 1.2528×10^4, and 6.9015×10^7. We also chose 3 different values of σ:

$$\sigma = 0,\ 0.01,\ 0.02.$$

Table 4 and Table 5 summarize our results of the BFGS method and the DW method respectively. From these two tables we see that the DW method performs better than the BFGS method in all the cases for this quartic problem. Even when D is very ill-conditioned ($\epsilon = 0.2$), the DW method is superior to the BFGS method.

It has been believed that a good member of the Broyden family should possess the property that it can correct both large and small eigenvalues of the approximation matrix B_k efficiently (see, for example, Byrd, et al. [3], Han and Liu [13], and Nocedal [15]). This property is called the self-correction property. A main drawback of the DFP update is that it can not correct large eigenvalues efficiently. In our last experiment, we tested and compared the abilities of the DW update, the BFGS update, and the DFP update on correcting both large and small eigenvalues of the approximation matrx B_k. In this test, we

Table 5. DW method.

	ϵ		
	0.0	0.1	0.2
σ	$it/f-g$	$it/f-g$	$it/f-g$
0.0	2/6	477/480	1043/1046
0.01	464/465	1742/1743	1680/1681
0.02	482/483	1765/1768	1765/1768

Table 6. Self-correction property of the DW, BFGS, and DFP updates.

q	DFP	BFGS	DW
10^{-6}	5.050337×10^{-1}	5.000005×10^{-1}	5.033708×10^{-1}
10^{-5}	5.049531×10^{-1}	5.000051×10^{-1}	5.033003×10^{-1}
10^{-4}	5.051058×10^{-1}	5.000510×10^{-1}	5.034459×10^{-1}
10^{-3}	5.053574×10^{-1}	5.005097×10^{-1}	5.037267×10^{-1}
10^{-2}	5.098492×10^{-1}	5.050954×10^{-1}	5.082210×10^{-1}
10^{-1}	5.546244×10^{-1}	5.508733×10^{-1}	5.531494×10^{-1}
10^{0}	1.000000×10^{0}	1.000000×10^{0}	1.000000×10^{0}
10^{1}	5.455331×10^{0}	5.418495×10^{0}	5.412108×10^{0}
10^{2}	4.998156×10^{1}	4.951908×10^{1}	4.950946×10^{1}
10^{3}	4.957219×10^{2}	4.905210×10^{2}	4.905111×10^{2}
10^{4}	4.950998×10^{3}	4.900520×10^{3}	4.900510×10^{3}
10^{5}	4.950239×10^{4}	4.900052×10^{4}	4.900051×10^{4}
10^{6}	4.949195×10^{5}	4.900005×10^{5}	4.900005×10^{5}

chose the following quadratic function

$$f(x) = \frac{1}{2}x^T x$$

with the dimension $n = 100$.

We chose B_1 as the diagonal matrix with the first 50 diagonal entries being q and the last 50 diagonal entries being 1. q was chosen as $10^{-6}, 10^{-5}, \ldots, 10^5, 10^6$, which can cause that B_1 has small, moderate, or large eigenvalues. We randomly chose s_1 and used each of the DW update, BFGS update, and the DFP update to update B_1 once. The eigenvalues of the resulted matrices B_2^{DW}, B_2^{BFGS}, and B_2^{DFP} were found. The averages of the eigenvalues of B_2^{DW}, B_2^{BFGS}, and B_2^{DFP} were evaluated.

For each case of the above procedure, we ran our program on 10 randomly chosen s_1's. We then took the mean of ten averages of eigenvalues obtained by each update and report the results in Table 6. For the above quadratic function and the matrix B_1, the closer the mean is to 1, the more efficient the self-correction of the update is.

We see from Table 6 that the DW update can correct both large and small eigenvalues of B_k more efficiently than the BFGS update. We believe that this is a main reason for the superiority of the DW method over the BFGS method. Although the DFP update can correct small eigenvalues more efficiently than both the BFGS and the DW updates, it is obvious from Table 6 that the DFP update works much less efficiently than the BFGS and the DW updates in correcting large eigenvalues. Since difficulties most often arise due to large eigenvalues (see Byrd, et al. [3] and Nocedal [15]), the lack of efficiently correcting large eigenvalues of the DFP update determines its inefficiency in optimization calculation.

Acknowledgment

We are grateful to the referees for their insightful comments and suggestions. In particular, we thank them for pointing out some mistakes in the proofs of Theorem 1 and Lemma 11 of the earlier version of this paper, which has led to the reformulation and the correct proofs of Theorem 1 and Lemma 11. We also thank Professor Jorge Nocedal of Northwestern University for his insightful suggestions.

References

1. M. Al-Baali, “Highly efficient Broyden methods of minimization with variable parameter,” Optimization Methods and Software, vol. 1, pp. 301–310, 1992.
2. R.H. Byrd, D.C. Liu, and J. Nocedal, “On the behavior of Broyden’s class of quasi-Newton methods,” SIAM J. Opt., vol. 2, pp. 533–557, 1992.
3. R.H. Byrd and J. Nocedal, “A tool for the analysis of quasi-Newton methods with application to unconstrained minimization,” SIAM J. Numer. Anal., vol. 26, pp. 727–739, 1989.
4. R.H. Byrd, J. Nocedal, and Y. Yuan, “Global convergence of a class of quasi-Newton methods on convex problems,” SIAM J. Numer. Anal., vol. 24, pp. 1171–1191, 1987.
5. R.H. Byrd, J. Nocedal, and C. Zhu, “Towards a discrete Newton method with memory for large-scale optimization,” Technical Report OTC 95/01, 1996.
6. A.R. Conn, N.I.M. Gould, and Ph. L. Toint, “Convergence of quasi-Newton matrices generated by the symmetric rank one update,” Math. Prog., vol. 2, pp. 177–195, 1991.
7. M. Contreras and R.A. Tapia, “Sizing the BFGS and DFP updates: Numerical study,” JOTA, vol. 78, pp. 93–108, 1993.
8. W.C. Davidon, “Optimally conditioned optimization algorithms without line searches,” Math. Prog, vol. 9, pp. 1–30, 1975.
9. J.E. Dennis Jr., D.M. Gay, and R.E. Welsch, “An adaptive nonlinear least-squares algorithm,” TOMS, vol. 7, pp. 348–368, 1981.
10. J.E. Dennis Jr. and R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall: Englewood Cliffs, NJ, 1983 (in English). Mir Publishing Office, Moscow, 1988 (in Russian).
11. J.E. Dennis Jr. and H. Wolkowicz, “Sizing and least-change secant methods,” SIAM J. Numer. Anal., vol. 30, pp. 1291–1314, 1993.
12. R. Fletcher, “A new variational result for quasi-Newton formulae,” SIAM J. Opt., vol. 1, pp. 18–21, 1991.
13. R. Fletcher, Practical Methods of Optimization, 2nd ed., John Wiley and Sons: New York, 1987.
14. J.C. Gilbert and J. Nocedal, “Global convergence properties of conjugate gradient methods for optimization,” SIAM J. Optimization, vol. 2, pp. 21–42, 1992.
15. L.X. Han and G.H. Liu, “Global analysis of the Dennis-Wolkowicz least-change secant algorithm,” SIAM J. Opt., vol. 8, pp. 813–832, 1998.

16. H. Khalfan, R.H. Byrd, and R.B. Schnabel, "A theoretical and experimental study of the symmetric rank one update," University of Colorado, Boulder, Technical Report CU-CS-489-90, 1990.
17. D. Li and M. Fukushima, "On the global convergence of BFGS method for nonconvex unconstrained optimization problems," Department of Applied Mathematics and Physics, Kyoto University, Technical Report 99007, 1999.
18. L. Luksan, "Computational experience with known variable metric updates," JOTA, vol. 83, pp. 27–47, 1994.
19. J. Nocedal, "Theory of algorithms for unconstrained optimization," in Acta Numerica, A. Iserles, (Ed.), Cambridge University Press: Cambridge, England, 1992, pp. 199–242.
20. S.S. Oren and D.G. Luenberger, "Self-scaling veriable metric (SSVM) algorithms, Part I. Criteria and sufficient conditions for scaling a class of algorithms," Management Sci., vol. 20, pp. 845–862, 1974.
21. S.S. Oren and E. Spedicato, "Optimal conditioning of self-scaling variable metric algorithms," Math. Prog., vol. 10, pp. 70–90, 1976.
22. J.D. Pearson, "Variable metric methods of minimization," The Computer J., vol. 12, pp. 171–178, 1969.
23. K.H. Phua and D.F. Shanno, "Numerical comparison of several variable metric algorithms," JOTA, vol. 25, pp. 507–518, 1978.
24. M.J.D. Powell, "Some properties of the variable metric algorithm," in Numerical Methods for Nonlinear Optimization, F.A. Lootsma, (Ed.), Academia Press: London, 1972.
25. M.J.D. Powell, "Some global convergence properties of a variable metric algorithm for minimization without exact line searches in Nonlinear Programming," in SIAM-AMS Proceedings, 1976, vol. IX., R.W. Cottle and C.E. Lemke (Eds.), American Mathematical Society: Providence, RI.
26. D. Pu and W. Tian, "A class of modified Broyden algorithms," J. of Computational Mathematics, vol. 12, pp. 366–379, 1994.
27. J. Werner, "Global convergence of quasi-Newton methods with practical line searches," NAM-Bericht Nr. 67, Marz, Technical Report, 1989.
28. H. Wolkowicz, "Measures for symmetric rank-one updates," Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Tech. Report CORR 90-03, 1990.
29. Y. Zhang and R.P. Tewarson, "Quasi-Newton algorithms with updates from the pre-convex part of Broyden's family," IMA J. Numer. Anal., vol. 8, pp. 487–509, 1988.