Convergence of DFP algorithm*

YUAN Yaxiang (袁亚湘)

(State Key Laboratory of Scientific and Engineering Computing, ICMSEC, Chinese Academy of Sciences, Beijing 100080, China)

Received November 8, 1994; revised March 15, 1995

Abstract The DFP method is one of the most famous numerical algorithms for unconstrained optimization. For uniformly convex objective functions convergence properties of the DFP method are studied. Several conditions that can ensure the global convergence of the DFP method are given.

Keywords: global convergence, DFP algorithm, line search.

1 DFP algorithm

The DFP algorithm is the first quasi-Newton method for unconstrained optimization.

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

The method was given by ref. [2], and then modified and stated as follows by reference [5].

Algorithm 1.1 (*The Davidon-Fletcher-Powell algorithm*).

Step 0. Given $x_1 \in \mathscr{R}^n$; $B_1 \in \mathscr{R}^{n\times n}$ positive definite;
$k := 1$.

Step 1. Compute $g_k = \nabla f(x_k)$;
if $g_k = 0$ then stop;
set $d_k = -B_k^{-1} g_k$;

Step 2. Carry out a line search along d_k, getting $\alpha_k > 0$;
set $x_{k+1} = x_k + \alpha_k d_k$;

Step 3. Set

$$B_{k+1} = B_k - \frac{B_k s_k y_k^T + y_k s_k^T B_k}{s_k^T y_k} + \left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right) \frac{y_k y_k^T}{s_k^T y_k}, \tag{1.2}$$

where

$$s_k = \alpha_k d_k, \tag{1.3}$$

* This work was partially supported by the project "large-scale scientific computing" of State Commission of Science and Technology, China.

$$y_k = g_{k+1} - g_k. \tag{1.4}$$

Step 4. $k := k+1$; go to Step 1.

The line search in Step 2 requires step lengths d_k to satisfy certain line search conditions. If exact line search is used, α_k satisfies

$$f(x_k + \alpha_k d_k) = \min_{\alpha > 0} f(x_k + \alpha d_k). \tag{1.5}$$

For inexact line search, one normally requires that

$$f(x_k + \alpha_k d_k) \leqslant f(x_k) + c_1 \alpha_k d_k^T g_k \tag{1.6}$$

and

$$d_k^T \nabla f(x_k + \alpha_k d_k) \geqslant c_2 d_k^T g_k, \tag{1.7}$$

where $c_1 \leqslant c_2$ are two constants in (0, 1). Usually $c_1 \leqslant 0.5$. Condition (1.7) implies that

$$s_k^T y_k \geqslant -(1 - c_2) s_k^T g_k. \tag{1.8}$$

The DFP method is one of the two most famous variable metric methods. The other is the BFGS method, which is the same as the DFP method except that (1.2) is replaced by

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k}. \tag{1.9}$$

Powell[6] showed that if $f(x)$ is a convex function and if line searches are exact in all iterations, then the DFP algorithm will either stop at a global minimum or generate a sequence that converges to a global minimum. For the inexact line search, Powell[7] showed that the BFGS method is globally convergent. Byrd, Nocedal and Yuan[1] showed the global convergence of all methods in the convex Broyden family except the DFP method. However, the global convergence of the DFP method with inexact line searches has been an open question for the last twenty years. In this paper, several conditions are given to show the global convergence of the DFP method. And it is hoped that our analyses will help us to find the correct answer for the convergence of the DFP method.

For a general line search algorithm, one can show the following result:

Lemma 1.1. *Let x_k $(k=1,\cdots)$ be generated by any line search algorithm in the form of*

$$x_{k+1} = x_k + \alpha_k d_k. \tag{1.10}$$

Assume that the inexact line search conditions (1.6) *and* (1.7) *are satisfied in all iterations. If $f(x_k)$ are bounded below, then*

$$\sum_{k=1}^{\infty} -\alpha_k d_k^T g_k < \infty. \tag{1.11}$$

If x_k are all in a closed convex set Ω in which $\nabla f(x)$ is uniformly Lipschitz continuous, then

$$\sum_{k=1}^{\infty} \cos^2\theta_k \|g_k\|_2^2 < \infty, \tag{1.12}$$

where θ_k is the angle between $-g_k$ and d_k; namely,

$$\cos\theta_k = \frac{-g_k^T d_k}{\|g_k\|_2 \|d_k\|_2}. \tag{1.13}$$

Proof. It follows from (1.6) that

$$\sum_{k=1}^{\infty} -\alpha_k d_k^T g_k \leqslant \sum_{k=1}^{\infty} [f(x_k) - f(x_{k+1})]/c_1$$

$$\leqslant c_1^{-1}[f(x_1) - \lim_{k\to\infty} f(x_k)] < \infty. \tag{1.14}$$

Due to our assumptions, there exists a positive constant M such that

$$s_k^T y_k \leqslant M\|s_k\|_2^2. \tag{1.15}$$

Thus, it follows from the above inequality and (1.7) that

$$\sum_{k=1}^{\infty} \cos^2\theta_k \|g_k\|_2^2 = \sum_{k=1}^{\infty} (s_k^T g_k)^2/\|s_k\|_2^2 \leqslant \sum_{k=1}^{\infty} M(s_k^T g_k)^2/s_k^T y_k$$

$$\leqslant \frac{M}{1-c_2} \sum_{k=1}^{\infty} -s_k^T g_k < \infty. \tag{1.16}$$

Hence the lemma is true.

Line search conditions (1.6) and (1.7) are called weak Wolfe line search conditions. Condition (1.6) and the following inequality

$$|d_k^T \nabla f(x_k + \alpha_k d_k)| \leqslant -c_2 d_k^T g_k \tag{1.17}$$

are called strong Wolfe line search conditions. It is easy to see that (1.17) is stronger than (1.7). A direct consequence of (1.17) is that

$$(1-c_2) \leqslant \frac{s_k^T y_k}{-s_k^T g_k} \leqslant (1+c_2). \tag{1.18}$$

It can be easily shown that

$$H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{s_k s_k^T}{s_k^T y_k}. \tag{1.19}$$

For example see ref. [3]. From (1.2) and (1.19), we have

$$\operatorname{tr}(B_{k+1}) = \operatorname{tr}(B_k) - 2\frac{s_k^T B_k y_k}{s_k^T y_k} + \left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right)\frac{\|y_k\|_2^2}{s_k^T y_k}, \tag{1.20}$$

$$\operatorname{tr}(H_{k+1}) = \operatorname{tr}(H_k) - \frac{\|H_k y_k\|_2^2}{y_k^T H_k y_k} + \frac{\|s_k\|_2^2}{s_k^T y_k}. \tag{1.21}$$

It is well known that[4]

$$\operatorname{Det}(H_{k+1}) = \operatorname{Det}(H_k)\frac{s_k^T y_k}{y_k^T H_k y_k}. \tag{1.22}$$

Throughout this paper we make the following assumptions.

Assumption 1.1.

(i) *$f(x)$ is twice continuously differentiable and uniformly convex. Namely, there exists $c_3>0$ such that*

$$d^T\nabla^2 f(x)d \geqslant c_3\|d\|_2^2, \quad \forall d, \ x\in \mathscr{R}^n. \tag{1.23}$$

(ii) *The line search conditions* (1.6) and (1.7) *are satisfied in all iterations.*

Under the condition of Assumption 1.3, it can be shown that there exists a constant c_4 such that

$$c_3 \leqslant \frac{s_k^T y_k}{\|s_k\|_2^2} \leqslant \frac{\|y_k\|_2^2}{s_k^T y_k} \leqslant c_4. \tag{1.24}$$

From Assumption 1.1 and Lemma 1.1 we have the following fundamental results.

Lemma 1.2. *Let x_k $\{k=1, 2,\cdots\}$ be generated by the DFP method. If the conditions in Assumption 1.1 are satisfied, then*

$$\sum_{k=1}^{\infty}\alpha_k g_k^T H_k g_k < \infty, \tag{1.25}$$

and

$$\sum_{k=1}^{\infty}\frac{(g_k^T H_k g_k)^2}{\|d_k\|_2^2} < \infty. \tag{1.26}$$

Proof. Assumption (1.23) implies that $f(x)$ is bounded below. Therefore, (1.25) follows from (1.11) and

$$d_k = -H_k g_k. \tag{1.27}$$

It follows from inequality (1.23) that the set

$$\Omega = \{x \mid f(x) \leqslant f(x_1), \ x\in \mathscr{R}^n\} \tag{1.28}$$

$$s_k^T y_k \leqslant -\frac{2(1-c_1)c_4}{c_3} s_k^T g_k. \tag{1.29}$$

Proof. It follows from (1.23) and (1.24) that

$$\begin{aligned} f(x_{k+1}) &\geqslant f(x_k) + s_k^T g_k + \frac{1}{2} c_3\|s_k\|_2^2 \\ &\geqslant f(x_k) + s_k^T g_k + \frac{1}{2}\frac{c_3}{c_4} s_k^T y_k. \end{aligned} \tag{1.30}$$

Now (1.29) follows from the above inequality and (1.6).

2 Convergence analyses

Under extra conditions, we can use the trace relation (1.20) to prove the global convergence of the DFP method.

$$\sum_{k=1} g_k^T H_k g_k < \infty. \tag{2.1}$$

If the theorem is not true, there exists a positive constant δ such that $\|g_k\| > \delta$ for all k. From (1.20) we have

$$\frac{s_k^T y_k}{\alpha_k}(\mathrm{Tr}(B_{k+1}) - \mathrm{Tr}(B_k)) = g_k^T(g_{k+1} - g_k) + \left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right)\frac{\|y_k\|_2^2}{\alpha_k}. \tag{2.2}$$

From the Cauchy-Schwarz inequality, the sum

$$\sum_{k=1}^{N} g_k^T(g_{k+1} - g_k) \leqslant \frac{1}{2}(\|g_{N+1}\|_2^2 - \|g_1\|_2^2) \tag{2.3}$$

is uniformly bounded above for all $N \geqslant 1$. The assumption that α_k are bounded away from zero and inequalities (1.18) and (1.24) implies that

$$\begin{aligned}\sum_{k=1}^{\infty}\left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right)\frac{\|y_k\|_2^2}{\alpha_k} &\leqslant \sum_{k=1}^{\infty}\left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right)\frac{c_4 s_k^T y_k}{\alpha_k} \\ &= \sum_{k=1}^{\infty} c_4\left(\frac{s_k^T y_k}{\alpha_k} - s_k^T g_k\right) \\ &\leqslant c_4(2 + c_2)\sum_{k=1}^{\infty} - s_k^T g_k < \infty.\end{aligned} \tag{2.4}$$

Therefore, from (2.2)—(2.4) and $\mathrm{Tr}(B_{k+1}) - \mathrm{Tr}(B_k) \geqslant 0$ we have

$$\sum_{k=1}^{\infty} \frac{s_k^T y_k}{\alpha_k}(\mathrm{Tr}(B_{k+1}) - \mathrm{Tr}(B_k)) < \infty. \tag{2.5}$$

Because

$$\begin{aligned}\frac{s_k^T y_k}{\alpha_k}\mathrm{Tr}(B_k) &\geqslant \frac{-(1 - c_2)s_k^T g_k}{\alpha_k}\mathrm{Tr}(B_k) \\ &= (1 - c_2) g_k^T H_k g_k \mathrm{Tr}(B_k) \geqslant (1 - c_2)\|g_k\|_2^2 \\ &\geqslant (1 - c_2)\delta > 0,\end{aligned} \tag{2.6}$$

it follows from (2.5) that

$$\sum_{k=1}^{\infty}\left(\frac{\operatorname{Tr}(B_{k+1})}{\operatorname{Tr}(B_k)}-1\right)<\infty. \tag{2.7}$$

Therefore

$$\prod_{k=1}^{\infty}\frac{\operatorname{Tr}(B_{k+1})}{\operatorname{Tr}(B_k)}<\infty, \tag{2.8}$$

and consequently $\operatorname{tr}(B_k)$ are uniformly bounded. This contradicts (2.1) because $\|g_k\|_2$ are bounded away from zero.

In order to continue our convergence analyses, we need to establish some lemmas.

Lemma 2.1. *There exists a positive constant c_5 such that*

$$\operatorname{tr}(H_{k+1})\leqslant c_5 k, \tag{2.9}$$

$$\sum_{i=1}^{k}\frac{\|H_i y_i\|_2^2}{y_i^T H_i y_i}\leqslant c_5 k. \tag{2.10}$$

Proof. It follows from (1.21) and (1.24) that

$$\operatorname{tr}(H_{k+1})=\operatorname{tr}(H_1)-\sum_{i=1}^{k}\frac{\|H_i y_i\|_2^2}{y_i^T H_i y_i}+\sum_{i=1}^{k}\frac{\|s_i\|_2^2}{s_i^T y_i}$$

$$\leqslant\operatorname{tr}(H_1)-\sum_{i=1}^{k}\frac{\|H_i y_i\|_2^2}{y_i^T H_i y_i}+\frac{k}{c_3}. \tag{2.11}$$

The above inequality shows that (2.9) is true if $c_5=\operatorname{tr}(H_1)+1/c_3$. From (2.11) it follows that

$$\sum_{i=1}^{k}\|H_i y_i\|_2^2/y_i^T H_i y_i\leqslant c_5 k. \tag{2.12}$$

This completes our proof.

Lemma 2.2.

$$\sum_{k=1}^{\infty}\frac{(g_k^T H_k y_k)^2}{y_k^T H_k y_k}<\infty. \tag{2.13}$$

Proof. From (1.19), we have

$$g_{k+1}^T H_{k+1} g_{k+1}=g_k^T H_k g_k-\frac{(g_k^T H_k y_k)^2}{y_k^T H_k y_k}+\frac{(g_{k+1}^T s_k)^2}{s_k^T y_k}. \tag{2.14}$$

It follows from (1.8), (1.29) and (1.25) that

$$\sum_{k=1}^{\infty}\frac{(g_{k+1}^T s_k)^2}{s_k^T y_k}=\sum_{k=1}^{\infty}\left[\frac{(s_k^T g_k)^2}{s_k^T y_k}+2s_k^T g_k+s_k^T y_k\right]$$

$$\leqslant\sum_{k=1}^{\infty}\left[-\frac{1}{(1-c_2)}+2-\frac{2(1-c_1)c_4}{c_3}\right]s_k^Tg_k<\infty. \tag{2.15}$$

Therefore, we have

$$\sum_{k=1}^{\infty}\frac{(g_k^TH_ky_k)^2}{y_k^TH_ky_k}\leqslant g_1^TH_1g_1+\sum_{k=1}^{\infty}\frac{(g_{k+1}^Ts_k)^2}{s_k^Ty_k}<\infty. \tag{2.16}$$

Thus (2.13) is true.

Lemma 2.3. *There exists a positive constant c_6 such that*

$$\sum_{i=1}^{k}\frac{1}{\alpha_i}\leqslant c_6k, \tag{2.17}$$

$$\sum_{i=1}^{k}\alpha_i\geqslant k/c_6. \tag{2.18}$$

Proof. It follows from (1.8), (1.24) and (2.10) that

$$\sum_{i=1}^{k}\frac{1}{\alpha_i}\leqslant\sum_{i=1}^{k}\frac{s_i^Ty_i}{-(1-c_2)\alpha_ks_i^Tg_i}=\frac{1}{1-c_2}\sum_{i=1}^{k}\frac{s_i^Ty_i}{s_i^TB_is_i}$$

$$\leqslant\frac{1}{1-c_2}\sum_{i=1}^{k}\frac{y_i^TH_iy_i}{s_i^Ty_i}\leqslant\frac{c_4}{(1-c_2)}\sum_{i=1}^{k}\frac{y_i^TH_iy_i}{\|y_i\|_2^2}$$

$$\leqslant\frac{c_4}{(1-c_2)}\sum_{i=1}^{k}\frac{\|H_iy_i\|_2^2}{y_i^TH_iy_i}\leqslant\frac{c_4c_5}{(1-c_2)}k. \tag{2.19}$$

Hence (2.17) is true if $c_6=c_4c_5/(1-c_2)$. (2.18) follows from (2.17) and the following inequality

$$\sum_{i=1}^{k}\alpha_i\sum_{i=1}^{k}\frac{1}{\alpha_i}\geqslant k^2. \tag{2.20}$$

This completes our proof.

Theorem 2.2. *Assume that the conditions in Assumption 1.3 hold. If the sequence x_k generated by the DFP algorithm satisfies*

$$\|g_{k+1}\|_2\leqslant\|g_k\|_2 \tag{2.21}$$

for all large k, then x_k converges to the unique minimum of $f(x)$.

Proof. If the theorem is not true, x_k $(k=1, 2, \cdots)$ generated by the DFP algorithm satisfies (2.21) for all large k and there exists a positive constant δ such that

$$\|g_k\|_2\geqslant\delta \tag{2.22}$$

for all k. Without loss of generality, we assume that (2.21) holds for all k. Thus

$$s_k^TB_ky_k=\alpha_k[\|g_k\|_2^2-g_k^Tg_{k+1}]\geqslant\alpha_k\|g_k\|_2[\|g_k\|_2-\|g_{k+1}\|_2]\geqslant 0. \tag{2.23}$$

The above inequality, (1.20), (1.24) and (2.18) imply that

$$\operatorname{tr}(B_{k+1}) \leqslant \operatorname{tr}(B_1) + c_4\left[k + \frac{1}{(1-c_2)}\sum_{i=1}^{k}\alpha_i\right]$$

$$\leqslant \operatorname{tr}(B_1) + c_4\left[c_6 + \frac{1}{(1-c_2)}\right]\sum_{i=1}^{k}\alpha_i. \tag{2.24}$$

It follows from (2.22) and (1.25) that

$$\sum_{k=1}^{\infty}\frac{\alpha_k}{\operatorname{tr}(B_k)} < \infty. \tag{2.25}$$

The above relation and inequality (2.24) indicate

$$\sum_{k=1}^{\infty}\alpha_k < \infty, \tag{2.26}$$

which contradicts (2.18). This shows that the theorem is true.

Lemma 2.4. *There exists a positive constant* c_7 *such that*

$$\sum_{i=1}^{k}\|d_i\|_2 \leqslant c_7\sqrt{k}. \tag{2.27}$$

Proof. The lemma follows from the previous lemma and the inequality

$$\sum_{i=1}^{k}\|d_i\|_2 \leqslant \sqrt{\sum_{i=1}^{k}\frac{\|d_i\|_2^2\|y_i\|_2^2}{y_i^T H_i y_i}\sum_{i=1}^{k}\frac{y_i^T H_i y_i}{\|y_i\|_2^2}}$$

$$\leqslant \sqrt{\frac{c_4}{c_3}}\sqrt{\sum_{i=1}^{k}\frac{(d_i^T y_i)^2}{y_i^T H_i y_i}\sum_{i=1}^{k}\frac{\|H_i y_i\|_2^2}{y_i^T H_i y_i}}$$

$$\leqslant \sqrt{\frac{c_4}{c_3}}\sqrt{c_5\sum_{i=1}^{\infty}\frac{(g_i^T H_i y_i)^2}{y_i^T H_i y_i}}\sqrt{k}. \tag{2.28}$$

It follows from (2.13) that

$$c_7 = \sqrt{\frac{c_4}{c_3}}\sqrt{c_5\sum_{i=1}^{\infty}\frac{(g_i^T H_i y_i)^2}{y_i^T H_i y_i}} \tag{2.29}$$

is a finite number. Therefore, the lemma follows from (2.28) and (2.29).

Corollary 2.1. *There exists a positive constant* c_7 *such that*

$$\sum_{i=1}^{k}\frac{1}{\|d_i\|_2} \geqslant \frac{k^{1.5}}{c_7}, \tag{2.30}$$

$$\sum_{i=1}^{k}\frac{1}{\|d_i\|_2^2} \geqslant \frac{k^2}{c_7^2}. \tag{2.31}$$

Proof. Inequality (2.30) follows directly from the previous lemma and the inequality

$$\sum_{i=1}^{k}\frac{1}{\|d_i\|}\sum_{i=1}^{k}\|d_i\|\geqslant k^2. \tag{2.32}$$

Using (2.30) and the inequality

$$\left(\sum_{i=1}^{k}\frac{1}{\|d_i\|}\right)^2\leqslant k\sum_{i=1}^{k}\frac{1}{\|d_i\|_2^2}, \tag{2.33}$$

we can obtain inequality (2.31).

Lemma 2.5. *There exists a positive constant c_8 such that*

$$\mathrm{tr}(B_{k+1})\leqslant c_8\sum_{i=1}^{k}\frac{1}{\|d_i\|_2^2}. \tag{2.34}$$

Proof. From (1.20), (1.24) and (2.18) we have

$$\begin{aligned}\mathrm{tr}(B_{k+1})&\leqslant \mathrm{tr}(B_1)+\sum_{i=1}^{k}\frac{2\alpha_i\|g_i\|_2\|y_i\|_2}{s_i^Ty_i}+\sum_{i=1}^{k}\left(1+\frac{\alpha_i}{1-c_2}\right)c_4\\ &\leqslant \mathrm{tr}(B_1)+2\eta\sqrt{c_4/c_3}\sum_{i=1}^{k}\frac{1}{\|d_i\|_2}+c_4k+\frac{c_4}{(1-c_2)}\sum_{i=1}^{k}\alpha_i\\ &\leqslant \mathrm{tr}(B_1)+2\eta\sqrt{\frac{c_4}{c_3}}\sum_{i=1}^{k}\frac{1}{\|d_i\|_2}+\frac{c_4[1+c_6(1-c_2)]}{(1-c_2)}\sum_{i=1}^{k}\alpha_i,\end{aligned} \tag{2.35}$$

where η is the upper bound of $\{\|\nabla f(x)\|_2, |x\in\Omega\}$, and Ω is defined by (1.28). It follows from (1.25), (1.29) and (1.24) that

$$\lim_{k\to\infty}\|s_k\|_2=0, \tag{2.36}$$

which implies that

$$\sum_{i=1}^{k}\alpha_i=\sum_{i=1}^{k}\|s_i\|_2\frac{1}{\|d_i\|_2}\leqslant\sum_{i=1}^{k}\frac{1}{\|d_i\|_2}, \tag{2.37}$$

for all large k. The above inequality and relation (2.35) show that there exists a positive constant c_8 such that (2.34) holds for all k.

Lemma 2.6. *If $\|g_k\|_2\geqslant\delta>0$ for all k, there exists a positive constant c_9 such that*

$$\sum_{i=1}^{k}\frac{1}{\|d_i\|_2}\leqslant c_9\left(\sum_{i=1}^{k}\sqrt{\alpha_i}\right)^2. \tag{2.38}$$

Proof. Define

$$D_k=\sum_{i=1}^{k-1}\frac{1}{\|d_i\|_2}. \tag{2.39}$$

It follows from (1.24) that

$$\frac{1}{\|d_k\|_2}=\frac{\alpha_k}{\|s_k\|_2}\leqslant\alpha_k\sqrt{\frac{c_4}{s_k^Ty_k}}$$

$$\leqslant \alpha_k \sqrt{\frac{c_4}{-(1-c_2)s_k^T g_k}} = \sqrt{\frac{c_4}{1-c_2}} \sqrt{\frac{\alpha_k}{g_k^T H_k g_k}}$$

$$\leqslant \sqrt{\frac{c_4}{(1-c_2)\delta^2}} \sqrt{\alpha_k \mathrm{tr}(B_k)} \leqslant \sqrt{\frac{c_4 c_8}{(1-c_2)\delta^2}} \sqrt{\alpha_k D_k}. \tag{2.40}$$

Therefore, by setting $c_9 = \sqrt{\frac{c_4 c_8}{(1-c_2)\delta^2}}$, we have

$$D_{k+1} = D_k + \frac{1}{\|d_k\|} \leqslant D_k + c_9 \sqrt{\alpha_k D_k}, \tag{2.41}$$

which implies that

$$\sqrt{D_{k+1}} \leqslant \sqrt{D_k} + c_9 \sqrt{\alpha_k}. \tag{2.42}$$

Therefore the lemma is true.

Theorem 2.3. *If*

$$\sum_{k=1}^{\infty} \|s_k\| < \infty, \tag{2.43}$$

then the DFP algorithm converges to the solution.

Proof. If the theorem is not true, there exists a positive constant δ such that $\|g_k\|_2 \geqslant \delta$ for all k. From (2.43) and (2.40) we have

$$\sum_{k=1}^{\infty} \frac{\sqrt{\alpha_k}}{\sqrt{\mathrm{Tr}(B_k)}} < \infty. \tag{2.44}$$

It follows from the above relation, (2.34) and (2.38) that

$$\sum_{k=1}^{\infty} \frac{\sqrt{\alpha_k}}{\sum_{i=1}^{k} \sqrt{\alpha_i}} < \infty. \tag{2.45}$$

Therefore, we can show that

$$\sum_{k=1}^{\infty} \sqrt{\alpha_k} < \infty, \tag{2.46}$$

which contradicts (2.18).

Finally we show that the DFP method is also globally convergent if inequality (2.47) holds for all k.

Theorem 2.4. *Assume that the conditions in Assumption 1.3 hold. Let the sequence $\{x_k, k=1, 2, \cdots\}$ be generated by the DFP method. If*

$$y_k^T B_k g_k \leqslant 0 \tag{2.47}$$

for all k, then $\{x_k\}$ either terminates at the unique minimum x^* of $f(x)$ or converges to x^*.

Proof. We prove the theorem by contradiction. If the theorem is not true, there exists a positive number δ such that (2.22) holds for all k.

We consider the updating formula for B_{k+1}^2. From (1.2), we can easily obtain

$$B_{k+1}^2 = B_k^2 - \frac{B_k^2 s_k y_k^T + B_k y_k s_k^T B_k}{s_k^T y_k} + \left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right) \frac{B_k y_k y_k^T}{s_k^T y_k}$$

$$- \frac{B_k s_k y_k^T B_k + y_k s_k^T B_k^2}{s_k^T y_k} + \frac{(B_k s_k y_k^T + y_k s_k^T B_k)(B_k s_k y_k^T + y_k s_k^T B_k)}{(s_k^T y_k)^2}$$

$$- \left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right) \frac{B_k s_k y_k^T \|y_k\|_2^2 + y_k s_k^T B_k y_k y_k^T}{(s_k^T y_k)^2}$$

$$+ \left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right) \frac{y_k y_k^T B_k}{s_k^T y_k} - \left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right) \frac{y_k y_k^T B_k s_k y_k^T + \|y_k\|_2^2 y_k s_k^T B_k}{(s_k^T y_k)^2}$$

$$+ \left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right)^2 \frac{\|y_k\|_2^2 y_k y_k^T}{(s_k^T y_k)^2}. \tag{2.48}$$

This implies the following relation

$$\operatorname{tr}(B_{k+1}^2) = \operatorname{tr}(B_k^2) - 4 \frac{y_k^T B_k^2 s_k}{s_k^T y_k} + 2\left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right) \frac{y_k^T B_k y_k}{s_k^T y_k}$$

$$+ 2 \frac{\|B_k s_k\|_2^2 \|y_k\|_2^2 + (s_k^T B_k y_k)^2}{(s_k^T y_k)^2} - 4\left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right) \frac{y_k^T B_k s_k \|y_k\|_2^2}{(s_k^T y_k)^2}$$

$$+ \left(1 + \frac{s_k^T B_k s_k}{s_k^T y_k}\right)^2 \frac{\|y_k\|_2^4}{(s_k^T y_k)^2}. \tag{2.49}$$

Therefore, using the above relation and inequalities (1.24) we have

$$\operatorname{tr}(B_{k+1}^2) \leqslant \operatorname{tr}(B_k^2) - 4 \frac{y_k^T B_k^2 s_k}{s_k^T y_k} + 2c_4\left(1 + \frac{\alpha_k}{1 - c_2}\right) \operatorname{tr}(B_k)$$

$$+ 4\eta^2 \frac{c_4}{c_3} \frac{1}{\|d_k\|_2^2} + 4\left(1 + \frac{\alpha_k}{1 - c_2}\right) c_4 \eta \sqrt{\frac{c_4}{c_3}} \frac{1}{\|d_k\|_2}$$

$$+ \left(1 + \frac{\alpha_k}{1 - c_2}\right)^2 c_4^2$$

$$\leqslant \left(1 + \frac{2c_4 \alpha_k \operatorname{tr}(B_k)}{(1 - c_2)\operatorname{tr}(B_k^2)}\right) \operatorname{tr}(B_k^2)$$

$$-4\frac{y_k^T B_k^2 s_k}{s_k^T y_k}+4\eta\sqrt{\frac{c_4}{c_3}}\left(c_4+\frac{c_4\alpha_k}{1-c_2}\right)\frac{1}{\|d_k\|_2}$$

$$+2c_4\mathrm{tr}(B_k)+4\eta^2\frac{c_4}{c_3}\frac{1}{\|d_k\|_2^2}+\left(1+\frac{\alpha_k}{1-c_2}\right)^2 c_4^2. \tag{2.50}$$

It follows from (2.25) that

$$\sum_{k=1}^{\infty}\frac{\alpha_k\mathrm{tr}(B_k)}{\mathrm{tr}(B_k^2)}<\infty. \tag{2.51}$$

This inequality, (2.47) and (2.50) imply that there exists a positive constant c_{10} such that

$$\mathrm{tr}(B_{k+1}^2)\leqslant\mathrm{tr}(B_1^2)+c_{10}\sum_{i=1}^{k}\left[\mathrm{tr}(B_i)+\alpha_i^2+(1+\alpha_i)\left(1+\frac{1}{\|d_i\|_2}\right)+\frac{1}{\|d_i\|)_2^2}\right]. \tag{2.52}$$

Due to $\|s_k\|\to 0$, it is easy to see that

$$\sum_{i=1}^{k}\alpha_i^2=\sum_{i=1}^{k}\frac{\|s_i\|_2^2}{\|d_i\|_2^2}\leqslant\sum_{i=1}^{k}\frac{1}{\|d_i\|_2^2} \tag{2.53}$$

holds for all large k. Similarly

$$\sum_{i=1}^{k}\frac{\alpha_i}{\|d_i\|_2}\leqslant\sum_{i=1}^{k}\frac{\|s_i\|}{\|d_i\|_2^2}\leqslant\sum_{i=1}^{k}\frac{1}{\|d_i\|_2^2} \tag{2.54}$$

holds for all large k. Therefore, there exists a positive constant c_{11} such that

$$\mathrm{tr}(B_{k+1}^2)\leqslant c_{11}\left(\sum_{i=1}^{k}\mathrm{tr}(B_i)+\sum_{i=1}^{k}\frac{1}{\|d_i\|_2^2}\right) \tag{2.55}$$

holds for all k. Thus, it follows that

$$(\mathrm{tr}(B_{k+1}))^2\leqslant c_{11}n^2\left(\sum_{i=1}^{k}\mathrm{tr}(B_i)+\sum_{i=1}^{k}\frac{1}{\|d_i\|_2^2}\right). \tag{2.56}$$

Define a positive sequence $\{\overline{D}_k,\ k=1,\ 2,\cdots\}$ by

$$\overline{D}_1=\mathrm{tr}(B_1), \tag{2.57}$$

$$\overline{D}_{k+1}=\sqrt{c_{11}n^2\sum_{i=1}^{k}\left(\overline{D}_i+\frac{1}{\|d_i\|_2^2}\right)}. \tag{2.58}$$

Then it follows from (2.56) that

$$\mathrm{tr}(B_k)\leqslant\overline{D}_k \tag{2.59}$$

for all k. From (2.58) we see that $\overline{D}_k$ is monotonically increasing. Thus, it follows that

$$\overline{D}_{k+1}^2\leqslant c_{11}n^2k\overline{D}_{k+1}+c_{11}n^2\sum_{i=1}^{k}\frac{1}{\|d_i\|_2^2}. \tag{2.60}$$

Therefore, one of the following two inequalities

$$\overline{D}_{k+1}^2 \leqslant 2c_{11}n^2k\overline{D}_{k+1} \tag{2.61}$$

and

$$\overline{D}_{k+1}^2 \leqslant 2c_{11}n^2\sum_{i=1}^{k}\frac{1}{\|d_i\|_2^2} \tag{2.62}$$

must hold. If (2.62) does not hold, from (2.61) and (2.31) we have

$$\overline{D}_{k+1}^2 \leqslant 4c_{11}^2n^4k^2 \leqslant 4c_7^2c_{11}^2n^4\sum_{i=1}^{k}\frac{1}{\|d_i\|_2^2}. \tag{2.63}$$

It follows from (2.62) or (2.63) that there exists a constant c_{12} such that

$$\overline{D}_k^2 \leqslant c_{12}\sum_{i=1}^{k}\frac{1}{\|d_i\|_2^2}. \tag{2.64}$$

From the above inequality, (2.59), (1.26) and (2.22), we can show that

$$\sum_{k=1}^{\infty}\frac{1/\|d_k\|_2^2}{\overline{D}_k^2} < \infty. \tag{2.65}$$

Now the above relation and inequality (2.64) imply that

$$\sum_{k=1}^{\infty}\frac{1}{\|d_k\|_2^2} < \infty, \tag{2.66}$$

which contradicts (2.31). This shows that the theorem is true.

3 Discussion

We have shown that the DFP method is globally convergent if the object function is uniformly convex, if line search conditions are weak Wolfe inexact line search conditions, and if some extra conditions are satisfied. From our analyses, we can see that, if either

$$\sum_{i=1}^{k} - \frac{y_i^TB_is_i}{s_i^Ty_i} \leqslant c_{12}\sum_{i=1}^{k}\alpha_i \quad . \tag{3.1}$$

or

$$\sum_{i=1}^{k}\left|\frac{y_i^TB_i^2s_i}{s_i^Ty_i}\right| \leqslant c_{13}\sum_{i=1}^{k}\frac{1}{\|d_i\|_2^2} \tag{3.2}$$

holds, then the global convergence of the DFP method can be established. But we have not yet been able to prove any of the above two inequalities.

References

1 Byrd, R., Nocedal, J., Yuan, Y., Global convergence of a class of variable metric algorithms, *SIAM J. Numer. Anal.*, 1987, 24:1171.

2 Davidon, W. C., Variable metric methods for minimization, *Argonne National Labs Report ANL-5990*, 1959.

3 Dennis, J. E., Jr., Moŕe, J. J., Quasi-Newton method, motivation and theory, *SIAM Review*, 1977, 19:46.

4 Dennis, J. E., Jr., Schnabel, R. B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations,*

Englewood Cliffs: Prentice-Hall, 1983.

5 Fletcher, R., Powell, M. J. D., A rapid convergent descent method for minimization, *The Computer Journal*, 1963, 6: 163.

6 Powell, M. J. D., On the convergence of the variable metric algorithm, *J. Inst. Maths. Appl.*, 1971, 7: 21.

7 Powell, M. J. D., Some global convergence properties of a variable metric algorithm for minimization without exact line searches, in *Nonlinear Programming, SIAM-AMS Proceedings vol. IX* (eds. Cottle, R. W., Lemke, C. E.), Philadelphia: SIAM Publications, 1976, 53—72.