American Journal of Applied Mathematics
2017; 5(1): 19-30
http://www.sciencepublishinggroup.com/j/ajam
doi: 10.11648/j.ajam.20170501.13
ISSN: 2330-0043 (Print); ISSN: 2330-006X (Online)

On Modified DFP Update for Unconstrained Optimization

Saad Shakir Mahmood, Samira Hassan Shnywer

Department of Mathematics, College of Education, Almustansiryah University, Baghdad, Iraq

Email address:
saadshakirmahmood@yahoo.com (S. S. Mahmood), samira.hassan75@yahoo.com (S. H. Shnywer)

To cite this article:
Saad Shakir Mahmood, Samira Hassan Shnywer. On Modified DFP Update for Unconstrained Optimization. *American Journal of Applied Mathematics.* Vol. 5, No. 1, 2017, pp. 19-30. doi: 10.11648/j.ajam.20170501.13

Received: December 25, 2016; **Accepted**: January 9, 2017; **Published**: February 6, 2017

Abstract: In this paper, we propose a new modify of *DFP* update with a new extended quasi-Newton condition for unconstrained optimization problem so called $(\alpha - DFP)$ update. This update is based on a new Zhang Xu condition we show that $(\alpha - DFP)$ update preserves the value of determinant of the next Hessian matrix equal to the value of determinant of current Hessian matrix theoretically and practically. Global convergence of the modify is established. Local and super linearly convergence are obtained for the proposed method. Numerical results are given to compare a performance of the modify $(\alpha - DFP)$ method with the standard *DFP* method on same function is selected.

Keywords: Quasi-Newton Equation, the *DFP* Updating Formula, Global Convergence and Super Linearly Convergence

1. Introduction

The quasi-Newton methods are very useful and efficient methods for solving the unconstrained minimization problem

$$\min_{x \in R^n} f(x) \tag{1}$$

Where $f: R^n \to R$ is twice continuously differentiable. Starting from point x_0 and a symmetric positive definite matrix B_0, quasi-Newton method generates sequence $\{x_k\}$ and $\{B_k\}$ by the iteration of the form

$$x_{x+1} = x_k - B_k^{-1} \nabla f(x_k), k = 0, 1, \cdots \tag{2}$$

where the update $B_k \in R^{n \times n}$ satisfies the following famous quasi-Newton equation or (secant equation):

$$B_{k+1} s_k = y_k \tag{3}$$

with

$$s_k = \lambda_k d_k, y_k = g_{k+1} - g_k \tag{4}$$

where λ_k is the step length and d_k is the search direction that is obtained by solving the equation:

$$d_k = -B_k^{-1} \nabla f(x_k) \tag{5}$$

in which $g_k = \nabla f(x_k)$ is the gradient of $f(x)$ at x_k and B_k is an approximation to the Hessian matrix $G_k = \nabla^2 f(x_k)$. The updating matrix B_k is required to satisfy the usual quasi-Newton equation (3) with equation (4). So B_{k+1} is reasonable approximation to G_{k+1}.

The $(\alpha - DFP)$ update consist of iteration of the form (2) where d_k is the search direction which of the form (5) and the Hessian approximation B is update by the $(-DFP)$ formula with quasi-Newton equation (3).

$$B_{k+1} = B - \frac{ys^TB + Bsy^T}{y^Ts} + \frac{yy^T}{y^Ts}\left(\frac{1}{\alpha} + \frac{s^TBs}{y^Ts}\right) \tag{6}$$

$$\text{where } \alpha = \frac{y^Ts\,[Det(B)]^2}{s^TBs} \tag{7}$$

The formula is the modifying of *DFP* update which is satisfy equation (3) and in the next section.

In the following discussion, we shall use $\|\cdot\|$ and $\|\cdot\|_F$ to denote the ι_2-norm and the frobenius norm, respectively. For a symmetric positive definite matrix $M \in R^{n \times n}$, we show also use the following weighted norm

$$\|X\|_M = \|MXM\|_F, \forall X \in R^{n \times n} \tag{8}$$

2. $\alpha - DFP$ Update

Let $\delta = \alpha - 1$ and from Zhang Xu condition, we have:

$$s^* = s + (\alpha - 1)s,$$

$$\text{which gives } s^* = \alpha s \tag{9}$$

From quasi-Newton equation (3), we get:

$B_{k+1}s_k^* = y_k$ which is extended of quasi-Newton equation (10)

Now, substitution (9) in (10), we obtain:

$$B_{k+1}\alpha s_k = y_k \tag{11}$$

Let we consider the determinant of Hessian matrix for the *DFP* update [7] with replace each s_k by s_k^*, we get:

$$\text{Det}\,(B_{k+1}) = \frac{y_k^T s_k^* [Det(B_k)]}{s_k^{*T} B_k s_k^*} \tag{12}$$

And from equation (9), we obtain:

$$\text{Det}\,(B_{k+1}) = \frac{y_k^T \alpha s_k [Det(B_k)]}{\alpha s_k^T B_k \alpha s_k} \tag{13}$$

Which gives:

$$\text{Det}\,(B_{k+1}) = \frac{y_k^T s_k [Det(B_k)]}{\alpha s_k^T B_k s_k} \tag{14}$$

we suppose that Det $(B_{k+1}) = \frac{1}{Det(B_k)}$.

Then the equation (14) becomes:

$$\frac{1}{Det(B_k)} = \frac{y_k^T s_k [Det(B_k)]}{\alpha s_k^T B_k s_k}$$

Now, by multiplying both sides with $\frac{1}{Det(B_k)}$, we get:

$$\frac{1}{[Det(B_k)]^2} = \frac{y_k^T s_k}{\alpha s_k^T B_k s_k} \tag{15}$$

Hence

$$\alpha(s_k^T B_k s_k) = y_k^T s_k\ [Det(B_k)]^2.$$

Finally:

$$\alpha = \frac{y_k^T s_k [Det(B_k)]^2}{s_k^T B_k s_k} \tag{16}$$

In addition, by more simplifying we can write $B_{k+1(\alpha-DFP)}$ as follows:

$$B_{k+1(\alpha-DFP)} = B_k - \frac{y_k s_k^{*T} B_k + B_k s_k^* y_k^T}{y_k^T s_k^*} + \frac{y_k y_k^T}{y_k^T s_k^*}\left(1 + \frac{s_k^{*T} B_k s_k^*}{y_k^T s_k^*}\right) \tag{17}$$

which is equivalent the following formula:

$$B_{k+1(\alpha-DFP)} = B_k + \frac{(y_k - B_k s_k^*) y_k^T + y_k (y_k - B_k s_k^*)^T}{y_k^T s_k^*} - \frac{(y_k - B_k s_k^*)^T s_k^* y_k y_k^T}{(y_k^T s_k^*)^2} \tag{18}$$

and also the formula (17) is equivalent the following formula:

$$B_{k+1(\alpha-DFP)} = \left(I - \frac{y_k s_k^{*T}}{y_k^T s_k^*}\right) B_k \left(I - \frac{s_k^* y_k^T}{y_k^T s_k^*}\right) + \frac{y_k y_k^T}{y_k^T s_k^*} \tag{19}$$

where s_k^* is defined by (9). It is clear that any formula is symmetric and satisfies the quasi-Newton equation.

3. Convergence Analysis

Now, we study the global convergence of the $\alpha - DFP$ update:

At the first, we need the following assumptions:

Assumption (3.1)

(A) : $f: R^n \to R$ is twice continuously differentiable on convex set $D \subset R^n$

(B): $f\ (x)$ is uniformly convex, i.e., there exist positive constants n and N such that for all $x \in L(x) = \{x\ | f\ (x) \le f(x_0)\}$, which is convex, where x_0 is starting point, we have:

$$n\ \|v\|^2 \le v^T \nabla^2 f(x) v \le N \|v\|^2,\ \forall\ v \in R^n \tag{20}$$

The assumption (B) implies that $\nabla^2 f(x)$ is positive definite on (x),

and that f has a unique minimizer x^* in $L(x)$.

By definition of weighted norm (8) and equation (9), satisfy extended (Q-N) equation then $W y_k = s_k^*$ and $y_k = W^{-1} \alpha\ s_k$ that is then

$$y_k = \widetilde{G_k} \alpha s_k \tag{21}$$

where $\widetilde{G_k} = W^{-1}$.

Now by return to the property (20) and from definition of weighted norm, we get:

$$n \le \frac{y_k^T \alpha s_k}{\|\alpha s_k\|^2} = \frac{\alpha s_k^T \widetilde{G_k} \alpha s_k}{\|\alpha s_k\|^2} \le N \tag{22}$$

Where $\widetilde{G_k}$ is the average Hessian, which is defined as:

$$\widetilde{G_k} = [\ \int_0^1 \nabla^2 f\ (x_k + \vartheta s_k^*)\ d\vartheta \tag{23}$$

and

$$\frac{1}{N} \le \frac{\|\alpha s_k\|^2}{y_k^T \alpha s_k} \le \frac{1}{n} \tag{24}$$

Since also

$$\frac{\|y_k\|^2}{\alpha s_k^T y_k} = \frac{y_k^T y_k}{\alpha s_k^T y_k} = \frac{\alpha s_k^T \widetilde{G_k} \widetilde{G_k} \widetilde{\alpha s_k}}{\alpha s_k^T \alpha \widetilde{G_k} s_k} = \frac{\alpha^2 s_k^T \widetilde{G_k}^2 s_k}{\alpha^2 s_k^T \widetilde{G_k} s_k} \tag{25}$$

Assumption (B) of (3.1) means that $\widetilde{G_k}$ is positive definite proven, thus its square root is well defined. There is a symmetric square root $\widetilde{G_k}^{\frac{1}{2}}$ is satisfying

$$\widetilde{G_k} = \widetilde{G_k}^{\frac{1}{2}}.\ \widetilde{G_k}^{\frac{1}{2}}$$

If we let $w_k = \widetilde{G_k}^{\frac{1}{2}} \alpha s_k$, then (26)

$$\frac{\|y_k\|^2}{\alpha s_k^T y_k} = \frac{\alpha s_k^T \widetilde{G_k}^{\frac{1}{2}} \widetilde{G_k}^{\frac{1}{2}} \widetilde{G_k}^{\frac{1}{2}} \widetilde{G_k}^{\frac{1}{2}} \alpha s_k}{\alpha s_k^T \widetilde{G_k}^{\frac{1}{2}} \widetilde{G_k}^{\frac{1}{2}} \alpha s_k}$$

$$= \frac{\left(\widetilde{G_k}^{\frac{1}{2}} \alpha s_k\right)^T \widetilde{G_k} \left(\widetilde{G_k}^{\frac{1}{2}} \alpha s_k\right)}{\left(\alpha s_k \widetilde{G_k}^{\frac{1}{2}}\right)^T \left(\widetilde{G_k}^{\frac{1}{2}} \alpha s_k\right)} \tag{27}$$

Substitution equation (26) in equation (27), we get

$$\frac{\|y_k\|^2}{\alpha s_k^T y_k} = \frac{w_k^T \widetilde{G_k} w_k}{w_k^T w_k} \tag{28}$$

And from (20) $v^T \nabla^2 f\,(x^*) v \le N \|v\|^2$, we know $\|v\|^2 > 0$. Then we can divide both sides from this, we get:

$$\frac{v_k^T \widetilde{G_k} v_k}{\|v_k\|^2} \le N$$

That mean

$$\frac{w_k^T \widetilde{G_k} w_k}{w_k^T w_k} \le N \tag{29}$$

Then from equation (28), we get:

$$\frac{\|y_k\|^2}{\alpha s_k^T y_k} \le N \tag{30}$$

In addition, from equations (21) and (9), we get:

$$\|y_k\| = \left\|\widetilde{G_k} s_k^*\right\|$$

$$\|y_k\| \le \left\|\widetilde{G_k}\right\| \|s_k^*\|,\ \|s_k^*\| \le \left\|\widetilde{G_k}^{-1}\right\| \|y_k\| \tag{31}$$

Which gives:

$$\frac{\|y_k\|}{\|s_k^*\|} \le N \tag{32}$$

And

$$\frac{\|s_k^*\|}{\|y_k\|} \le \frac{1}{n} \tag{33}$$

Therefore, from the above discussion, we have:

Lemma (3.2)

Let *f*: $R^n \to R$ satisfy Assumption (3.1). Then

$$\frac{\|s_k\|}{\|y_k\|}, \frac{\|y_k\|}{\|s_k\|}, \frac{s_k y_k}{\|s_k\|^2}, \frac{s_k^T y_k}{\|y_k\|^2}, \frac{\|y_k\|^2}{s_k^T y_k}$$

Are bounded.

Hence, we have

$$\frac{\|s_k^*\|}{\|y_k\|}, \frac{\|y_k\|}{\|s_k^*\|}, \frac{s_k^* y_k}{\|s_k^*\|^2}, \frac{{s_k^*}^T y_k}{\|y_k\|^2}, \frac{\|y_k\|^2}{{s_k^*}^T y_k}$$

Are bounded.

Lemma (3.3)

Under exact line search, $\sum\|s_k\|^2$ and $\sum\|y_k\|^2$ are convergent.

Which gives:

$\frac{\|s_k\|^2}{\|y_k\|^2}$ is convergent and bounded.

From lemma (3.2), we have

$\frac{\|s_k^*\|^2}{\|y_k\|^2}$ is convergent and bounded.

Since $\sum\|y_k\|^2$ is convergent, we get:

$\|s_k^*\|^2$ is convergent, then

$$\sum\|s_k^*\|^2 \le \frac{1}{n}\sum\|y_k\|^2$$

Which implies $\sum\|s_k^*\|^2$ is convergent, where $f\,(x^*)$ is the minimum of $f\,(x)$.

Lemma (3.4)

For all vector *x*, the inequality

$$\|g(x)\|^2 \ge n\,[\,f(x) - f(x^*)] \tag{34}$$

holds, where $f(x^*)$ is the minimum of $f(x)$.

Proof:

Since the function

$$\phi(\tau) = f\big(x + \tau(x^* - x)\big), (0 \le \tau \le 1)$$

is a convex function, we have:

$$f\big(x + \tau(x^* - x)\big) \ge f(x) + \tau(x^* - x)^T g(x)$$

In particular, set $\tau = 1$, then we have

$$f\big(x + (x^* - x)\big) \ge f(x) + (x^* - x)^T g(x)$$

$$f(x^*) \ge f(x) + (x^* - x)^T g(x)$$

Which gives

$$f(x^*) - f(x) \ge (x^* - x)^T g(x)$$

By multiplying both sides with (-1) we get

$$f(x) - f(x^*) \le -(x^* - x)^T g(x)$$

By Cauchy-Schwarz inequality, we get

$$f(x) - f(x^*) \le |(x^* - x)^T g(x)| \le \|g(x)\|\|x^* - x\| \tag{35}$$

From (24) and (9), we have

$$|\alpha|\|s_k\|^2 \le \frac{y_k^T s_k}{n}$$

$$|\alpha|\|s_k\|^2 \le \frac{\big(g(x^*) - g(x)\big)^T s_k}{n} \tag{36}$$

$$\le \frac{\|g(x)\|\|x^* - x\|}{n}$$

Hence

$$\alpha\|x^* - x\|^2 \le \frac{\|g(x)\|\|x^* - x\|}{n}$$

$$\|x^* - x\|^2 \le \frac{\|g(x)\|\|x^* - x\|}{\alpha n}$$

Let $\alpha n = n$

$$\|x^* - x\|^2 \le \|g(x)\|\|x^* - x\|/n$$

$$\|x^* - x\| \le \|g(x)\|/n \tag{37}$$

Substituting (37) into (35) establishes (34).

Theorem (3.5)

Suppose that $f(x)$ satisfies Assumption (3.1). Then under exact line search the sequence $\{\,x_k\}$ generated by $\alpha - DFP$ method converges to the minimizer x^* of *f*.

Proof: Consider $\alpha - DFP$ formula of inverse Hessian approximation

$$H_{k+1} = H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} + \frac{\alpha s_k \alpha s_k^T}{\alpha s_k^T y_k} \quad (38)$$

and from $\alpha - DFP$ formula (19) of Hessian approximation

$$B_{k+1} = \left(I - \frac{y_k \alpha s_k^T}{\alpha s_k^T y_k}\right) B_k \left(I - \frac{\alpha s_k y_k^T}{\alpha s_k^T y_k}\right) + \frac{y_k y_k^T}{\alpha s_k^T y_k} \quad (39)$$

Obviously, $B_{k+1}H_{k+1} = I$. By computing the trace of (39), we have

$$Tr(B_{k+1}) = Tr(B_k) - 2\frac{\alpha s_k^T B_k y_k}{\alpha s_k^T y_k} + \frac{\alpha s_k^T B_k \alpha s_k (y_k^T y_k)}{(\alpha s_k^T y_k)^2} + \frac{y_k y^T}{\alpha s_k^T y_k} \quad (40)$$

The middle two terms can be written as:

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

From equation (4) and (5), we have

$$s_k^T y_k = s_k^T (g_{k+1} - g_k) = s_k^T g_{k+1} - s_k^T g_k$$

From the property of the *DFP* method [5] $g_{k+1}^T s_k = s_k^T g_{k+1} = 0$, we obtain:

$$= \frac{2 \lambda_k \alpha g_k^T y_k}{\alpha s_k^T y_k} + \frac{\lambda_k \alpha^2 s_k^T y_k y_k^T y_k}{(\alpha s_k^T y_k)^2}$$

$$= \lambda_k \left[\frac{2 g_k^T y_k + y_k^T y_k}{s_k^T y_k}\right]$$

From equation (4)and (5) again, we get

$$= \frac{2 g_k^T y_k + y_k^T y_k}{g_k^T H_k g_k}$$

$$= \frac{g_{k+1}^T g_{k+1} - g_k^T g_k}{g_k^T H_k g_k}$$

$$= \frac{\|g_{k+1}\|^2 - \|g_k\|^2}{g_k^T H_k g_k} \quad (41)$$

From the positive definiteness property of H_{k+1}, then (38) becomes:

$$g_{k+1}^T H_{k+1} g_{k+1} = g_{k+1}^T \left[H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}\right] g_{k+1}$$

$$= g_{k+1}^T \left[H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}\right] g_{k+1}, \left(g_{k+1}^T \frac{s_k^* s_k^{*T}}{s_k^{*T} y_k} g_{k+1}\right) = 0$$

$$= (y_k + g_k)^T \left[H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}\right] (y_k + g_k)$$

$$= \left[y_k^T H_k - \frac{y_k^T H_k y_k y_k^T H_k}{y_k^T H_k y_k} + g_k^T H_k - \frac{g_k^T H_k y_k y_k^T H_k}{y_k^T H_k y_k}\right] (y_k + g_k)$$

$$= \left[(y_k^T H_k - y_k^T H_k) + g_k^T H_k - \frac{g_k^T H_k y_k y_k^T H_k}{y_k^T H_k y_k}\right] (y_k + g_k)$$

$$= \left(g_k^T H_k - \frac{g_k^T H_k y_k y_k^T H_k}{y_k^T H_k y_k}\right) (y_k + g_k)$$

$$= g_k^T H_k y_k - \frac{g_k^T H_k y_k y_k^T H_k y_k}{y_k^T H_k y_k} + g_k^T H_k g_k - \frac{g_k^T H_k y_k y_k^T H_k g_k}{y_k^T H_k y_k}$$

$$= g_k^T H_k y_k - g_k^T H_k y_k + g_k^T H_k g_k - \frac{g_k^T H_k y_k y_k^T H_k g_k}{y_k^T H_k y_k}$$

$$= g_k^T H_k g_k - \frac{g_k^T H_k y_k y_k^T H_k g_k}{y_k^T H_k y_k}$$

$$= g_k^T \left[H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}\right] g_k$$

$$= g_k^T \left[H_K - \frac{H_k (g_{k+1} - g_k)(g_{k+1} - g_k)^T H_k}{(g_{k+1} - g_k)^T H_k (g_{k+1} - g_k)}\right] g_k$$

$$= g_k^T \left[H_k - \frac{H_k (g_{k+1} g_{k+1}^T - g_{k+1} g_k^T - g_k g_{k+1}^T + g_k g_k^T) H_k}{(g_{k+1} H_k - g_k H_k)^T (g_{k+1} - g_k)}\right] g_k$$

$$= g_k^T \left[H_k - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}\right] g_k$$

Which gives:

$$g_{k+1}^T H_{k+1} g_{k+1} = \frac{[g_k^T H_k g_k][g_{k+1}^T H_k g_{k+1}]}{[g_k^T H_k g_k] + [g_{k+1}^T H_k g_{k+1}]}$$

By finding the inverse number of the expression, we obtain

$$\frac{1}{g_{k+1}^T H_{k+1} g_{k+1}} = \frac{1}{g_{k+1}^T H_k g_{k+1}} + \frac{1}{g_k^T H_k g_k} \quad (42)$$

using (41) and (42), then (40) becomes:

$$Tr(B_{k+1}) = Tr(B_k) + \frac{\|g_{k+1}\|^2}{g_{k+1}^T H_{k+1} g_{k+1}} - \frac{\|g_k\|^2}{g_k^T H_k g_k} - \frac{\|g_{k+1}\|^2}{g_{k+1}^T H_k g_{k+1}} + \frac{\|y_k\|^2}{\alpha s_k^T y_k} \quad (43)$$

By recurrence, we obtain:

$$Tr(B_{k+1}) = Tr(B_0) + \frac{\|g_{k+1}\|^2}{g_{k+1}^T H_{k+1} g_{k+1}} - \frac{\|g_0\|^2}{g_0^T H_0 g_0} - \sum_{j=0}^{k} \frac{\|g_{j+1}\|^2}{g_{j+1}^T H_j g_{j+1}} + \sum_{j=0}^{k} \frac{\|y_j\|^2}{\alpha s_j^T y_j} \quad (44)$$

Therefore, by lemma (3.2), there exists a positive number N which is independent of k, such that

$$Tr(B_{k+1}) \le \frac{\|g_{k+1}\|^2}{g_{k+1}^T H_{k+1} g_{k+1}} - \sum_{j=0}^{k} \frac{\|g_{j+1}\|^2}{g_{j+1}^T H_j g_{j+1}} + Nk \quad (45)$$

In the left part, we will prove that if the theorem does not hold, then the sum of the last two term in (45) is negative.

Now consider the trace of H_{k+1}, from (38), we have

$$Tr(H_{k+1}) = Tr(H_0) - \sum_{j=0}^{k} \frac{\|H_j y_j\|^2}{y_j^T H_j y_j} + \sum_{j=0}^{k} \frac{\|s_j^*\|^2}{s_j^{*T} y_j} \quad (46)$$

Since H_{k+1} is positive definite, the right hand side of (46)is positive. By lemma (3.2) there exists $n > 0$ which is independent of k such that

$$\sum_{j=0}^{k} \frac{\|H_j y_j\|^2}{y_j^T H_j y_j} \leq \frac{k}{n} \quad (47)$$

Note that

$$\left(y_j^T H_j y_j\right)^2 \leq \|H_j y_j\|^2 \|y_j\|^2 \quad (48)$$

and

$$y_j^T H_j y_j > g_{j+1}^T H_j g_{j+1} \quad (49)$$

By the positive definiteness of H_j and exact line search, then by using (49), (48) and (47) in turn, we obtain:

$$\sum_{j=0}^{k} \frac{g_{j+1}^T H_j g_{j+1}}{\|y_j\|^2} \leq \sum_{j=0}^{k} \frac{y_j^T H_j y_j}{\|y_j\|^2} \leq \frac{\|H_j y_j\|^2}{y_j^T H_j y_j} \leq \frac{k}{n} \quad (50)$$

By using Cauchy-Schwarz inequality and (50)

$$\sum_{j=0}^{k} \frac{\|g_{j+1}\|^2}{g_{j+1}^T H_j g_{j+1}} \geq \left(\sum_{j=0}^{k} \frac{\|g_{j+1}\|}{\|y_j\|}\right)^2 / \sum_{j=0}^{k} \frac{g_{j+1}^T H_j g_{j+1}}{\|y_j\|^2}$$

$$\geq \frac{n}{k}\left(\sum_{j=0}^{k} \frac{\|g_{j+1}\|}{\|y_j\|}\right)^2 \quad (51)$$

Now suppose that the theorem is not true, that is, there exists $\varepsilon > 0$ such that for all sufficiently large *k*,

$$\|g_k\| \geq \varepsilon \quad (52)$$

Also, by lemma (3.3), there exists a constant $\eta > 0$ such that

$$\|x^* - x\| \leq (x^* - x)^T\left(g(x^*) - g(x)\right) / \eta$$

$$f(x_k) - f(x_{k+1}) \geq \frac{1}{2}\eta\|s_k^*\|^2$$

which gives $\|s_k^*\| \to 0$ and further $\|y_k\| \to 0$. Then, by (51) and (52) we deduce, for any *k* sufficiently large, that

$$\sum_{j=0}^{k} \frac{\|g_{j+1}\|^2}{g_{j+1}^T H_j g_{j+1}} > Nk \quad (53)$$

The above inequality implies that the sum of the last two terms in (45) is negative.

By (53) and (45), we immediately obtain:

$$Tr(B_{k+1}) < \frac{\|g_{k+1}\|^2}{g_{k+1}^T H_{k+1} g_{k+1}} \quad (54)$$

Note that, for a symmetric positive definite matrix, the inverse of trace is the lower bound of the last eigenvalue of inverse of the matrix. Then, it follows from (54) that

$$\frac{g_{k+1}^T H_{k+1} g_{k+1}}{\|g_{k+1}\|^2} < \mu, \quad (55)$$

Where μ is the lower bound of the last eigenvalue of H_{k+1}. However, from the property of Rayleigh quotient [9], we have

$$\frac{g_{k+1}^T H_{k+1} g_{k+1}}{\|g_{k+1}\|^2} > \mu, \quad (56)$$

which contradicts (55). This contradiction proves that $\{x_k\}$ converges to x^* and that our theorem holds.

4. Local Linear Convergence of $\alpha - DFP$ Method

Now, we shall prove the local linear convergence of $\alpha - DFP$ method of equivalent formula (18) for α>0 under exact line search.

The $\alpha - DFP$ iteration we consider is:

$$x_{k+1} = x_k - \alpha B_k^{-1} F_{(x)} \quad (57)$$

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k^*) y_k^T + y_k (y_k - B_k s_k^*)^T}{y_k^T s_k^*} - \frac{(y_k - B_k s_k^*)^T s_k^{*T}}{(y_k^T s_k^*)^2}(y_k y_k^T) \quad (58)$$

So that replace $\nabla f(x)$ and $\nabla^2 f(x)$ by $F(x)$ and $F'(x)$ respectively.

In this discussion of this subsection, we need the following assumption:

Assumption (4.1)

(A): The mapping $F: R^n \to R$ is continuously differentiable in open convex set $D \subset R^n$.

(B): There is x^* in D such that $F(x^*) = 0$ and $F'(x^*)$ is nonsingular. $(\boldsymbol{C})$: $F'(x)$ satisfies the Lipshiz condition at x^*, that is, there exists a constant $\gamma > 0$ such that

$$\|F'(x) - F'(x^*)\| \leq \gamma\|x - x^*\|, x \in D$$

We begin with some general converges results.

Lemma (4.2) [9]

Let $F: R^n \to R^n$satisfy assumption (A). Then for any $u, v, x^* \in D \subset R^n$ we have

$$\|\mathrm{F}(u) - F(v) - F'(x^*)(u - v)\|$$

$$\leq \left[sup_{\alpha \leq 1}\left\|F'\left(v + t(u - v)\right) - F'(x^*)\right\|\right]\|u - v\| \quad (59)$$

Furthermore, assume that F' satisfy assumption (C), then

$$\|F(u) - F(v) - F'(x^*)(u - v)\| \leq \gamma\sigma(u, v)\|u - v\| \quad (60)$$

and

$$\|F(u) - F(v) - F'(x^*)\| \leq \gamma \frac{\|u - x^*\| + \|v - x^*\|}{2}\|u - v\|, \quad (61)$$

where $(u, v) = \max\{\|u - x^*\|, \|v - x^*\|\}. \quad (62)$

Lemma (4.3)

Let $y, s^* \in R^n, y, s^* \neq 0,$ and $\rho \in (0, 1)$. If $\|y - s^*\| \leq \rho\|y\|,$

then $y^T s^*$ is positive and

$$\left|1-\frac{\|s^*\|}{\|y\|}\right| \le \rho, 1-\left(\frac{y^T s^*}{\|y\|\|s^*\|}\right)^2 \le \rho^2 \quad (63)$$

Conversely, if $y^T s^* > 0$ and (63) holds, then

$$\|y-s^*\| \le 3\rho\|y\| \quad (64)$$

Theorem (4.4) [9]

Let $F: R^n \to R^n$ is satisfy the assumptions (A), (B) and (C) in (4.1), U an update function such that for all $(x_k, B_k) \in dom\ U$ and

$$B_{k+1} \in U(x_k, B_k), \text{ we have} \quad (65)$$

$$\|B_{k+1}-F'(x^*)\| \le \|B_k - F^*(x^*)\| + \frac{\gamma}{2}(\|x_{k+1}-x^*\| + \|x_k - x^*\|), \quad (66)$$

where γ is some constant, or that

$$\|B_{k+1}-F'(x^*)\| \le [1+\rho_1\sigma(x_k,x_{k+1})]\|B_k - F'(x^*)\| + \rho_2\sigma(x_k,x_{k+1}) \quad (67)$$

where ρ_1 and ρ_2 are some constants, and

$$\sigma(x_k,x_{k+1}) = \max\{\|x_k - x^*\|, \|x_{k+1}-x^*\|\}. \quad (68)$$

Then, there exist constants ϵ and δ, such that, for all $\|x_0 - x^*\| < \epsilon$ and $\|B_0 - F'(x^*)\| < \delta$, the iteration (57) and (65) is well-defined, and $\{x_k\}$ converges to x^* linearly.

To study the local convergence of $\alpha - DF$ method, it is required to estimate $\|B_{k+1} - \nabla^2 f(x^*)\|$.

As show in the following theorem, there is a matrix $P = I - \frac{s_k^* y_k^T}{y_k^T s_k^*}$ in $B_{k+1} - \nabla^2 f(x^*)$. Since

$$\|P\|_2 = \frac{\|s_k^*\|\|y_k\|}{y_k^T s_k^*}, \quad (69)$$

It is a secant of the angle between y_k and s_k^*. In general, y_k and s_k^* is not parallel. So $\|P\|_2$ may be quite big, and it is not suitable to estimate $\|B_{k+1} - \nabla^2 f(x^*)\|$ by means of ℓ_2 −norm. However, near x^*,

$f(x)$ closes a quadratic function, and hence $A^{-\frac{1}{2}} y_k$ and $A^{\frac{1}{2}} s_k^*$ are approximately parallel, where $A = \nabla^2 f(x^*)$. In motivates us some weighted norm to estimate $\|B_{k+1} - \nabla^2 f(x^*)\|$. Then, we have

$$\|E\|_{\alpha-DFP} = \|E\|_{A^{-\frac{1}{2}},F} = \left\|A^{-\frac{1}{2}} E A^{-\frac{1}{2}}\right\|_F. \quad (70)$$

Below, we first develop the linear convergence of $\alpha - DFP$.

Theorem (4.5)

Let $f: R^n \to R$ satisfy Assumption (c) in (4.1). Also let

$$\mu\gamma\sigma(x_k,x_{k+1}) \le \frac{1}{3} \quad (71)$$

In a neighborhood of x^*, where $\mu = \|[\nabla^2 f(x^*)]^{-1}\|$, $\sigma(x_k,x_{k+1}) = \max\{\|x_k - x^*\|, \|x_{k+1}-x^*\|\}\|$. then, there exist $\epsilon > 0$ and $\delta > 0$ such that for $\|x_0 - x^*\| < \epsilon$ and $\|B_0 - \nabla^2 f(x^*)\|_{\alpha-DFP} < \delta$, the iteration (57) and equivalent formula (18) of $\alpha - DFP$ method is well-defined, and the produced sequence $\{x_k\}$ converges to x^* linearly.

Proof:

Based on the lemma (4.4), to prove the linearly convergence of $\alpha - DFP$ method, it is enough to prove

$$\|B_{k+1} - \nabla^2 f(x^*)\|_{\alpha-DFP} < [1 + \rho_1\sigma(x_k,x_{k+1})]\|B_k - \nabla^2 f(x^*)\|_{\alpha-DFP} + \rho_2\sigma(x_k,x_{k+1}) \quad (72)$$

where ρ_1 and ρ_2 are positive constants independent of x_k and x_{k+1}, σ is defined by (68).

Le $A = \nabla^2 f(x^*)$ and $M = [\nabla^2 f(x^*)]^{-\frac{1}{2}}$, which is $F'(x^*)$ and $[F'(x^*)]^{-\frac{1}{2}}$ respectively, also are symmetric positive definite matrices.

From (57) and the formula (18) of $\alpha - DFP$, it follows that

$$B_{k+1} - A = P^T(B_k - A)P + \frac{(y_k - As_k^*)y_k^T + y_k(y_k - As_k^*)^T P}{y_k^T s_k^*} \quad (73)$$

where

$$P = I - \frac{s_k^* y_k^T}{y_k^T s_k^*} \quad (74)$$

Thus, from (73), one has

$$\|B_{k+1} - A\|_{\alpha-DFP} \le \|P^T(B_k - A)P\|_{\alpha-DFP} + \left\|\frac{(y_k - As_k^*)y_k^T}{y_k^T s_k^*}\right\|_{\alpha-DFP} + \left\|\frac{y_k(y_k - As_k^*)^T P}{y_k^T s_k^*}\right\|_{\alpha-DFP} \quad (75)$$

Note that $\|P\|_2$ is defined by (69).

The first term on the right hand side of (75) can be estimate as:

$$\|P^T(B_k - A)P\|_{\alpha-DFP} \le \left\|A^{\frac{1}{2}} P A^{-\frac{1}{2}}\right\|_2^2 \|B_k - A\|_{\alpha-DFP}$$

$$\le \frac{1}{z^2}\|B_k - A\|_{\alpha-DFP} \quad (76)$$

moreover, for the rest two terms on the right hand side of (75) and by (70) we have:

$$\left\|\frac{(y_k - As_k^*)y_k^T}{y_k^T s_k^*}\right\|_{\alpha-DFP} = \left\|\frac{A^{-\frac{1}{2}}(y_k - As_k^*)A^{-\frac{1}{2}} y_k^T}{y_k^T s_k^*}\right\|_F$$

$$\left\|\frac{(y_k - As_k^*)y_k^T}{y_k^T s_k^*}\right\|_{\alpha-DFP} \le \frac{1}{z}\frac{\left\|A^{-\frac{1}{2}} y_k - A^{\frac{1}{2}} s_k^*\right\|}{\left\|A^{\frac{1}{2}} s_k^*\right\|} \quad (77)$$

and

$$\left\|\frac{y_k(y_k - As_k^*)^T P}{y_k^T s_k^*}\right\|_{\alpha-DFP} = \left\|\frac{A^{-\frac{1}{2}} y_k(y_k - As_k^*)^T P A^{-\frac{1}{2}}}{y_k^T s_k^*}\right\|_F$$

$$\left\|\frac{y_k(y_k - As_k^*)^T P}{y_k^T s_k^*}\right\|_{\alpha-DFP} \le \frac{1}{z^2} \cdot \frac{\left\|A^{-\frac{1}{2}} y_k - A^{\frac{1}{2}} s_k^*\right\|}{\left\|A^{\frac{1}{2}} s_k^*\right\|} \tag{78}$$

where

$$z = \frac{y_k^T s_k^*}{\left\|A^{-\frac{1}{2}} y_k\right\| \left\|A^{\frac{1}{2}} s_k^*\right\|} = \frac{\left(A^{-\frac{1}{2}} y_k\right)^T \left(A^{\frac{1}{2}} s_k^*\right)}{\left\|A^{-\frac{1}{2}} y_k\right\| \left\|A^{\frac{1}{2}} s_k^*\right\|} \tag{79}$$

which by lemma (4.3) implies the curvature condition

$$\left(A^{-\frac{1}{2}} y_k\right)^T \left(A^{\frac{1}{2}} s_k^*\right) = y_k^T s_k^* > 0$$

Now, we estimate $\|B_{k+1}\|_{\alpha-DFP}$ by using (76), (77) and (78), we have

$$\|B_{k+1} - A\|_{\alpha-DFP} \le \frac{1}{z^2} \|B_k - A\|_{\alpha-DFP} + \left(\frac{1}{z^2} + \frac{1}{z}\right) \frac{\left\|A^{-\frac{1}{2}} y_k - A^{\frac{1}{2}} s_k^*\right\|}{\left\|A^{\frac{1}{2}} s_k^*\right\|} \tag{80}$$

Note from lemma (4.2) that

$$\frac{\left\|A^{-\frac{1}{2}} y_k - A^{\frac{1}{2}} s_k^*\right\|}{\left\|A^{\frac{1}{2}} s_k^*\right\|} \le \frac{\left\|A^{-\frac{1}{2}}\right\| \|y_k - As_k^*\|}{\|s_k^*\| \,/\, \left\|A^{-\frac{1}{2}}\right\|}$$

$$= \|A^{-1}\| \frac{\|y_k - As_k^*\|}{\|s_k^*\|}$$

$$= \mu \frac{\|y_k - As_k^*\|}{\|s_k^*\|}$$

$$\le \mu\, \gamma\, \sigma(x_k, x_{k+1}) \le \frac{1}{3} \tag{81}$$

Since $\|y_k - As_k^*\| \le \|s_k^*\| \mu\, \gamma\, \sigma(x_k, x_{k+1}) \le \frac{1}{3} \|s_k^*\|$

By lemma (4.3), we have:

$$1 - z^2 \le \left[\mu \frac{\|y_k - As_k^*\|}{\|s_k^*\|}\right]^2 \le [\mu\, \gamma\, \sigma(x_k, x_{k+1})]^2.$$

Consequently, if x_k and x_{k+1} are in the neighborhood of x^*, then

$$1 - z^2 \le [\mu\, \gamma\, \sigma(x_k, x_{k+1})]^2 < \frac{1}{2},$$

$$\frac{1}{z^2} < \frac{1}{2} < \mu\, \gamma\, \sigma(x_k, x_{k+1})$$

$$\frac{1}{z^2} = 1 + \frac{1 - z^2}{z^2} < 1 + \frac{[\mu\, \gamma\, \sigma(x_k, x_{k+1})]^2}{\mu\, \gamma\, \sigma(x_k, x_{k+1})}$$

$$= 1 + \mu\, \gamma\, \sigma(x_k, x_{k+1})$$

So, the two terms in (80) satisfy respectively.

$$\frac{1}{z^2} \|B_k - A\|_{\alpha-DFP} < [1 + \mu\, \gamma\, \sigma(x_k, x_{k+1})] \|B_k - A\|_{\alpha-DFP} \tag{82}$$

and

$$\left(\frac{1}{z^2} + \frac{1}{z}\right) \frac{\left\|A^{-\frac{1}{2}} y_k - A^{\frac{1}{2}} s_k^*\right\|}{\left\|A^{\frac{1}{2}} s_k^*\right\|}$$

$$\le \left[\left(1 + \mu\gamma\sigma(x_k, x_{k+1})\right) + \left(\sqrt{1 + \mu\gamma\sigma(x_k, x_{k+1})}\right)\right] \mu\gamma\sigma(x_k, x_{k+1}) \tag{83}$$

combining (82) with (83) into (80), we have:

$$\|B_{k+1} - A\|_{\alpha-DFP}$$

$$\le [1 + \mu\, \gamma\, \sigma(x_k, x_{k+1})] \|B_k - A\|_{\alpha-DFP} + \left(\frac{4}{3} + \sqrt{\frac{4}{3}}\right) [\mu\, \gamma\, \sigma(x_k, x_{k+1})]$$

$$< [1 + \mu\, \gamma\, \sigma(x_k, x_{k+1})] \|B_k - A\|_{\alpha-DFP} + 3\, \mu\, \gamma\, \sigma(x_k, x_{k+1})$$

Which completes the proof by applying lemma (4.4) with $\rho_1 = \mu\, \gamma$ and $\rho_2 = 3\, \mu\, \gamma$.

5. Super Linear Convergence of $\alpha - DFP$ Method

Now, we shall prove the super linear convergence of the $\alpha - DFP$ method. the convergence analysis in this section mainly Dennis and Mor'e [2]. The super linear convergence of the sequence $\{x_k\}$ generated by the iteration (57) is generally characterized by the following theorem.

Theorem (5.1) [2]

Let $F: R^n \to R^n$ is satisfy (A) and (B) in Assumption (4.1). Let $\{B_k\}$ be a sequence, of nonsingular matrices. Suppose for $x_0 \in D$, that the iteration generated by (57) remain in D. $x_k \ne x^* (\forall k \ge 0)$. Suppose also that $\{x_k\}$ converges to x^*. Then $\{x_k\}$ converges to x^* at super linear rate if and only if

$$\lim_{k\to\infty} \frac{\left\|\left(B_k - F'(x^*)\right)(s_k^*)\right\|}{\|s_k^*\|} = 0 \tag{84}$$

Theorem (5.1) indicates that if B_k converges to $F'(x^*)$ along the direction s_k^*, then $\alpha - DFP$ method converges super linearly. This theorem is very important in analysis of $\alpha - DFP$. Equation (84)is called the Dennis and Mor'e characterization of super linear convergence.

To apply theorem (5.1), we need a refinement estimate $\|B_{k+1} - F'(x^*)\|$ which is established with the help of the following lemmas.

Lemma (5.2)

Let $M \in R^{n\times n}$ be a nonsingular symmetric matrix, If, for

$\lambda \in \left[0, \frac{1}{3}\right]$,

The inequality

$$\|My_k - M^{-1}s_k^*\| \le \lambda \|M^{-1}s_k^*\| \quad (85)$$

holds, then for any non-zero matrix $E \in R^{n\times n}$, we have:

$$\text{(a)}: (1-\lambda)\|M^{-1}s_k^*\|^2 \le y_k^T s_k^* \le (1+\lambda)\|M^{-1}s_k^*\|^2. \quad (86)$$

$$\text{(b)}: \left\|E\left[I - \frac{(M^{-1}s_k^*)(M^{-1}s_k^*)^T}{y_k^T s_k^*}\right]\right\|_F \le \sqrt{1-\rho\vartheta^2}\|E\|_F. \quad (87)$$

$$\text{(C)}: \left\|E\left[I - \frac{(M^{-1}s_k^*)(My_k)^T}{y_k^T s_k^*}\right]\right\|_F$$

$$\le \left[\sqrt{1-\rho\vartheta^2} + (1-\lambda)^{-1}\frac{\|My_k - M^{-1}s_k^*\|}{\|M^{-1}s_k^*\|}\right]\|E\|_F \quad (88)$$

Where

$$\rho = \frac{1-2\lambda}{(1-\lambda)^2} \in \left[1, \frac{4}{3}\right], \vartheta = \frac{\|EM^{-1}s_k^*\|}{\|E\|_F\|M^{-1}s_k^*\|} \in [0,1] \quad (89)$$

Proof:

Note that

$$y_k^T s_k^* = (My_k)^T(M^{-1}\alpha s_k) = (My_k - M^{-1}\alpha s_k)^T M^{-1}\alpha s_k + \|M^{-1}\alpha s_k\|^2 \quad (90)$$

Also, it follows from Cauchy-Schwarz inequality and (85) that

$$|(My_k - M^{-1}\alpha s_k)^T M^{-1}\alpha s_k| \le \|My_k - M^{-1}\alpha s_k\|\|M^{-1}\alpha s_k\|$$

$$= \lambda\|M^{-1}\alpha s_k\|^2 \quad (91)$$

From (90) and (91), we get

$$y_k^T s_k^* \le \lambda\|M^{-1}\alpha s_k\|^2 + \|M^{-1}\alpha s_k\|^2 = (1+\lambda)\|M^{-1}\alpha s_k\|^2$$

And, in same way, we have

$$(1-\lambda)\|M^{-1}\alpha s_k\|^2 \le y_k^T s_k^*$$

Which gives the first result (a).

Now, we will prove (b) By using the property of the Frobenius norm of a rank-one update.

$$\|A + xy^T\|_F^2 = \|A\|_F^2 + 2y^T A^T x + \|x\|^2\|y\|^2$$

To prove (b) we need the following property

$$\|E(I - uv^T)\|_F^2 = \|E\|_F^2 - 2vE^T Eu + \|Eu\|^2\|v\|^2.$$

In particular,

$$\text{Let } u = \frac{M^{-1}\alpha s_k}{y_k^T \alpha s_k}, v = M^{-1}\alpha s_k$$

$$\left\|E\left[I - \frac{(M^{-1}\alpha s_k)(M^{-1}\alpha s_k)^T}{y_k^T \alpha s_k}\right]\right\|_F^2$$

$$= \|E\|_F^2 - 2\frac{(M^{-1}\alpha s_k)^T E^T E(M^{-1}\alpha s_k)}{y_k^T \alpha s_k} + \left\|\frac{EM^{-1}\alpha s_k}{y_k^T \alpha s_k}\right\|^2 \|M^{-1}\alpha s_k\|^2$$

$$= \|E\|_F^2 + (-2y_k^T \alpha s_k + \|M^{-1}\alpha s_k\|^2)\frac{\|EM^{-1}\alpha s_k\|^2}{(y_k^T \alpha s_k)^2}$$

By using (a) and (89), we get:

$$\|M^{-1}\alpha s_k\|^2 - 2y_k^T \alpha s_k \le \frac{1}{1-\lambda}y_k^T \alpha s_k - 2\frac{1-\lambda}{1-\lambda}y_k^T \alpha s_k = -\left(\frac{1-2\lambda}{1-\lambda}\right)y_k^T \alpha s_k$$

And therefore

$$\left\|E\left[I - \frac{(M^{-1}\alpha s_k)(M^{-1}\alpha s_k)^T}{y_k^T \alpha s_k}\right]\right\|_F^2 \le \|E\|_F^2 - \frac{1-2\lambda}{1-\lambda}y_k^T \alpha s_k\left(\frac{\|EM^{-1}\alpha s_k\|^2}{(y_k^T \alpha s_k)^2}\right)$$

$$\le \left[1 - \frac{1-2\lambda}{(1-\lambda)^2}\frac{\|EM^{-1}\alpha s_k\|^2}{\|M^{-1}\alpha s_k\|^2\|E\|_F^2}\right]\|E\|_F^2,$$

from (89) again, we get:

$$\left\|E\left[I - \frac{(M^{-1}\alpha s_k)(M^{-1}\alpha s_k)^T}{y_k^T \alpha s_k}\right]\right\|_F^2 \le 1 - \rho\vartheta^2\|E\|_F^2$$

Which shows (b):

$$\left\|E\left[I - \frac{(M^{-1}\alpha s_k)(M^{-1}\alpha s_k)^T}{y_k^T \alpha s_k}\right]\right\|_F \le \sqrt{1-\rho\vartheta^2}\,\|E\|_F.$$

Finally, we prove (c) by means of (b). It enough to prove that

$$\left\|E\frac{M^{-1}\alpha s_k(M^{-1}\alpha s_k - My_k)^T}{y_k^T \alpha s_k}\right\|_F \le (1-\lambda)^{-1}\left(\frac{\|My_k - M^{-1}\alpha s_k\|}{\|M^{-1}\alpha s_k\|}\right)\|E\|_F \quad (92)$$

from (a), we have:

$$\frac{1}{y_k^T \alpha s_k} \le (1-\lambda)^{-1}\frac{1}{\|M^{-1}\alpha s_k\|^2} \quad (93)$$

$$\frac{1}{y_k^T \alpha s_k}\|M^{-1}\alpha s_k\|\,\|M^{-1}\alpha s_k - My_k\|\|E\|_F$$

$$\le (1-\lambda)^{-1}\frac{\|M^{-1}\alpha s_k\|\|M^{-1}\alpha s_k - My_k\|}{\|M^{-1}\alpha s_k\|^2}\|E\|_F$$

$$\le \left[\sqrt{1-\lambda\vartheta^2} + (1-\lambda)^{-1}\frac{\|M^{-1}\alpha s_k - My_k\|}{\|M^{-1}\alpha s_k\|}\right]\|E\|_F$$

which proves (c).

We have known that if $f: R^n \to R$ satisfies Assumption (4.1), then (72) holds.

Then under Assumptions of the theorem (4.5), the preceding lemma can be applied with the sitting

$$A = F'(x^*) = \nabla^2 f(x^*),\, M = [F'(x^*)]^{-\frac{1}{2}} = A^{-\frac{1}{2}},$$

$$\psi_k = \|B_k - \nabla^2 f(x^*)\|_{\alpha-DFP},\, \psi_{k+1} = \|B_{k+1} - \nabla^2 f(x^*)\|_{\alpha-DFP,}$$

and $\xi_k = \max\{\rho_1\sigma(x_k, x_{k+1}), \rho_2\sigma(x_k, x_{k+1})\}$.

Due to the linear convergence of the sequence $\{x_k\}$ gives in theorem (4.5), we have $\sum_{k=1}^{\infty}\xi_k < +\infty$, and can sequnetly by lemma (5.3), there exists a constant $h \geq 0$ such that

$$\lim_{k\to\infty^+}\|B_k - \mathrm{F}'(x^*)\|_{\alpha-DFP} = h \tag{94}$$

Hence

$$\lim_{k\to\infty}\|B_k - F'(x^*)\|_{\alpha-DFP}\ ,$$

is exists.

Lemma (5.3):

Let $\{\phi_k\}$ and $\{\delta_k\}$ be sequences of nonnegative numbers satisfying

$$\phi_{k+1} \leq (1+\delta_k)\phi_k + \delta_k \tag{95}$$

and

$$\sum_{k=1}^{\infty}\delta_k < +\infty \tag{96}$$

then $\{\phi_k\}$ is converges.

These results together then give rise to a refinement estimate of $\|B_{k+1} - F'(x^*)\|_{\alpha-DFP}$ as follows:

Lemma (5.4)

Under the assumption of theorem(4.5), there exist positive constants λ_1, λ_2 and λ_3, such that $\forall x_{k+1} \in N(x^*, \epsilon)$, we have

$$\|B_{k+1} - \nabla^2 f(x^*)\|_{\alpha-DFP} \leq \left[\sqrt{1-\lambda_1\vartheta^2} + \lambda_2\sigma(x_k, x_{k+1})\right]\|B_k - \nabla^2 f(x^*)\|_{\alpha DFP} + \lambda_3\sigma(x_k, x_{k+1}), \tag{97}$$

where σ is defined by (68) and

$$\vartheta_k = \frac{\left\|[\nabla^2 f(x^*)]^{-\frac{1}{2}}\left(B_k - [\nabla^2 f(x^*)]^{-\frac{1}{2}}\right)\alpha s_k\right\|}{\|B_k - \nabla^2 f(x^*)\|_{\alpha-DFP}\left\|[\nabla^2 f(x^*)]^{\frac{1}{2}}\alpha s_k\right\|} \tag{98}$$

Proof:

First m we write $= \nabla^2 f(x^*)$. From (75), we have:

$$\|B_{k+1} - A\|_{\alpha-DFP} \leq \|P^T(B_k - A)P\|_{\alpha-DFP} + \left\|\frac{(y_k - A\alpha s_k)y_k^T}{y_k^T\alpha s_k}\right\|_{\alpha-DFP} + \left\|\frac{y_k(y_k - A\alpha s_k)^T}{y_k^T\alpha s_k}\right\|_{\alpha-DFP}$$

Let

$$R = I - \frac{A^{\frac{1}{2}}\alpha s_k y_k^T A^{-\frac{1}{2}}}{y_k^T\alpha s_k},\, E_k = A^{-\frac{1}{2}}(B_k - A)A^{-\frac{1}{2}} \tag{99}$$

And

$$\|P^T(B_k - A)P\|_{\alpha-DFP} = \left\|\left(A^{-\frac{1}{2}}P^T A^{\frac{1}{2}}\right)\left(A^{-\frac{1}{2}}(B_k - A)A^{-\frac{1}{2}}\right)\left(A^{\frac{1}{2}}PA^{-\frac{1}{2}}\right)\right\|_F$$

$$= \|R^T\, E\, R\|_F$$

Similar to the proof of the theorem (4.5), we known that there exists $\rho_3 > 0$ and $\rho_4 > 0$ such that

$$\left\|\frac{(y_k - A\alpha s_k)y_k^T}{y_k^T\alpha s_k}\right\|_{\alpha-DFP} \leq \frac{1}{z}\frac{\left\|A^{-\frac{1}{2}}y_k - A^{\frac{1}{2}}\alpha s_k\right\|}{\left\|A^{\frac{1}{2}}\alpha s_k\right\|} \leq \rho_3\sigma(x_k, x_{k+1}),$$

$$\left\|\frac{y_k(y_k - A\alpha s_k)^T P}{y_k^T\alpha s_k}\right\|_{\alpha-DFP} \leq \frac{1}{z^2}\frac{\left\|A^{-\frac{1}{2}}y_k - A^{\frac{1}{2}}\alpha s_k\right\|}{\left\|A^{\frac{1}{2}}\alpha s_k\right\|} \leq \rho_4\sigma(x_k, x_{k+1}).$$

If we let $\lambda_3 = \rho_3 + \rho_4$, then (75) becomes:

$$\|B_{k+1} - A\|_{\alpha-DFP} \leq \|R^T E R\|_F + \lambda_3\sigma(x_k, x_{k+1}) \tag{100}$$

Since

$$\frac{\left\|A^{-\frac{1}{2}}y_k - A^{\frac{1}{2}}\alpha s_k\right\|}{\left\|A^{\frac{1}{2}}\alpha s_k\right\|} \leq \mu\,\gamma\,\sigma(x_k, x_{k+1}) \leq \frac{1}{3},$$

Then, by use of lemma (5.2), and from (99), we get:

$$\|R^T E R\|_F \leq \left[\sqrt{1-\rho\vartheta^2} + (1-\lambda)^{-1}\frac{\left\|A^{-\frac{1}{2}}y_k - A^{\frac{1}{2}}\alpha s_k\right\|}{\left\|A^{\frac{1}{2}}\alpha s_k\right\|}\right]\|R^T E\|_F$$

Note that $\|R^T E\|_F = \|E^T R\|_F = \|ER\|_F$, thus, by using lemma (5.2) again, we obtain:

$$\|R^T E R\|_F \leq \left[\sqrt{1-\rho\vartheta^2} + (1-\lambda)^{-1}\frac{\left\|A^{-\frac{1}{2}}y_k - A^{\frac{1}{2}}\alpha s_k\right\|}{\left\|A^{\frac{1}{2}}\alpha s_k\right\|}\right]\left[\sqrt{1-\rho\vartheta^2} + (1-\lambda)^{-1}\frac{\left\|A^{-\frac{1}{2}}y_k - A^{\frac{1}{2}}\alpha s_k\right\|}{\left\|A^{\frac{1}{2}}\alpha s_k\right\|}\right]\|E\|_F.$$

Where ϑ_k is defined by (98), From lemma (5.2) again and by (89), we get

$$\|R^T ER\|_F \leq \left[\sqrt{1-\rho\vartheta^2} + \frac{5}{2}(1-\lambda)^{-1}\frac{\left\|A^{-\frac{1}{2}}y_k - A^{\frac{1}{2}}\alpha s_k\right\|}{\left\|A^{\frac{1}{2}}\alpha s_k\right\|}\right]\|E\|_F$$

$$\leq \left[\sqrt{1-\lambda_1\vartheta^2} + \lambda_2\,\mu\,\gamma\,\sigma(x_k,x_{k+1})\right]\|E\|_F \qquad (101)$$

Where $\lambda_1 = \rho, \lambda_2 = \frac{15}{4}\mu\gamma$. Substitution (101) into (100) and from (99), we deduce the desired result (97). The proof is complete

Finally, using the above four lemmas, we can establish the following super linear convergence theorem of $\alpha - DFP$ method.

Theorem (5.5)

Under the assumption of theorem (4.5), $\alpha - DFP$ method defined by (57) and (58) is convergent super linearly.

Proof:

Since $(1-\lambda_1\vartheta^2)^{\frac{1}{2}} \leq \left(1-\frac{1}{2}\lambda_1\vartheta_k^2\right)$, then (95) can be written as:

$$\|B_{k+1} - A\|_{\alpha-DFP} \leq \sqrt{1-\lambda_1\vartheta_k^2}\,\|B_k - A\|_{\alpha-DFP} + \lambda_2\sigma(x_k,x_{k+1})\,\|B_k - A\|_{\alpha-DFP} + \lambda_3\sigma(x_k,x_{k+1})$$

$$\leq \left(1-\frac{\lambda_1}{2}\vartheta_k^2\right)\|B_k - A\|_{\alpha-DFP} + \lambda_2\sigma(x_k,x_{k+1})\|B_k - A\|_{\alpha-DFP} + \lambda_3\sigma(x_k,x_{k+1})$$

$$\left(\frac{\lambda_1}{2}\right)\vartheta_k^2\|B_k - A\|_{\alpha-DFP} \leq \|B_k - A\|_{\alpha-DFP} - \|B_{k+1} - A\|_{\alpha-DFP} + \lambda_2\|B_k - A\|_{\alpha-DFP}\sigma(x_k,x_{k+1}) + \lambda_3\sigma(x_k,x_{k+1})$$

Summing the above from $k = 1$ to infinity gives:

$$\frac{1}{2}\lambda_1\sum_{k=1}^{\infty}\vartheta_k^2\|B_k - A\|_{\alpha-DFP} \leq \sum_{k=1}^{\infty}\|B_k - A\|_{\alpha-DFP} - \sum_{k=1}^{\infty}\|B_{k+1} - A\|_{\alpha-DFP} + \lambda_2\sum_{k=1}^{\infty}\sigma(x_k,x_{k+1})\,\|B_k - A\|_{\alpha-DFP} + \lambda_3\sum_{k=1}^{\infty}\sigma(x_k,x_{k+1})$$

Since, from theorem (4.5), $\{x_k\}$ is linearly convergent, then $\sum_{k=1}^{\infty}\sigma(x_k,x_{k+1}) < \infty$.

Also, since $\{\|B_k - A\|_{\alpha-DFP}\}$ is bounded, then

$$\frac{1}{2}\lambda_1\sum_{k=1}^{\infty}\vartheta_k^2\|B_k - A\|_{\alpha-DFP} < \infty$$

By (94), the $\lim_{k\to\infty}\|B_k - A\|_{\alpha-DFP}$ exists.

Hence, if some subsequence of $\{\|B_k - A\|_{\alpha-DFP}\}$ converges to zero.

The whole sequence converges to zero.

Therefore,

$$\lim_{k\to\infty}\frac{\|(B_k - A)s_k^*\|}{\|s_k^*\|} = 0,$$

Which proves the super linear convergence of $\{x_k\}$ by theorem (5.1). Otherwise, there exists a positive constant τ such that

$\|B_k - A\|_{\alpha-DFP} \geq \tau, \forall\, k \geq k_{0,}$ then

$$\frac{1}{2}\lambda_1\tau\sum_{k=1}^{\infty}\vartheta_k^2 < \infty$$

Since $\lambda_1 = \frac{4}{3} > 0$, it follows that $\lim_{k\to\infty}\vartheta_k = 0$

Furthermore, we have:

$$\lim_{k\to\infty}\frac{\|(B_k - A)s_k^*\|}{\|s_k^*\|} \leq \lim_{k\to\infty}\frac{\left\|A^{\frac{1}{2}}\right\|\left\|A^{-\frac{1}{2}}(B_k - A)\alpha s_k\right\|}{\left\|A^{\frac{1}{2}}\right\|^{-1}\left\|A^{\frac{1}{2}}\alpha s_k\right\|}$$

$$= \lim_{k\to\infty}\|A\|\cdot\|B_k - A\|\frac{\left\|A^{-\frac{1}{2}}(B_k - A)\alpha s_k\right\|}{\|B_K - A\|\left\|A^{\frac{1}{2}}\alpha s_k\right\|}$$

$$= \lim_{k\to\infty}\|A\|\cdot\|B_k - A\|\,\vartheta_k$$

Where ϑ_k is defined by (98).

Then, by using $\vartheta_k \to 0$, we immediately obtain:

$$\lim_{k\to\infty}\frac{\|(B_k - A)\alpha s_k\|}{\|\alpha s_k\|} = 0.$$

Hence, $\{x_k\}$ is convergent super linearly, we complete the proof.

6. Numerical Results

This section is devoted to numerical experiments. Our purpose was to check whether the modified $\alpha - DFP$algorithm provide improvements on the corresponding standard DFP algorithm. The programs were written in MATLAP. The reason for their selection is that the problems appear to have been used in standard problems in most the literature these functions represent a result of application in the branch of technology and industry.

The test functions are chosen as follows:

$1 - f(x) = (1-x_1)^2 + (1-x_2)^2$. [1]

2 − A quadratic function. [10]

$$f(x) = \sum_{i=1}^{4} (10^{i-1}x_i^4 + x_i^3 + 10^{1-i}x_i^2)$$

3 − Rosen brook's function. [4]

$$f(x) = (1-x_1)^2 + (x_2 - x_1)^2.$$

4 − Rosenbroc'k cliff function [8]

$$f(x) = 10^{-4}(x_1-3)^2 - (x_1-x_2) + e^{20(x_1-x_2)}.$$

5 − Generalized Edger function. [1]

$$f(x) = \sum_{i=1}^{n/2}[(x_{2i-1}-2)^4 + (x_{2i-1}-2)^2x_{2i}^2 + (x_{2i}+1)^2].$$

6 − Extended Himmelbla function [1]

$$f(x) = \sum_{i=1}^{n/2}(x_{2i-1}^2 + x_{2i} - 11)^2 + (x_{2i-1} + x_{2i}^2 - 7)^2.$$

7 − Rosen rock's function [6]

$$f(x) = \sum_{i=1}^{n/2}[100(x_i - x_i^3)^2 + (1-x_i)^2].$$

8 − Trigonometric function [1]

$$f(x) = \sum_{i=1}^{n}\left[n - \sum_{j=1}^{n}\cos x_j + i(1-\cos x_i) - \sin x_i + e^{xi} - 1\right]^2.$$

9 − Extended Rosen rock function [1]

$$f(x) = \sum_{i=1}^{n/2} c(x_{2i} - x_{2i-1}^2)^2 + (1-x_{2i-1})^2. \ c = 100$$

10 − Watson function [6]

$$F(x) = \sum_{i=1}^{j} f_i^2(x)$$

$$f_i(x) = \sum_{j=2}^{3}(j-1)x_j t_j^{j-2} - \left(\sum_{j=1}^{3} x_j t_j^{j-1}\right)^2 - 1.$$

$$t_j = \frac{i}{29}. \text{ and}$$

11 − Freudenstein and Roth function [3]

$$f(x) = \{-13 + x_1 + [(5-x_2)x_2 - 2]x_2\}^2$$
$$+\{-29 + x_1 + [(x_2+1)x_2 - 14]x_2\}^2$$

Table 1. *Numerical results for DFP and α − DFP update.*

Fun.	Starting point	Dim.	*DFP*		*α − DFP*		The Best
			Feval	Iter.	Feval	Iter.	
1	$[0;0]^T$	2	6.0962e-016	2	1.7909e-017	2	*DFP*
1	$[-1;-1]^T$	2	2.7900e-020	2	2.7900e-020	2	Same
1	$[0;1]^T$	2	3.2494e-016	2	2.9082e-017	2	*DFP*
2	$[-1;0;-1;0]^T$	4	1.5076e-007	9	9.0146e-007	7	α − *DFP*
2	$[-1;0;0;0]^T$	4	1.8142e-013	3	7.3965e-011	3	α − *DFP*
2	$[-30;10;\cdots]^T$	4	7.9747e-008	34	7.2012e-006	9	α − *DFP*
3	$[0;0]^T$	2	1.7525e-018	20	1.6887e-016	14	α − *DFP*
3	$[0;-5]^T$	2	1.5615e-016	3	7.5140e-015	3	α − *DFP*
3	$[-3;-3]^T$	2	3.0198e-o17	14	2.4041e-017	14	Same
4	$[-0.5;\cdots]^T$	4	0.2011	3	0.2011	3	Same
4	$[0.5;\cdots]^T$	12	0.2004	3	0.2004	3	Same
4	$[0;\cdots]^T$	12	0.2007	3	0.2007	3	Same
5	$[-3;0]^T$	2	5.5433e-015	8	8.8715e-014	3	*α − DFP*
5	$[0;5;\cdots]^T$	18	2.9846e-008	9	3.6384e-008	6	*α − DFP*
5	$[-1;\cdots]^T$	18	1.4714e-010	7	5.3611e-009	6	*α − DFP*
6	$[5;10]^T$	2	8.1785e-012	7	3.2157e-011	6	*α − DFP*
6	$[0;0]^T$	2	1.3697e-013	8	3.1675e-009	8	*α − DFP*
6	$[0;-1]^T$	2	1.1910e-012	8	2.7507e-012	8	*α − DFP*
7	$[-1;1]^T$	8	2.0658e-011	6	1.9815e-010	24	*α − DFP*
7	$[0;1;\cdots]^T$	8	5.1583e-012	6	3.1487e-011	8	*α − DFP*
7	$[0;\cdots]^T$	4	1.7728e-010	2	1.7728e-010	2	Same
8	$[-0.5;\cdots]^T$	12	7.6464e-007	13	4.6941e-006	5	*α − DFP*
8	$[0.5;\cdots]^T$	12	3.7119e-006	13	3.7598e-006	4	*α − DFP*
8	$[2;\cdots]^T$	12	6.8965e-007	21	1.8229e-006	5	*α − DFP*
9	$[1;1;1]^T$	3	0	1	0	1	Same
9	$[-4;\cdots]^T$	3	0.0220	40	0.0118	9	*α − DFP*
9	$[10;10]^T$	2	3.5522	31	40.2765	8	*α − DFP*
10	$[1;1;1;1]^T$	4	6.0707e-017	3	1.0838e-016	3	*α − DFP*
10	$[1;0;1;0]^T$	4	2.1626e-013	3	3.5351e-008	3	*α − DFP*
10	$[0;\cdots]^T$	10	1.6053e-010	2	5.5796e-010	2	*α − DFP*
11	$[5;10]^T$	2	5.7785e-010	12	2.2977e-010	8	*α − DFP*
11	$[-10;20]^T$	2	2.3185e-013	15	1.7572e-013	9	α − *DFP*

Fun.	Starting point	Dim.	DFP		$\alpha-DFP$		The Best
			Feval	Iter.	Feval	Iter.	
11	$[-7;15]^T$	2	1.6761e-012	13	8.3077e-012	10	$\alpha-DFP$

7. Conclusion

In this thesis, we introduce a new modified of the DFP say $\alpha-DFP$ update, we show that under certain circumstances this update preserve the value of the determinant of hessian matrix and without Quasi-Newton or based on the Zhang Xu condition.

Global convergence of the proposed method establishes under exact line search. The proposed method possesses local linearly convergence and super linearly convergence for unconstrained optimization problem.

Numerical results show that the proposed is efficient for unconstrained optimization problem compared the modified $\alpha-DFP$ method with the standard DFP method on same function is selected, which suggests that a good improvement has been achieved.

References

[1] Al-Bayati, A., (1991), A new family of self-scaling variable metric Algorithms for unconstraint optimization, Journal of Educ. and Sci., Iraq, Vol. 12, pp. 25-54.

[2] Dennis J. E., More J., (1974), "A characterization of super linear Convergence and its application to quasi-Newton methods Math and Computation 28 (6) 549-60.

[3] F. Freudenstein and B. Roth,(1962), Numerical solution of system of nonlinear equations, Journal of ACM, Vol. 10, No. 4, pp. 550-556.

[4] H. H. Rosen brock, (1960), An automatic method for finding the greatest least value of a function, Computer Journal, Vol. 3, No. 3, pp. 175-184.

[5] J. E. Dennis, Jr. and Robert B. Schnabel, (1996), Numerical Methods for Unconstrained Optimization and Nonlinear Equations, SIAM, Classics in Applied Mathematics.

[6] Oren, S. S (1973), Self-scaling Variable Metric Algorithms without line search for Unconstrained minimization. Mathematics of computation, 27: 873-885.

[7] Saad S. Mahmood, (2011), $\alpha-BFGS$ update for Unconstrained Opt-imization, Journal of college of Education, No. 1, Almustansiriya University.

[8] Todd M. J., (1984), Quasi-Newton updates in abstract spaces, SIAM Review, 26: 367-377.

[9] W. Sun and Y. Yuan, (2006), "Optimization Theory and Method: Nonlinear Programing", Vol. 1 of Springer optimization and its Applications, Springer, New York, NY, USA.

[10] Yuan, Y., (1990), On a Modified Algorithm for Unconstrained Optimization, Computing Center, Academia Sinica, Beijing, China.