International
Journal
Of Advanced Research in Engineering & Management (IJAREM)

A Modification of Quasi-Newton (DFP) Method for Solving Unconstrained Optimization Problems

Salah Gazi Shareef[1], Ahmed Anwer Mustafa[2]
[1]*Deperatment of Mathematics, Faculty of Science, University of Zakho*
[2]*Deperatment of Mathematics, Faculty of Science, University of Zakho*
Kurdistan Region-Iraq

Abstract— The Quasi-Newton method is a very useful technique for solving minimization problems and has wide applications in many fields. In this paper we develop a new class of DFP method for unconstrained optimization. The given method satisfies the Quasi-Newton condition and positive definite theorem under strong Wolfe line search. Numerical results based on the number of iterations (NOI) and number of function (NOF), have shown that the new method (New5) performs better than standard method of ($DFP^{H/S}$) method.
Keywords— Unconstrained optimizations, DFP method, Quasi-Newton condition, positive definite theorem.

1. INTRODUCTION

Quasi-Newton methods are recognized today as one of the most efficient ways to solve nonlinear unconstrained or bound constrained optimization problems. These methods are mostly used when the second derivative matrix of the objective function is either unavailable or too costly to compute. They are very similar to Newton's method, but avoid the need of computing Hessian matrices by recurring, from iteration to iteration, a symmetric matrix which can be considered as an approximation of the Hessian. They allow, therefore, the curvature of the problem to be exploited in the numerical algorithm, despite the fact that only first derivatives (gradients) and function values are required. We refer the reader to [4, 5, 10, 11] for further motivation and analysis concerning these now classical algorithms.

In this study we consider the unconstrained minimization problem

$$\text{Min } f(x) \tag{1.1}$$

$x \in R^n$

where $f: R^n \to R$ is continuously differentiable. The DFP method for solving (1.1) generates a sequence of iterates $\{x_k\}$ by

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0,1,2,\dots \tag{1.2}$$

where $x_k \in R^n$ is the current iterative and α_k is a step size., we denote $\nabla f(x_k)$ by g_k $f(x_k)$ by f_k, $\nabla^2 f(x_k)$ by G_k and d_k is a search direction, generally descent and generated from either $B_k d_k = -g_k$ is an approximation to the Hessian of f) or $d_k = -H_k g_k$ is used to generate the search direction.

.

$$d_k = \begin{cases} -H_k g_k, & k = 0 \\ -H_k g_{k+1} + \beta_k d_k, & k \ge 1 \end{cases} \tag{1.3}$$

There are mainly different in the choice of the parameter β_k . Some well-known formulas for β_k given below: β_k

$$\beta_k^{HS} = \frac{g_{k+1}^T H_k y_k}{d_k^T y_k} \tag{1.4}$$

$$\beta_k^{FR} = \frac{g_{k+1}^T H_k g_k}{g_k^T g_k} \tag{1.5}$$

$$\beta_k^{PR} = \frac{g_{k+1}^T H_k y_k}{g_k^T g_k} \tag{1.6}$$

$$\beta_k^{CD} = \frac{g_{k+1}^T H_k g_{k+1}}{g_k^T d_k} \tag{1.7}$$

$$\beta_k^{BA2} = \frac{y_k^T H_k y_k}{g_k^T g_k} \quad (1.8)$$

$$\beta_k^{LS} = \frac{g_{k+1}^T H_k y_k}{-d_k^T g_k} \quad (1.9)$$

$$\beta_k^{DY} = \frac{g_{k+1}^T H_k g_{k+1}}{d_k^T y_k} \quad (1.10)$$

$$\beta_k^{HZ} = \left(y_k - 2d_k \frac{\|y_k\|^2}{d_k^T y_k}\right)^T H_k \frac{g_{k+1}}{d_k^T y_k} \quad (1.11)$$

$$\beta_k^{RAMI} = \frac{g_{k+1}^T H_k \left(g_{k+1} - \frac{\|g_{k+1}\|}{\|g_k\|} g_k\right)}{d_k^T (d_k - g_{k+1})} \quad (1.12)$$

$$\beta_k^{AMRI} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} |g_{k+1}^T H_k g_k|}{\|d_k\|^2} \quad (1.13)$$

where g_k and g_{k+1} g are the gradients of $f\ (x)$ at the point x_k and x_{k+1} respectively.The above corresponding methods, HS is known as Hestenes and Steifel [6], FR is Fletcher and Reeves [8], PR is Polakand Ribiere [3], CD is Conjugate Descent [9],BA2 is AL - Bayati, A.Y. and AL-Assady[2],
LS is Liu and Storey [14], DY is Dai and Yuan [13], HZ is Hager and Zhang [12], RAMI is M. Rivaie, A. Abashar, M. Mamat and I. Mohd [7] and lastly AMRI denotes Abdelrhaman Abashar, Mustafa Mamat, Mohd Rivaie and Ismail Mohd[1].
The matrix H_k is required to satisfy the Quasi-Newton equation

$$H_{k+1}\, y_k = v_k \quad (1.14)$$

The remaining sections of the paper are arranged as follows. In section 2 the New proposed method, in section 3, Derivative of New5 Algorithm of Quasi–Newton (DFP) and algorithm. In section 4 numerical results, percentages, graphics and discussion. Lastly, in section 5 conclusions.

2. NEW PROPOSED METHOD OF DFP (NEW5)

The main idea of the method is the use of modified of quasi Newton condition approximation to find a minimum of multiple dimensional nonlinear functions. The following is the modified of quasi Newton condition.

$$H_{k+1} z_k = v_k \quad (2.1)$$

Where $z_k = y_k + \alpha_k (\frac{\theta}{v_k^T y} g_k)$ (2.2)

$$\alpha_k = \frac{v_k^T y_k}{y_k^T y_k}$$

$$\theta = 2(f(x_{k+1}) - f(x_k) - g_k{}^T v_k) \quad (2.3)$$

Where $v_k^T y_k \neq 0$.

3. DERIVATIVE OF NEW5 ALGORITHM OF QUASI–NEWTON (DFP) AND ALGORITHM

DFP update is another typical update which is a rank-two update, i.e., H_{k+1} is formed by adding to H_k two symmetric matrices, each of rank one. Let us Consider the symmetric rank-two update

$$H_{k+1} = H_k + \alpha_1 a a^T + \alpha_2 b b^T \quad (3.1)$$

Clearly, a and b are not uniquely determined, but their obvious choices are

$$a = v_k,\ b = H_k z_k \quad (3.2)$$

Then, from (3.1), we have

$$\alpha_1 = \frac{1}{a^T y_k} = \frac{1}{v_k{}^T z_k} \tag{3.3}$$

$$\alpha_2 = -\frac{1}{b^T z_k} = \frac{1}{z_k{}^T H_k z_k} \tag{3.4}$$

Put (3.3) and (3.4) in (3.1) we get

$$H_{k+1} = H_k + \frac{1}{v_k{}^T z_k} a a^T + \frac{1}{z_k{}^T H_k z_k} b b^T \tag{3.5}$$

Put (3.2) in (3.5) we have

$$H_{k+1} = H_k + \frac{v_k v_k{}^T}{v_k{}^T z_k} - \frac{(H_k z_k)(H_k z_k)^T}{z_k{}^T H_k z_k} \tag{3.6}$$

The formula (3.6) is a New (New5) Algorithm of Quasi–Newton Symmetric Rank Two (DFP) update as follows.

3.1 ALGORITHM OF (NEW5)

Step (1): Given $x_0 \in R^n$ an initial point, $H_0 \in R^{n\times n}$ a symmetric and
Positive definite matrix, $\epsilon > 0$ a termination scalar. $k = 0$ and compute $f(x_0), g_0$

Step (2): Compute $g_k = \nabla f(x_k)$

Step (3): Compute $d_k = -H_k g_k$.

Step (4): Compute the step size α_k=arg min $f(x_k + \alpha_k d_k)$ and compute $g_{k+1} = \nabla f(x_{k+1})$,

If $\|g_{k+1}\| \leq \epsilon$ stop

Step (5): Set $v_k = \alpha_k d_k,\ x_{k+1} = x_k + v_k,\ y_k = g_{k+1} - g_k,$

$$z_k = y_k + \alpha_k \left(\frac{\theta}{v_k^T y} g_k\right),\quad \alpha_k = \frac{v_k^T y_k}{y_k^T y_k}$$

$$\theta = 2(f(x_{k+1}) - f(x_k) - g_k{}^T v_k)$$

$$H_{k+1} = H_k + \frac{v_k v_k{}^T}{v_k z_k{}^T} - \frac{(H_k z_k)(H_k z_k)^T}{z_k{}^T H_k z_k}.$$

Step (6): $d_{k+1} = -H_{k+1} g_{k+1} + \beta_k d_k$

Step (7): If $|g_{k+1}{}^T g_k| \geq 0.2\|g_{k+1}\|^2$ go to step (3) else continue.
Step (8): Set $k = k + 1$, go to step (4).

Theorem3.1: In the New5 algorithm applied to the quadratic with Hessian $Q = Q^T$, we have

$$H_{k+1} z_i = v_i\ , 0 \leq i \leq k.$$

Proof: Multiplying both sides of equation (3.6) by z_k from left we have

$$H_{k+1}z_k = H_k z_k + \frac{v_k v_k^T}{v_k z_k^T} z_k - \frac{(H_k z_k)(H_k z_k)^T}{z_k^T H_k z_k} z_k \quad (3.7)$$

$$\Rightarrow \quad H_{k+1}z_k = H_k z_k + \frac{v_k (v_k^T z_k)}{v_k^T z_k} - \frac{(H_k z_k)(z_k^T H_k z_k)}{z_k^T H_k z_k} \quad (3.8)$$

Since $v_k^T z_k$ and $z_k^T H_k z_k$ are scalars, then

$H_{k+1}z_k = H_k z_k + v_k - H_k z_k$ So we have $H_{k+1}z_k = v_k$ ∎

Theorem3.2: Positive definiteness of H matrix

If $H_k$is positive definite, then the matrix H_{k+1} generated by the New5 method is also positive definite.

Proof: We first write the following quadratic form:

$$x^T H_{k+1} x = x^T H_k x + x^T \frac{v_k v_k^T}{v_k z_k^T} x - x^T \frac{(H_k z_k)(H_k z_k)^T}{z_k^T H_k z_k} x \quad (3.9)$$

$$\Rightarrow x^T H_{k+1} x = x^T H_k x + \frac{(x^T v_k)^2}{v_k z_k^T} - \frac{(x^T H_k z_k)^2}{z_k^T H_k z_k} \quad (3.10)$$

Define $a = H_k^{\frac{1}{2}} x$

$b = H_k^{\frac{1}{2}} z_k$

Where $H_K = H_k^{\frac{1}{2}} H_k^{\frac{1}{2}}$

Note that because H_K >0, Using the Definitions of a and b, we obtain

$$x^T H_k x = x H_k^{\frac{1}{2}} H_k^{\frac{1}{2}} x = a^T a$$

$$x^T H_k z_k = x H_k^{\frac{1}{2}} H_k^{\frac{1}{2}} z_k = a^T b$$

And

$$z_k^T H_k z_k = z_k H_k^{\frac{1}{2}} H_k^{\frac{1}{2}} z_k = b^T b$$

Hence

$$x^T H_{k+1} x = a^T a + \frac{(x^T v_k)^2}{v_k z_k^T} - \frac{(a^T b)^2}{b^T b}$$

$$x^T H_{k+1} x = \frac{\|a\|^2 - (a^T b)^2}{\|b\|^2} + \frac{(x^T v_k)^2}{v_k^T z_k}$$

We also have

$$v_k{}^T z_k = v_k{}^T \left(y_k + \alpha_k \left(\frac{\theta}{v_k^T y_k} g_k \right) \right) = v_k{}^T y_k + \alpha_k \frac{\theta}{v_k^T y_k} v_k{}^T g_k$$

$$v_k{}^T z_k = v_k{}^T y_k + \frac{v_k^T y_k}{y_k^T y_k} \frac{\theta}{v_k^T y_k} v_k{}^T g_k$$

Since $v_k{}^T y_k$ is scalar then

$$v_k{}^T z_k = v_k{}^T y_k + \frac{\theta}{y_k^T y_k} v_k{}^T g_k$$

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

$$\Rightarrow \; v_k{}^T y_k = \alpha_k d_k{}^T g_{k+1} + \alpha_k g_k{}^T H_k g_k$$

By Wolfe condition

$$v_k{}^T y_k \geq c_2 \alpha_k g_k{}^T d_k + \alpha_k g_k{}^T H_k g_k = -c_2 \alpha_k g_k{}^T H_k g_k + \alpha_k g_k{}^T H_k g_k$$

$$\Rightarrow \; v_k{}^T y_k \geq (1 - c_2) \alpha_k g_k{}^T H_k g_k$$

Since $0 < c_2 < 1$, $\alpha_k > 0$ and H_k is positive then

$$(1 - c_2) \alpha_k g_k{}^T H_k g_k > 0$$

$$\Rightarrow \; v_k{}^T y_k > 0 \qquad (3.11)$$

Now we will prove that $\frac{\theta}{y_k^T y_k} > 0$

$$\theta = 2(f(x_{k+1}) - f(x_k) - g_k{}^T v_k)$$

$$\theta = 2(f(x_{k+1}) - f(x_k) - g_k{}^T v_k)$$

$$\Rightarrow \; \theta = -2(f(x_k) - f(x_{k+1}) + g_k{}^T v_k)$$

By Wolfe condition

$$f(x_{k+1}) - f(x_k) \leq c_1 \alpha_k g_k{}^T d_k \Rightarrow f(x_k) - f(x_{k+1}) \geq -c_1 \alpha_k g_k{}^T d_k$$

So we have

$$\theta \geq -2(-c_1 \alpha_k g_k{}^T d_k + \alpha_k g_k{}^T d_k)$$

$$\Rightarrow \; \theta \geq -2(c_1 \alpha_k g_k{}^T H_k g_k - \alpha_k g_k{}^T H_k g_k)$$

$\Rightarrow \ \theta \geq 2(-c_1 \alpha_k g_k{}^T H_k g_k + \alpha_k g_k{}^T H_k g_k)$

$\Rightarrow \ \theta \geq 2\alpha_k(1 - c_1) g_k{}^T H_k g_k$

Since $0 < c_2 < 1$, $\alpha_k > 0$ and H_k is positive then

$2\alpha_k(1 - c_1) g_k{}^T H_k g_k > 0 \ \Rightarrow \ \theta > 0$

Hence $y_k^T y = \|y_k\|^2 > 0$

Therefore

$$\frac{\theta}{y_k^T y_k} > 0 \qquad (3.12)$$

Now we will prove that $v_k{}^T g_k > 0$

This impels that

$v_k{}^T g_k = v_k{}^T g_{k+1} - v_k{}^T y_k = \alpha_k d_k{}^T g_{k+1} - v^T y_k$

We know that $v^T y_k$ is positive and $\alpha_k g_k{}^T d_k = -\alpha_k g_k{}^T H_k g_k$ is negative

So, we can write $v^T y_k > \alpha_k g_k{}^T d_k$

Depending on the above inequality and by Wolfe condition, we obtain

$v_k{}^T g_k \geq c_2 \alpha_k g_k{}^T d_k - \alpha_k g_k{}^T d_k$

$\Rightarrow \ v_k{}^T g_k \geq -c_2 \alpha_k g_k{}^T H_k g_k + \alpha_k g_k{}^T H_k g_k$

$\Rightarrow \ v_k{}^T g_k \geq (1 - c_2) \alpha_k g_k{}^T H_k g_k$

Since $0 < c_2 < 1$, $\alpha_k > 0$ and H_k is positive then

$(1 - c_2) \alpha_k g_k{}^T H_k g_k > 0$

$$\Rightarrow v_k{}^T g_k > 0 \qquad (3.13)$$

From (3.11), (3.12) and (3.13), we get

$v_k{}^T z_k > 0$

These yields

$$x^T H_{k+1} x = \frac{\|a\|^2 - (a^T b)^2}{\|b\|^2} + \frac{(x^T v_k)^2}{{v_k}^T z_k}$$

Both terms on the right-hand side of the above equation are nonnegative-the

First term is nonnegative because of the Cauchy-Schwarz inequality, and the

Second term is nonnegative because ${v_k}^T z_k > 0$

Therefore, to show that $x^T H_{k+1} x > 0$ *for* $x \neq 0$, we only need to demonstrate that these terms do not both vanish simultaneously. The first term vanishes only if a and b are proportional, that is, if $a = \beta b$ for some scalar β.Thus, to complete the proof it is enough to show that if $a = \beta b$

Then $\frac{(x^T v_k)^2}{{v_k}^T z_k} > 0$. Indeed first observe that

$$H_k^{\frac{1}{2}} x = a = \beta b = \beta H_k^{\frac{1}{2}} z_k = H_k^{\frac{1}{2}} (\beta z_k)$$

Hence $x = \beta z_k$

Using the expression for x above and the expression

$$x^T H_{k+1} x = \frac{\|a\|^2 - (a^T b)^2}{\|b\|^2} + \frac{({v_k}^T \beta z_k)^2}{{v_k}^T z_k}$$

$$x^T H_{k+1} x = \frac{\|a\|^2 - (a^T b)^2}{\|b\|^2} + \frac{\beta^2 ({v_k}^T z_k)^2}{{v_k}^T z_k}$$

$$x^T H_{k+1} x = \frac{\|a\|^2 - (a^T b)^2}{\|b\|^2} + \beta^2 {v_k}^T z_k$$

$$\beta^2 {v_k}^T z_k > 0$$

Thus for all $x \neq 0$

$$x^T H_{k+1} x > 0$$

Which completes the proof.

■

4. NUMERICAL RESULTS AND DISCUSSIONS

This section is devoted to test the implement of the new method. We compare the new update of DFP (New5) and standard $DFP^{H/S}$. The comparative tests involve well known nonlinear problems (classical test function) with different function $4 \leq N \leq 5000$. all programs are written in FORTRAN 95 language and for all cases the stopping condition $\|g_{k+1}\|_\infty \leq 1 \times 10^{-5}$
and restart using Powell condition $|{g_k}^T g_{k+1}| > 0.2\|g_{k+1}\|$. The line search routine was a cubic interpolation which uses function and gradient values. The results given in tables (4.1) and (4.2) specifically quote the number of iteration NOI and the number of function NOF. Experimental results in tables (4.1) and (4.2) confirm that the new algorithm (New5) is superior to classical method of $DFP^{H/S}$ with respect to the number of iteration NOI and the number of function NOF.

Comparative Performance of Standard $DFP^{H/S}$ and New5

Table (4.1)

No. of test	Test function	N	Standard $DFP^{H/S}$		New5	
			NOI	NOF	NOI	NOF
1	Rosen	4	34	95	31	86
		100	34	94	32	88
		500	34	95	34	95
		1000	35	97	34	93
		5000	36	100	39	104
2	Cubic	4	14	40	14	40
		100	17	41	16	47
		500	16	46	17	49
		1000	17	49	16	47
		5000	16	46	16	46
3	Powell	4	35	96	33	90
		100	38	111	36	105
		500	38	111	37	107
		1000	39	113	37	107
		5000	39	113	37	107
4	Wolfe	4	11	24	9	21
		100	44	89	44	89
		500	47	95	47	95
		1000	50	101	50	101
		5000	107	215	107	215
5	Shallow	4	8	21	8	21
		100	9	24	8	21
		500	9	24	8	21
		1000	9	24	8	21
		5000	9	24	8	21

Table (4.2)

No. of test	Test function	N	Standard $DFP^{H/S}$		New5	
			NOI	NOF	NOI	NOF
6	Wood	4	20	50	20	50
		100	23	57	22	54
		500	23	57	22	54
		1000	23	57	22	54
		5000	23	57	22	54
7	Non-diagonal	4	27	74	26	72
		100	53	128	46	112
		500	49	118	49	118
		1000	49	119	49	119
		5000	49	119	49	119
8	OSP	4	8	45	8	45
		100	56	227	48	185
		500	104	334	104	338
		1000	125	389	124	388
		5000	297	1028	297	1028
9	Mile	4	66	300	29	124
		100	86	411	39	183
		500	94	454	49	254
		1000	100	469	57	310
		5000	112	575	65	369
10	Beal	4	11	29	9	22
		100	12	31	10	26
		500	12	31	10	26
		1000	12	31	10	26
		5000	12	31	10	26
11	G-centeral	4	34	237	23	149
		100	39	295	29	228
		500	49	423	30	241
		1000	54	479	30	241
		5000	71	682	31	255
Total			**2438**	**9225**	**2065**	**7107**

Comparing the Rate of Improvement between the new algorithm (New5) and the Standard Algorithm of $DFP^{H/S}$

Table (4.3) shows the rate of improvement in the new algorithm (New5) with the standard algorithm (DFP), The numerical results of the new algorithm is better than the standard algorithm, As we notice that (NOI), (NOF) of the standard algorithm are about 100%, That means the new algorithm has improvement on standard algorithm prorate (15.2994%) in (NOI) and prorate (22.9593%) in (NOF), In general the new algorithm (New5) has been improved prorate (19.1294%) compared with standard algorithm (DFP).

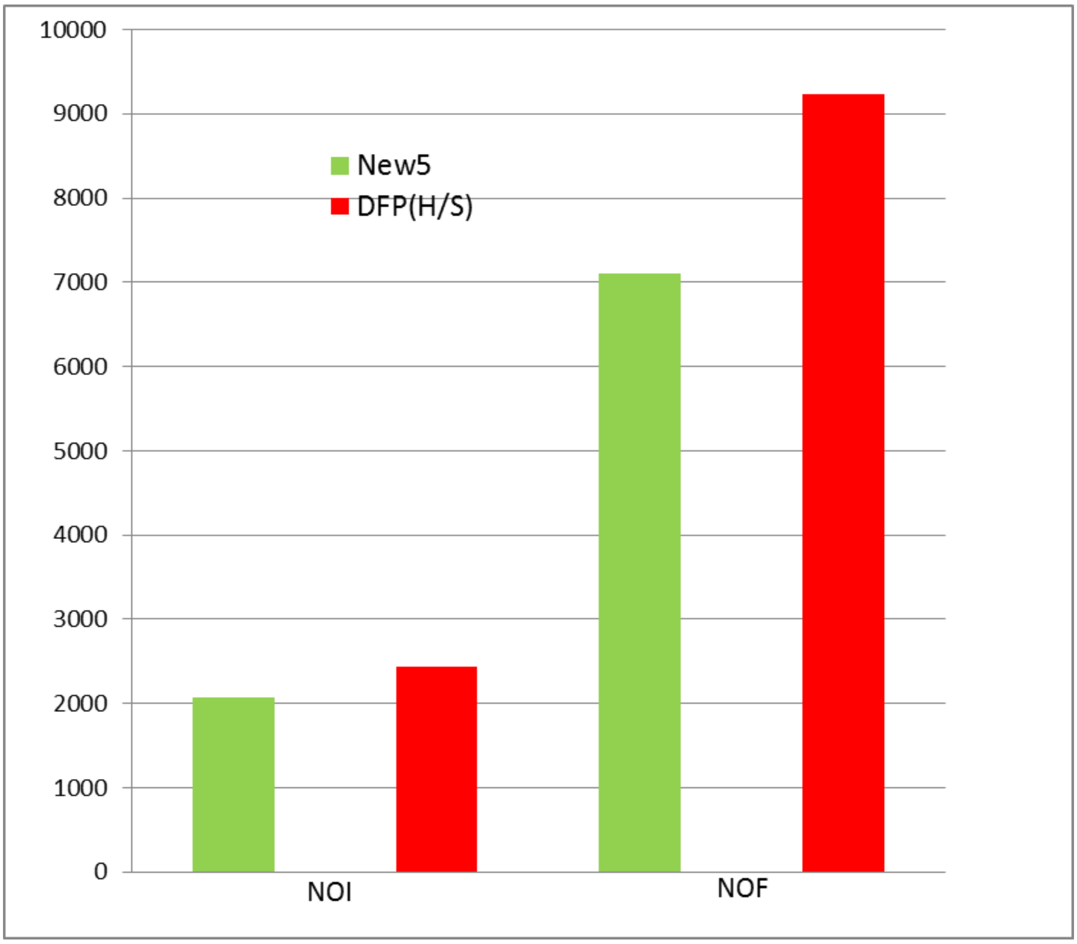

Figure (4.1) shows the comparison between new algorithm (New5) and the standard algorithm (DFP) according to the total number of iterations (NOI) and the total number of functions (NOF).

5. CONCLUSION

In this paper, we proposed a new algorithm of DFP method. The experimental results in tables (4.1) and (4.2) confirm that the new algorithm (New5) is superior to standard method of $DFP^{H/S}$ with respect to the number of iteration NOI and the number of function NOF and in the future can be applied we proposed on Rank one method and BFGS method with some amendments

References

[1]. A. Abashar, M. Mamat, M. Rivaie and I. Mohd, Global Convergence Properties of a New Class of Conjugate Gradient Method for Unconstrained Optimization, International Journal of Mathematical Analysis, 8 (2014), 3307-3319.

[2]. AL - Bayati, A.Y. and AL-Assady, N.H.,Conjugate gradient method,Technical Research , 1 (1986), school of computer studies, Leeds University.

[3]. E. Polak and G. Ribiere, Note sur la convergence de directionsconjugees, Rev. Francaise Informat Recherche Operationelle. 3(16) (1969), pp. 35-43.

[4]. J.E. Dennis and J.J. Mor6, "Quasi-Newton methods, motivation and theory," SIAM Review 19
a. (1977) 46-89.

[5]. J.E. Dennis and R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear
a. Equations (Prentice-Hall, Englewood Cliffs, NJ, 1983).

[6]. M.R. Hestenes and E. Steifel, Method of conjugate gradient for solving linear equations, J. Res. Nat. Bur. Stand. 49 (1952), pp. 409-436.

[7]. M. Rivaie, A. Abashar, M. Mamat and I. Mohd, The Convergence Properties of a New Type of Conjugate Gradient Methods, Applied Mathematical Sciences, 8 (2014),33-44.

[8]. R. Fletcher, and C. Reeves, Function minimization by conjugate gradients, Comput. J. 7(1964), pp. 149-154.

[9]. R. Fletcher, Practical method of optimization, vol 1, unconstrained
a. optimization, John Wiley & Sons, New York, 1987.

[10]. R. Fletcher, Practical Methods of Optimization: Unconstrained Optimization (Wiley, Chichester, 1980).

[11]. P.E. Gill, W. Murray and M.H. Wright, Practical Optimization (Academic Press, New York, 1981).

[12]. W.W. Hager, H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, SIAM Journal on Optimization,16 (2005) 170–192

[13]. Y. H. Dai and Y. Yuan, A note on the nonlinear conjugate gradient method, J. Compt. Appl. Math., 18(2002), 575-582.

[14]. Y. Liu and C. Storey, Efficient generalized conjugate gradient algorithms part 1: Theory, J. Comput. Appl. Math. 69 (1992), pp. 129-137.