Partial Davidon, Fletcher and Powell (DFP) of quasi newton method for unconstrained optimization

Basheer M. Salih[1] , Khalil K. Abbo[2] , Zeyad M. Abdullah[3]
[1] College of Education , University of Mosul , Mosul , Iraq
[2] College of Computers Sciences and Math. , University of Mosul , Mosul , Iraq
[3] College of Computers Sciences and Math. , University of Tikrit , Tikrit , Iraq

Abstract:
The nonlinear Quasi-newton methods is widely used in unconstrained optimization. However, In this paper, we developing new quasi-Newton method for solving unconstrained optimization problems. We consider once quasi-Newton which is (DFP) update formula, namely, Partial DFP. Most of quasi-Newton methods don't always generate a descent search directions, so the descent or sufficient descent condition is usually assumed in the analysis and implementations . Descent property for the suggested method is proved. Finally, the numerical results show that the new method is also very efficient for general unconstrained optimizations.

Key words: Unconstrained optimization; Davidon, Fletcher and Powell ; global convergence.

1.Introduction:

we consider the following unconstrained optimization problem

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

Where $f : R^n \rightarrow R$ is continuously differentiable function.

Quasi-Newton method is a well-known and useful method for solving unconstrained convex programming and the BFGS method is the most effective quasi-Newton type methods for solving unconstrained optimization problems from the computation point of view. For the current iterate $x_k \in R^n$ and symmetric positive definite matrix $B_k \in R^{n\times n}$, the next iterate is obtained by

$x_{k+1} = x_k + \alpha_k d_k$ (2)

where $\alpha_k > 0$ is a step-size obtained by a one-dimensional line search, and

$d_k = -B_k^{-1}\nabla f(x_k)$ (3)

Is a descent direction B_k^{-1} being available and approximating the inverse of the Hessian matrix of f at x_k . throughout this paper, we use $\| \ \|$ to denote Euclidean vector or matric norm and denote $\nabla f(x_k)$ by g_k.

Although the quasi-Newton type method is known to be remarkably robust in practice, one will not be able to establish truly convergence results for general nonlinear objective functions, that is, one cannot prove that the iterates generated by this method approach a stationary point of the problem from any starting point and any (suitable) initial Hessian approximation. Therefore, there has been an ever-expanding interest in quasi-Newton type methods since the first quasi-Newton method was suggested by Davidon (1959) and improved by Fletcher and Powell (1963) (hence the name "DFP" formula), and there is a vast literature on the problem of convergence properties of the quasi-Newton type method for solving problem (1).For details, see Refs[1], [2], [3] and references therein.

Algorithm 1.1 (A general quasi-Newton algorithm)

Step 1. Given $x_1 \in R^n$, $H_1 \in R^{n\times n}$, $0 \le \varepsilon < 1$ and $k = 1$

Step 2. If $\|g_k\| \le \varepsilon$, stop.

Step 3. Compute

$$d_k = -H_k g_k \quad (4)$$

Step 4. Find a step size $\alpha_k > 0$ by line search and set

$$x_{k+1} = x_k + \alpha_k d_k \quad (5)$$

Step 5. Update H_k into H_{k+1} such that the quasi-Newton equation

$H_{k+1} y_k = s_k$ holds.

Step 6. $k = k+1$ and go to Step 2. □

In the above algorithm, it is common to start the algorithm with $H_1 = I$, an identity matrix or set H_1 to be a finite-difference approximation to the inverse Hessian G_1^{-1}. If $H_1 = I$, the first iteration is just a steepest descent iteration. Sometimes, quasi-Newton method takes the form of Hessian approximation B_k. In this case, the Step 3 and Step 5 in Algorithm (1) have the following forms respectively.

Step 3*. Solve

$B_k d_k = -g_k$ for d_k

Step5*. Update B_k into B_{k+1} so that quasi-Newton equation $B_{k+1} s_k = y_k$ holds.

Since the metric matrices B_k are positive definite and always changed from iteration to iteration, the method is also called the variable metric method. In the following we consider some of the most popular formulas for updating inverse Hessian approximation H_k.

2. Rank-Two Update Quasi-Newton Methods.

Other quasi-Newton methods use rank-two updates, i.e. the difference between two consecutive approximations is a matrix of rank two. The reason for this approach is to preserve the symmetrical structure of the approximate to the invers Hessian matrix. It is thus not surprising that these methods are not meant as non-linear solvers but used to solve minimization problems. We will describe briefly the Davidon, Fletcher and Powell (DFP), for a good survey of these and other methods we refer to [4].

3. DFP update

DFP update is a rank-two update, i.e., H_{k+1} is formed by adding to H_k two symmetric matrices, each of rank one, and it has the following form

$$H_{k+1}^{DFP} = H_k + \frac{s_k y_k^T}{s_k^T s_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} \qquad (6)$$

The formula

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k) v_k^T}{v_k^T y_k} \qquad (7)$$

is the first quasi-Newton update proposed originally by
Davidon [5] and developed later by Fletcher and Powell [6]. Hence it is called DFP update.
DFP method has the following important properties:

A. For a quadratic function (under exact line search)

(a) DFP update has quadratic termination, i.e., $H_n = G^{-1}$.

(b) DFP update has hereditary property, i.e., $H_i y_j = s_j$, $j < i$.

(c) DFP method generates conjugate directions; when $H_1 = I$, the method generates conjugate gradients.

B. For a general function

(a) DFP update maintains positive definiteness.

(b) Each iteration requires $3n^2 + O(n)$ multiplications.

(c) DFP method is super-linearly convergent.

(d) For a strictly convex function, under exact line search, DFP method is globally convergent.

The fact that quasi-Newton update retains positive definiteness is of importance in efficiency, numerical stability and global convergence. If the Hessian $G(x^*)$ is positive definite, the stationary point x is a strong minimizer. Hence, we hope inverse Hessian approximation H_k is positive definite. In addition, if H_k is positive definite, the local quadratic model of f has a unique local minimizer, and the direction d_k is a descent direction. Usually, the update retaining positive definiteness means that if H_k is positive definite, then H_{k+1} is also positive definite. Such an update is also
called positive definite update. Next, we introduce some important theorems related to the DFP update for more details see [Op. theory and methods].

Theorem (1.1) (Positive Definiteness of DFP Update)

DFP update (6) retains positive definiteness if and only if $s_k^T y_k > 0$.

Corollary (1.1)

Each matrix H_k generated by DFP Algorithm is positive definite, and the directions $d_k = -H_k g_k$ are descent directions.

Theorem 1.2 (Quadratic Termination Theorem of DFP Method)

Let $f(x)$ be a quadratic function with positive definite Hessian G. Then,

if exact line search is used, the sequence $\{s_j\}$ generated from DFP method
satisfies hereditary property, conjugate property and quadratic termination,
that is, for $i = 0,1,\dots, m$ $\cdots$, where m ≤ n−1,

1. $H_{i+1} y_j = s_j$, $j = 0,1,\dots, i$; (hereditary property)
2. $s_i^T G\, s_j = 0$, $j = 0,1,\dots, i-1$; (conjugate direction property)
3. The method terminates at m+1≤ n steps. If m=n−1,then $H_n = G^{-1}$.

For the prof of the above theorems see [Op. theory and methods].
From the theorem (1.2) we see that DFP method is a conjugate direction
method. If the initial approximation $H_1 = I$, the method becomes a conjugate gradient method. DFP method is a seminal quasi-Newton method and has been widely used in many computer codes. It has played an important role in theoretical analysis and numerical computing. However, further studies indicate that DFP method is numerically unstable, and sometimes produces numerically singular Hessian approximations and requires $3n^2 + o(n)$ multiplication per iteration.
Powell in [7] analyzed the performance of the DFP algorithm on a very simple objective function of two variables. Through studying the eigenvalues of Hessian matrix, he found out that the DFP algorithm can be highly inefficient, could fail for general non-linear problems, it can stop at a saddle point. It is sensitive to inaccurate linear search.
In the next subsection will develop new method based on DFP and partial DFP algorithm quasi-Newton algorithms, to overcome these drawbacks.

4. Partial DFP method (PDFP).

Consider the DFP update formula defined in the equation (6) then the search direction based on equation (6) can be calculated as

$$d_{k+1} = -H_{k+1} g_{k+1}$$

Now

$$\begin{aligned} H_k g_{k+1} &= H_k g_{k+1} - H_k g_k + H_k g_k \\ &= H_k y_k - d_k \end{aligned}$$

$$H_k g_{k+1} = H_k y_k - \frac{1}{\alpha} s_k \qquad (8)$$

Therefore

$$d_{k+1} = -[H_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}] g_{k+1}$$

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

ISSN: 1813 – 1662 (Print)
E-ISSN: 2415 – 1726 (On Line)

$$d_{k+1} = -[H_k y_k - \frac{1}{\alpha} s_k + \frac{s_k^T g_{k+1}}{s_k^T y_k} s_k - \frac{y_k^T (H_k y_k - \frac{1}{\alpha} s_k)}{y_k^T H_k y_k} H_k y_k]$$

$$= -[\frac{s_k^T y_k}{y_k^T H_k y_k} H_k y_k + (\frac{s_k^T g_{k+1}}{s_k^T y_k} - \frac{1}{\alpha}) s_k]$$

$$d_{k+1} = -\frac{s_k^T y_k}{\alpha y_k^T H_k y_k} H_k y_k + \frac{s_k^T y_k - \alpha s_k^T g_{k+1}}{\alpha s_k^T y_k} s_k \quad (9)$$

We see from equation (9) that the computation of the search direction d_{k+1} not need any additional computations, since all terms in (9) have already been calculated during the iterations. Thus, the computing $H_{k+1} g_k$ is reduced at each iteration, hence we can use only $3n^2$ multiplications per iteration instead of $3n^2 + o(n)$.

At this point we summarize our algorithm (PDFP) in the following formal steps.

Algorithm (PDFP)

Step 1. Given an initial point x_1 and a positive definite matrix H_1, set $\varepsilon > 0$, $k = 1$.

Step2. If $\|g_k\| < \varepsilon$. Stop

Step 3. Compute search direction:

If $k = 1$ then

$$d_k = -H_k g_k$$

$$d_{k+1} = -\frac{s_k^T y_k}{\alpha y_k^T H_k y_k} H_k y_k + \frac{s_k^T y_k - \alpha s_k^T g_{k+1}}{\alpha s_k^T y_k} s_k$$

Step 4. Compute the step size α_k such that Wolfe conditions

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta\alpha_k g_k^T d_k, \quad (10)$$

$$|g(x_k + \alpha_k d_k)^T d_k| \le -\sigma g_k^T d_k, \quad (11)$$

and

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta\alpha_k g_k^T d_k, \quad (12)$$

$$\sigma g_k^T d_k \le g(x_k + \alpha_k d_k)^T d_k \le 0, \quad (13)$$

with $0 < \delta < \frac{1}{2}$, $\delta < \sigma < 1$.

satisfied.

Step 5. Compute new iterative point

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

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

Step 6. set $k = k + 1$. Go to step 2.

The following theorem indicates that, in the inexact case, the search direction d_k satisfies descent property:

$g_k^T d_k < 0$.

Theorem (4.1):

consider the algorithm PDFP, and assume that $\frac{s_k^T y_k}{\alpha^2 y_k^T H_k y_k} \le \frac{(s_k^T g_{k+1})^2}{s_k^T y_k}$, then the search direction generated by PDFP algorithm are descent directions.

Proof:

The proof is by induction for $k = 1$

$$g_1^T d_1 = -g_1^T H_1 g_1 = -\|g_1\|^2 < 0$$

Let $g_k^T d_k < 0$, note that, by definition of H_k we have $y_k^T H_k y_k > 0$ and by second Wolfe condition $s_k^T y_k \ge 0$

$$g_{k+1}^T d_{k+1} = -\frac{s_k^T y_k}{\alpha y_k^T H_k y_k} y_k^T H_k g_{k+1} + \frac{s_k^T y_k - \alpha g_{k+1}^T s_k}{\alpha s_k^T y_k} s_k^T g_{k+1}$$

$$= -\frac{s_k^T y_k}{\alpha} + \frac{s_k^T y_k}{\alpha^2 y_k^T H_k y_k} + \frac{s_k^T g_{k+1}}{\alpha} - \frac{(s_k^T g_{k+1}{}^2)}{s_k^T y_k}$$

Since

$$y_k^T s_k = g_{k+1}^T s_k - g_k^T s_k \ge g_{k+1}^T s_k$$

Hence

$$d_{k+1}^T g_{k+1} \le -\frac{s_k^T y_k}{\alpha} + \frac{s_k^T y_k}{\alpha^2 y_k^T H_k y_k} + \frac{g_{k+1}^T s_k}{\alpha} - \frac{(s_k^T g_{k+1}{}^2)}{s_k^T y_k},$$

$$= \frac{s_k^T y_k}{\alpha^2 y_k^T H_k y_k} - \frac{(g_{k+1}^T g_{k+1}{}^2)}{s_k^T y_k} < 0$$

The proof is complete.

Theorem (4.2)

The search directions generated by the equation (9) are conjugate directions.

Proof:

Consider the search direction given in equation (9):

$$d_{k+1} = -\frac{s_k^T y_k}{\alpha y_k^T H_k y_k} H_k y_k + \frac{s_k^T y_k - \alpha s_k^T g_{k+1}}{\alpha s_k^T y_k} s_k$$

Multiply both sides by y_k^T to get

$$y_k^T d_{k+1} = -\frac{s_k^T y_k}{\alpha y_k^T H_k y_k} y_k^T H_k y_k + \frac{(s_k^T y_k) - \alpha s_k^T g_{k+1}}{\alpha s_k^T y_k} s_k^T y_k$$

$$= -s_k^T g_{k+1}$$

This shows that the search directions generated by PDFP method are conjugate for all k with $t = 1$.

5. Convergence analysis of the PDFP method:

In order to establish the global convergence of PDFP method, we make the objective function:

Assumption (A)

(A1): The level set $\ell = \{x \in R^n : f(x) \le f(x_1)\}$ at the initial point x_1 is bounded namely, then exists positive constants μ_1 and μ_2 such that

$$\|x\| \le \mu_1 \text{ And } \|s\| \le \mu_2, \quad \forall x \in \ell \quad (14)$$

(A2): In some neighborhood N of ℓ, f is continuously differentiable and its gradient is Lipschitz continuous i.e there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \le L\|x - y\|, \quad \forall x, y \in N$$

$$\|g(x)\| \le \Gamma_1 \text{ And } \|y\| \le \Gamma_2, \quad \forall x \in \ell \quad (15)$$

Note that, since H_k is symmetric and positive then the following is true

$$m\|y\| \le y_k^T H_k y_k \le M\|y\| \quad (16)$$

ISSN: 1813 – 1662 (Print)
E-ISSN: 2415 – 1726 (On Line)

Where m and M are the smallest and largest eigenvalues.

The following is a useful lemma for proving the global convergence property of iterative methods (see lemma 1.1 of [8]).

Lemma (5.1)

Suppose that Assumption A is satisfied. Consider any iterative method of the form

$$x_{k+1} = x_k + \alpha_k d_k \qquad (17)$$

Where d_k and α_k satisfy the descent or (sufficient descent) condition and the Wolfe conditions (10), (11), (12) and (13) respectively. If

$$\sum_{k=1}^{\infty} \frac{1}{\|d_k\|^2} = +\infty$$

Then the following holds

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

Theorem (5.1)

Let $\{x_k\}$ be generated by PDFP method and Assumption A hold. Then we have

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

Proof:

The proof is by contradiction. Assume that

$$\sum_{k=1}^{\infty} \frac{1}{\|d_k\|^2} < \infty$$

Consider the search direction generate by PDFP method

$$d_{k+1} = -\frac{s_k^T y_k}{\alpha y_k^T H_k y_k} H_k y_k + \frac{s_k^T y_k - \alpha s_k^T g_{k+1}}{\alpha s_k^T y_k} s_k$$

Note that by assumption A we have

$$\frac{1}{y_k^T H_k y_k} \le \frac{1}{m\|y_k\|}, \|s_k\| \le \mu_2, \|y_k\| \le \Gamma_2, \|H_k\| \le \lambda_m$$

and $\|y_k\| \le \|s_k\|$, using the definitions we have

$$\|d_{k+1}\|^2 = \left\| \frac{-s_k^T y_k}{\alpha y_k^T H_k y_k} H_k y_k + \frac{s_k}{\alpha} - \frac{s_k^T g_{k+1}}{s_k^T y_k} s_k \right\|^2$$

$$\le \left| \frac{s_k^T y_k}{\alpha m \|y\|^2} \right| \|1 + y\|^2 + \frac{1}{\alpha}\|s_k\|^2 + \left| \frac{s_k^T g_{k+1}}{s_k^T y_k} \right| \|s_k\|$$

$$\le \frac{s_k^T y_k}{\alpha m} M + \frac{\mu_2}{\alpha} + \mu_2$$

$$\le \frac{2ML\|s_k\|^2}{\alpha m} \mu_2$$

$$= \frac{2ML\mu_2^2}{\alpha m} \to \sum_{k=0}^{\infty} d_{k+1} \le \sum_{k=0}^{\infty} \gamma = \infty$$

Hence,

$$\sum \frac{1}{\|d_{k+1}\|} > \sum \frac{1}{\gamma} > 0$$

6. Numerical results and comparisons methods

In this subsection we present the computation performance of a FORTRAN implementation of the DFP and PDFP algorithms on a set of unconstrained optimization test problems.

In this subsection we report numerical experiments of the proposed methods (partial DFP methods) and classical DFP Quasi-Newton methods. Our experiments are performed for 52 non-linear unconstrained optimization problems (functions) in the CUTEr library [9]. Each test problem is made ten experiments with the number of variable 100,200,…, 1000, respectively.

In table (1-1) method examined in our experiments

n.	Method name	Description
1	DFP	Davidon- Fletcher – Powell QN method
2	PDFP	Partial Davidon- Fletcher – Powell QN method

In the line search Procedure, the step-size α_k is chosen so that the Wolfe conditions (10), (11), (12) and (13) are satisfied with $\rho = 0.1$ and $\sigma = 0.9$. The stopping criterion was $\|g_k\| \le 10^{-6}$.

In this work, we used two codes of the codes are programmer by visual Fortran. The first code was developed by Andrie [10] and improved by Donal and more. The second code developed by Andrei [11] which uses CG algorithms, we improved this code and adapted by using QN algorithms. We developed the third code wing Matlab for results and graphic comparisons.

Table (1.1) gives the total number of iterations (toit), the total number of function evaluations (tfn) and total time (totime) for solving 520 test problems.

Table (1-1)

n.	Name of Algorithm	Toit	Tfn	Totime
1	DFP	89665	860578	150730
2	PDFP	64629	209701	102052

In this Figures (1- 3) we adopt the performance profiles by Donald and More [12] to compare the performance based on the number of iterations and CPU time. That is, for each method, we plot the fraction ρ of problems for which the method is within a factor tao of the best result. The left side of the figure gives the percentage of the test problems for whicg a method is the best result, the right side gives the percentage of the test problems that are successfully solved by each of the methods. The top curve is the method that solved the most problems in a result that is within a factor tao of the best results.

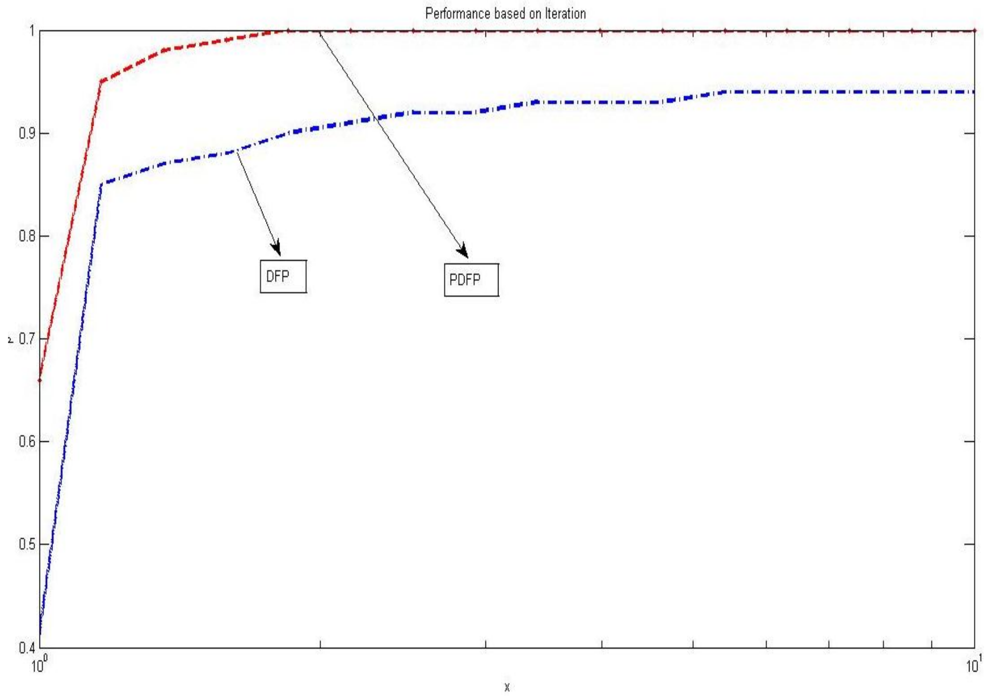

Figure (1-1) Comparison between DFP and PDFP based on iteration

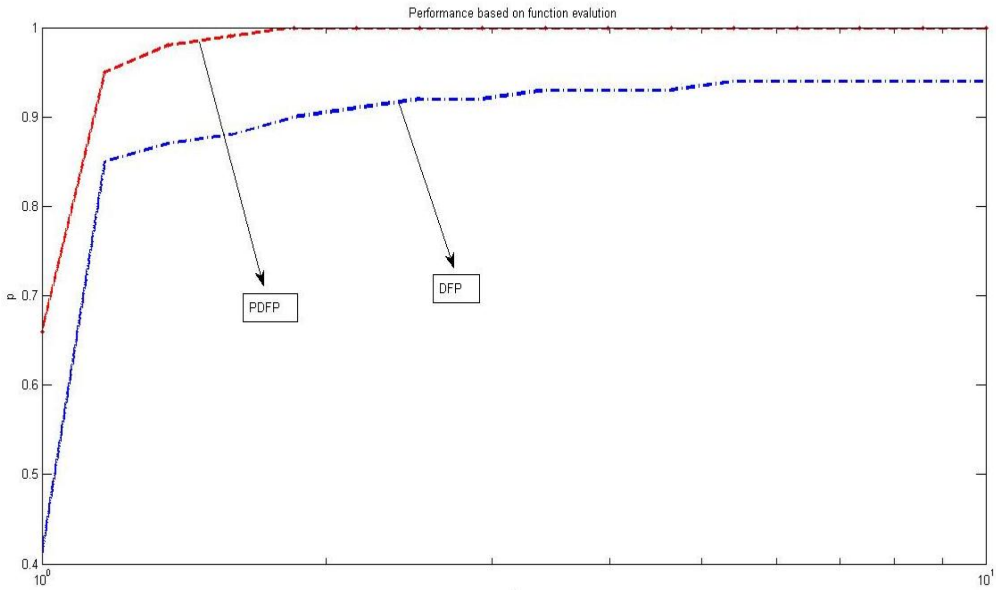

Figure (1-2) Comparison between DFP and PDFP based on evaluation function

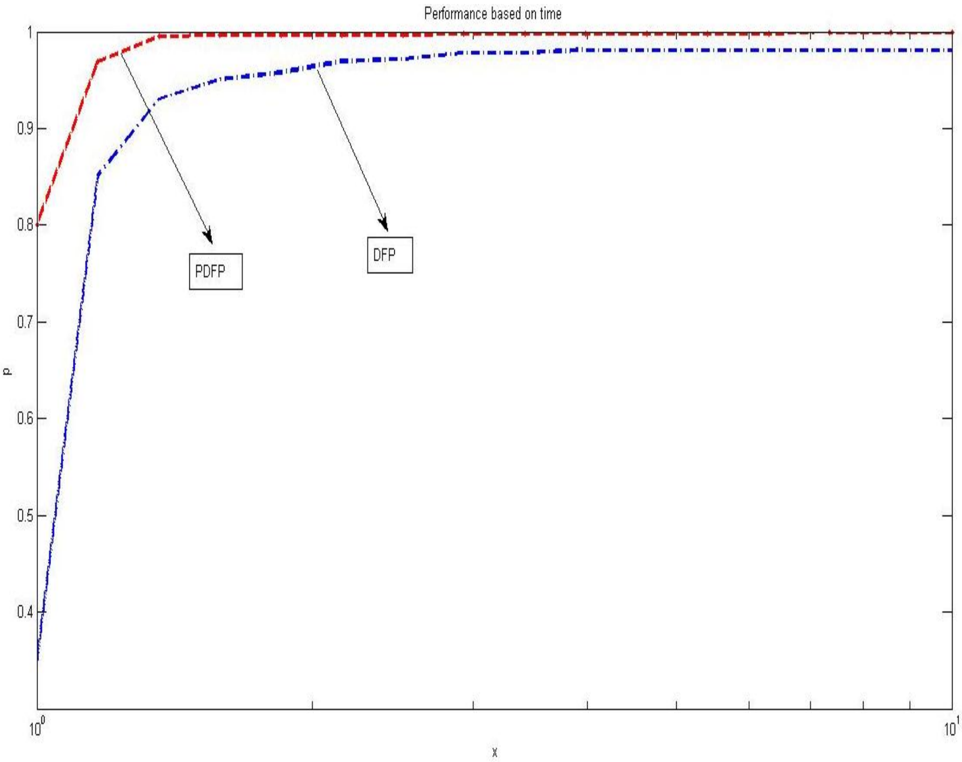

Figure (1-3) Comparison between DFP and PDFP based on Time

7. Conclusion:

In this study a new algorithm presented a new form DFP QN method developed for solving large-scale unconstrained optimization problems, in which the PDFP update based on the modified QN equation have applied. An important feature of the proposed method is that it preserves positive definiteness of the updates. The presented method owns with descent property and global convergence with the Wolfe line search. Numerical results showed that the proposed method is encouraging comparing with the methods DFP and PDFP.

References:

[1] Al-Baali, M. (1985). Descent property and global convergence of the Fletcher-Reeves method with inexact line search, IMA J. Num. Anal., 5, 121-124.

[2] W.C. Davidon, Variable-metric Method for Minimization, AEC Report ANL5990, Argonne National Laboratory, 1959.

[3] Pearson D.(1969):'Variable Metric methods for minimization'. Computer Journal vol.12.

[4] J.Z. Zhang, C.X. Xu, Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations, J. Comput. Appl. Math. 137 (2001) 269–278.

[5] Davidon, W. C. (1959). Variable metric method for minimization, AEC Res. and Report ANL-5990 (revised).

[6] Fletcher, R. and Powell, M. J. D. (1963). A rapidly convergent descent method for minimization, Computer. J., 6, 163-168.

[7] Powell, M. J. D. and Yuan, Y. (1986). A recursive quadratic programming algorithm that uses differentiable penalty functions, Math. Prog., 35, 265-278.

[8] Sugiki K. Narushima Y. and Yube H. (2012)' Globally convergence three- term conjugate gradient methods that uses secant conditions and generate descent search direction, for unconstrained optimization. J. of optimization theory and Applications 153.

[9] Bongartz, I., A.R. Conn, N.I.M. Gould and Ph. L. Toint, 1995. CUTE: Constrained and unconstrained testing environment: ACM Trans. Math. Software, 21: 123-160. http://portal.acm.org/citation.cfm?doid=200979.201043.

[10] Andrei, N., Scaled conjugate gradient algorithms for unconstrained optimization. Computational Optimization and Applications 38, 401-416 (2007).

[11] Andrei, N., A hybrid conjugate gradient algorithm with modified secant condition for unconstrained optimization. ICI Technical Report, February 6, 2008.

[12] Dolan, E.D. and J.J. Moré, 2002. Benchmarking optimization software with performance profiles. Math. Program., 91: 201-203, DOI: 10.1007/ s101070100263.

ISSN: 1813 – 1662 (Print)
E-ISSN: 2415 – 1726 (On Line)

صيغة (DFP) الجزئية لطريقة شبه نيوتن في الامثلية الغير مقيدة

بشير محمد صالح خلف[1] ، خليل خضر عبو[2] ، زياد محمد عبد الله[3]
[1] قسم الرياضيات ، كلية التربية للعلوم الصرفة ، جامعة الموصل ، الموصل ، العراق
[2] قسم الرياضيات ، كلية التربية الاساسية ، جامعة تلعفر ، تلعفر ، العراق
[3] قسم الرياضيات ، كلية علوم الحاسوب والرياضيات ، جامعة الموصل ، الموصل ، العراق

الملخص:

تعتبر طرق شبيهة نيوتن من اكثر الطرق انتشارا لحل مسائل الامثلية غير المقيدة. في هذا البحث تم تطوير طريقة جديدة من طرق شبيهة نيوتن (DFP) واسميناها بطريقة (PDFP) الجزئية ، ولان اغلب طرق شبيهة نيوتن لاتولد دائما شرط الانحدار ولذلك فان خاصية الانحدار والانحدار الكافي تفرض عند تحليل وتمثيل هذه الخوارزميات. تم اثبات خاصية الانحدار في الطريقة المقترحة. والنتائج العددية تبين ان الطريقة المقترحة هي ايضا فعالة جداً وممتازة بالمقارنة مع طريقة (DFP) الاصلية.