Research Journal of Applied Sciences 12 (1): 31-36, 2017
ISSN: 1815-932X


# A New Search Direction for Broyden's Family Method with Coefficient of Conjugate Gradient in Solving Unconstrained Optimization Problems

[1]Mohd Asrul Hery Ibrahim, [2]Zahratul Amani Zakaria, [2]Mustafa Mamat,
[3]Ummie Khalthum Mohd Yusof and [1]Azfi Zaidi Mohammad Sofi
[1]Faculty of Entrepreneurship and Business, University Malaysia Kelantan,
Kampus Kota, Kuala Lumpur, Malaysia
[2]Faculty of Informatics and Computing, University Sultan Zainal Abidin,
Kampus Tembila, Besut, Malaysia
[3]Faculty of Computer Science and Technology, University Malaya, Malaya, Malaysia

**Abstract:** In this study, we present a new search direction known as the CG-Broyden method which uses the search direction of the conjugate gradient method approach in the quasi-Newton methods. The new algorithm is compared with the quasi-Newton methods in terms of the number of iterations and CPU-time. The Broyden's family method is used as an updating formula for the approximation of the Hessian for both methods. Our numerical analysis provides strong evidence that our CG-Broyden method is more efficient than the ordinary Broyden method. Besides, we also prove that the new algorithm is globally convergent.

**Key words:** Broyden method, CG-Broyden method, CPU time, conjugate gradient method, globally convergent

## INTRODUCTION

Quasi-Newton methods are well-known methods in solving the unconstrained optimization method which uses the updating formulas for approximation of the Hessian. These methods were introduced by Davidon in 1959 and later popularised by Fletcher and Powell in 1963 but the Davidon-Fletcher-Davidon (DFP) method is rarely used nowadays. However, in 1970 Broyden, Fletcher, Goldfarb and Shanno developed the idea of a new updating formula, known as BFGS which has become widely used and recently the subject of many modifications. Then, Broyden (1970) proposed a family of quasi-Newton methods in 1970. In general, the unconstrained optimization problems are described as follows:

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

Where:

$R^n$ = An n-dimensional Euclidean space
$f: R^n \to R$ = Continuously differentiable

The gradient and Hessian for Eq. 1 are denoted as g and G, respectively. In order to display the updated formula of Broyden's family, the step-vectors and are defined as:

$$\begin{aligned} s_k &\overset{\text{def}}{=} x_{k+1} - x_k \\ y_k &\overset{\text{def}}{=} g(x_{k+1}) - g(x_k) = g_{k+1} - g_k \end{aligned} \tag{2}$$

The Broyden's algorithm for unconstrained optimization problem uses the matrices $B_i$ which is updated by the equation:

$$B_{k+1} = B_k - \left(\frac{B_k s_k s_k^T B_k}{s_k^T B_k y_k}\right) + \frac{y_k y_k^T}{s_k^T y_k} + \phi_k \left(s_k^T B_k s_k\right) v_k v_k^T \tag{3}$$

where, $\phi$ is a scalar and:

$$v_k = \left[\frac{y_k}{s_k^T y_k} - \frac{B_k s_k}{s_k^T B_k s_k}\right]$$

The choice of the parameter $\phi$ is important, since it can greatly affect the perfomance of the method (Xu, 2003). When in Eq. 3, we obtain the DFP algorithm and $\phi_1 = 0$ we get the BFGS algorithm. But Byrd and Nocedal (1989) extended his result to $\phi \in (0, 1)$ Based on (Chong and Zak, 2001), the Broyden's algorithm is one of the most efficient algorithm for solving the unconstrained optimization problem. It's also well known that the matrix $B_{k+1}$ is generated by Eq. 3 to satisfy the secant equation

**Corresponding Author:** Mohd Asrul Hery Ibrahim, Faculty of Entrepreneurship and Business, University Malaysia Kelantan, Kampus Kota, Kuala Lumpur, Malaysia

$$B_{k+1}s_k = y_k \quad (4)$$

which may be regarded as an approximate version of the relation. Note that it is only possible to fulfil the secant equation if:

$$s_k^T y_k > 0 \quad (5)$$

which is known as the curvature condition. Realising the possible non-convergence for general objective functions, some researchers have considered modifying quasi-Newton methods to enhance the convergence. For example, Li and Fukushima (2001) modify the BFGS method by skipping the update when certain conditions are not satisfied and prove the global convergence of the resulted BFGS method with a "cautious update" (which is called the CBFGS method). However, their numerical tests show that the CBFGS method does not perform better than the ordinary BFGS method. Then, Mamat *et al.* (2009) and Ibrahim *et al.* (2010) proposed a new search direction for quasi-Newton methods in solving unconstrained optimization problems. Generally, the search direction focused on the hybridization of quasi-Newton methods with the steepest descent method. The search direction proposed by Mamat *et al.* (2009) is $d_k = -\eta B_k^{-1} g_k - \delta g_k$ where $\eta > 0$ and $\delta > 0$. They realised that the hybrid method is more effective compared with the ordinary BFGS in terms of computational cost. Hence, the delicate relationships between the conjugate gradient and the BFGS method have been explored in the past.

In this study, motivated by the idea of conjugate gradient methods we propose a line search algorithm for solving (1) where the search direction of the quasi-Newton methods will be modified using the search direction of the conjugate gradient method approach. We prove that our algorithm with the Wolfe line search is globally convergent for general objective function. Then we test the new approach on standard test problems, comparing the numerical results with the results of applying the quasi-Newton methods to the same set of test problems.

## MATERIALS AND METHODS

**Iteration method:** The iterative method is used to solve unconstrained optimization problems in order to get the minimal value of the function where the gradient is 0. Hence, the iterative formula for the quasi-Newton methods will be defined as:

$$x_{k+1} = x_k + \alpha_k d_k \quad (6)$$

where the $a_k$ and $d_k$ denote the step size and the search direction, respectively. The step size must always have a positive value such that f(x) is sufficiently reduced. The success of a line search depends on the effective choices of both the search direction $d_k$ and the step size $a_k$. There are a lot of formulas in calculating the step size which are divided into an exact line search and an inexact line search.

The ideal choice would be the exact line search formula which is defined as $a_k = \arg\min (f(x_k + a_k d_k)\ \alpha > 0$ but in general it is too expensive to identify this value. Generally, it requires too many evaluations of the objective function f and also its gradient g. The inexact line search has a few formulas which have been presented by previous researchers such as the Armijo (1966) line search Wolfe (1969, 1970) condition and Goldstein (1965) condition. Shi (2006) claims that among several well-known inexact line search procedures, the Armijo line search is the most useful and the easiest to implement in the computational calculation. It is also easy to implement it in programming like Matlab and Fotran. The Armijo line search is described as follows. Given $s > 0, \lambda \in (0, 1), \sigma \in (0, 1)$ and $\alpha_i = \max\{s, s\lambda, s\lambda^2, \ldots\}$ such that:

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\sigma \alpha_k g_k^T d_k \quad (7)$$

k = 0, 1, 1, 2, 3, ....The reduction in f should be proportional to both the step size and directional derivative $g_k^T d_k$. The search directions are also important in order to determine the value of f which decreases along the direction. Moreover, the search direction of the quasi Newton methods often has the form:

$$d_k = -B_k^{-1} g_k \quad (8)$$

where, $B_k$ is a symmetric and non-singular matrix of approximation of the Hessian (Eq. 3). Initial matrix $B_0$ is chosen by an identity matrix which subsequently is updated by an update formula. When $d_k$ is defined by Eq. 8 and $B_k$ is a positive definite, we have $d_k^T g_k = -g_k^T B_k^{-1} g_k < 0$ and therefore $d_k$ is a descent direction. Hence, the algorithm for an iteration method of ordinary Broyden is described as follows:

**Algorithm 1 (Broyden method):**
Step 0: Given a starting point $x_0$ and $B_0 = I_n$. Choose values for s, $\beta$ and $\sigma$
Step 1: Terminate if $\|g(x_{k+1})\| < 10^{-6}$
Step 2: Calculate the search direction by Eq. 8
Step 3: Calculate the step size $a_k$ by the Armijo Line Search Eq. 7
Step 4: Compute the difference $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$
Step 5: Update $B_k$ by Eq. 3 to obtain $B_{k+1}$
Step 6: Set k = k+1 and go to Step 1

**A new search direction:** In this study, researchers will discuss the new search direction for the quasi Newton methods which will be proposed by using the concept of the conjugate gradient method. The search direction of conjugate gradient method is:

$$d_k = \begin{cases} -g_k & k=0 \\ -g_k + \beta_k d_{k-1} & k \geq 1 \end{cases} \quad (9)$$

where, $\beta_k$ is a coefficient of the conjugate gradient method. So, the concept of the conjugate gradient method's search direction will be implemented into the new search direction as introduced by Ibrahim *et al.* (2014). Therefore, the new search direction for the quasi-Newton method known as the CG-Broyden method is:

$$d_k = \begin{cases} -B_k^{-1} g_k & k=0 \\ -B_k^{-1} g_k + \lambda_k d_{k-1} & k \geq 1 \end{cases} \quad (10)$$

where, $\lambda_k = \eta g_k^T g_k / g_k^T d_{k-1}$ and $\eta \in (0,1)$ with these considerations in mind we shall now propose the algorithm for the CG-Broyden method as follows:

**Algorithm 2 (CG-Broyden method):**
Step 0: Given a starting point $x_0$ and $B_0 = I_n$. Choose values for s, $\beta$ and $\sigma$
Step 1: Terminate if $\|g(x_{k+1})\| < 10^{-6}$
Step 2:Calculate the search direction by Eq. 10
Step 3: Calculate the step size $\alpha_k$ by Eq. 7
Step 4: Compute the difference $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$
Step 5: Update $B_k$ by Eq. 3 to obtain $B_{k+1}$
Step 6: Set k = k+1 and go to Step 1

Based on Algorithms 1 and 2 we assume that every search direction $d_k$ satisfied the descent condition:

$$g_k^T d_k < 0 \quad (11)$$

for all k≥0. If there exists a constant $c_1 > 0$ such that:

$$g_k^T d_k \leq c_1 \|g_k\|^2 \quad (12)$$

for all k≥0, then the search directions satisfy the sufficient descent condition which can be proof in Theorem 3.2. Hence, we make a few assumptions based on the objective function.

**Assumption:**

- $H_1$: The objective function is twice continuously differentiable
- $H_2$: The level set is convex. Moreover, positive constants and exist, satisfying for all and where is the Hessian matrix for

$$c_1 \|z\|^2 \leq z^T F(x) z \leq c_2 \|z\|^2 \quad (13)$$

for all $z \in R^n$ and $x \in L$ where, F (x) is the Hessian matrix for f:

- $H_3$: The Hessian matrix is Lipschitz continuous at the point that is the positive constant exists, satisfying

$$\|G(x) - G(x^*)\| \leq c_3 \|x - x^*\| \quad (14)$$

for all x in a neighborhood of x*. If the iterates $\{x_k\}$ are converging to a point x*, it is to be expected that $y_k$ is approximately equal to $G(x^*)s_k$.

**Theorem 1 (Byrd and its proof):** Let $\{B_k\}$ be generated by the BFGS Eq. 3 where $B_1$ is symmetric and positive definite and where $y_k^T s_k > 0$ for all k. Furthermore, assume that $\{s_k\}$ and $\{y_k\}$ are such that:

$$\frac{\|(y_k - G^*) s_k\|}{\|s_k\|} \leq \varepsilon_k$$

for some symmetric and positive definite matrix G(x*) and for some sequence $\{\epsilon_k\}$ with the property $\sum_{k=1}^{\infty} \varepsilon_k < \infty$. Then:

$$\lim_{k \to \infty} \frac{\|(B_k - G_*) d_k\|}{\|d_k\|} = 0 \quad (15)$$

and the sequences $\{\|B_k\|\}$ or $\{\|B_k^{-1}\|\}$ are bounded.

**Theorem 2:** Suppose that Assumption 1 and 2 hold. Then, condition (Eq. 12) holds for all K≥0.

**Proof:** From Eq. 9, we see that:

$$\begin{aligned} g_k^T d_k &= -g_k^T B_k^{-1} g_k - \eta g_k^T d_{k-1} \\ &= -g_k^T B_k^{-1} g_k - \eta \frac{g_k^T g_k}{g_k^T d_{k-1}} g_k^T d_{k-1} \end{aligned}$$

and using the cauchy inequality we get:

$$\begin{aligned} g_k^T d_k &\leq -g_k^T \delta_k g_k - \eta g_k^T g_k \\ &\leq -\delta_k \|g_k\|^2 - \eta \|g_k\|^2 \\ &\leq c_1 \|g_k\|^2 \end{aligned}$$

where, $c_1 = -(\delta_k + \eta)$ which is bounded away from zero. Hence, Eq. 12 holds and the proof is completed.

**Lemma 1:** Under assumption 1, positive constants $c_2$ and $\varpi$ exist such that for any $x_k$ and any $d_k$ with $g_k^T d_k < 0$ the step size $a_k$, produced by Algorithm 2 will satisfy either:

$$f(x_k + \alpha_k d_k) - f_k \le -c_4 \frac{(g_k^T d_k)^2}{\|d_k\|^2} \quad (16)$$

Or:

$$f(x_k + \alpha_k d_k) - f_k \le c_5 g_k^T d_k$$

**Proof:**
Suppose that $a_k$<1 which means that (Eq. 7) failed for step size a' $\le$ a/$\tau$:

$$f(x_k + \alpha'_k d_k) - f(x_k) \le \varpi a' g_k^T d_k \quad (17)$$

Then, using the mean value theorem we obtain:

$$f(x_{k+1}) - f(x_k) = \bar{g}^T (x_{k+1} - x_k)$$

where, $\bar{g} = \nabla f(\bar{x})$ for some $\bar{x} \in (x_k, x_{k+1})$. Now, by the Cauchy-Schwartz inequality, we get:

$$\begin{aligned} \bar{g}^T (x_{k+1} - x_k) &= g^T (x_{k+1} - x_k) + (\bar{g} - g_k)^T (x_{k+1} - x_k) \\ &= g^T (x_{k+1} - x_k) + \|\bar{g} - g_k\| (x_{k+1} - x_k) \\ &\le g^T (x_{k+1} - x_k) + \mu \|x_{k+1} - x_k\|^2 \\ &\le g^T (a' d_k) + \mu \|a' d\|^2 \\ &\le g^T (a' d_k) + \mu (a' \|d\|)^2 \end{aligned}$$

Thus from $H_3$:

$$(\varpi - 1) a' g_k^T d_k < a' (\bar{g} - g_k)^T d_k \le M(a' \|d_k\|)^2$$

which implies that:

$$a_k \ge \tau a' > \tau(1 - \varpi) \frac{-g_k^T d_k}{M(a' \|d_k\|)^2}$$

Substituting this into Eq. 17, we have:

$$f(x_k + \alpha'_k d_k) - f(x_k) \le c_4 \frac{-g_k^T d_k}{(a' \|d_k\|)^2}$$

where, $c_6 = \tau(1 - \varpi)/M$ which gives Eq. 16.

**Theorem 3 (Global convergence):** Suppose that Assumption 1 and Theorem 1 hold. Then:

$$\lim_{k \to \infty} \|g_k\|^2 = 0$$

**Proof:** Combining the descent property (Eq. 12) and Lemma 1 gives:

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty \quad (18)$$

Hence, from Theorem 3 we can define that $\|d_k\| \le -c_1 \|g_k\|$. Then, Eq. 18 will be simplified as:

$$\sum_{k=0}^{\infty} \|g_k\|^2 < \infty$$

Therefore, the proof is completed.

## RESULTS AND DISCUSSION

**Numerical result:** In this study, researcher use a large number of test problem considered by Andrei (2008) and More *et al.* (1981) in Table 1 to analyse the improvement of the CG-Broyden method with the Broyden method. The dimensions of the tests range between 2 and 1,000 only.

The comparison between Algorithm 1 (Broyden) and Algorithm 2 (CG-Broyden) uses the cost of computation based on the number of iterations and CPU-time. As suggested by More *et al.* (1981) for each of the test problems, the initial point will take further away from the minimum point $x_0$ and we analyse three of initial points of each of test problems. In doing so, it leads us to test the

Table 1: A list of problem functions

| Test problem | N-dimensional | Sources |
|---|---|---|
| Powell badly scaled | 2 | More *et al.* (1981) |
| Beale | 2 | More *et al.* (1981) |
| Biggs exp6 | 6 | More *et al.* (1981) |
| Chebyquad | 4, 6 | More *et al.* (1981) |
| Colville polynomial | 4 | Michalewicz and Hartley (1996) |
| Variably dimensioned | 4, 8 | More *et al.* (1981) |
| Freudenstein and Roth | 2 | More *et al.* (1981) |
| Goldstein pricepolynomial | 2 | Michalewicz and Hartley (1996) |
| Himmelblau | 2 | Andrei (2008) |
| Penalty 1 | 2, 4 | More *et al.* (1981) |
| Extended powell singular | 4, 8 | More *et al.* (1981) |
| Extended rosenbrock | 2, 10, 100, 200, 500, 1000 | Andrei (2008) |
| Trigonometric | 6 | Andrei (2008) |
| Watson | 4, 8 | More *et al.* (1981) |
| Six-hump camel back polynomial | 2 | Michalewicz and Hartley (1996) |
| Extended shallow | 2, 4, 10, 100, 200, 500, 1000 | Andrei (2008) |
| Extended strait | 2, 4, 10, 100, 200, 500, 1000 | Andrei (2008) |
| Scale polynomial | 2 | Michalewicz and Hartley (1996) |
| Raydan 1 | 2, 4 | Andrei (2008) |
| Raydan 2 | 2, 4 | Andrei (2008) |
| Diagonal 3 | 2 | Andrei (2008) |
| Cube | 2, 10, 100, 200 | More *et al.* (1981) |

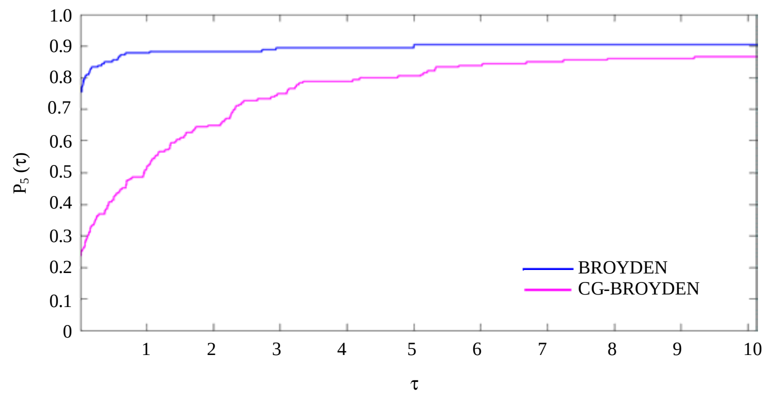


Fig. 1: Performance profile in a $\log_{10}$ scaled based on iteration

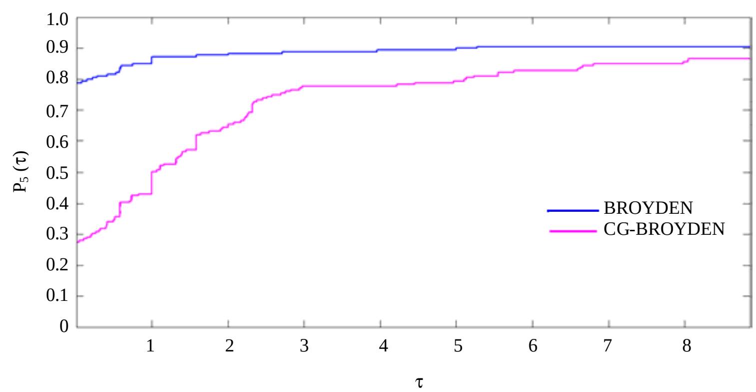


Fig. 2: Performance profile in a $\log_{10}$ scaled based on CPU time

global convergence properties and the robustness of our method. For the Armijo line search, we use s = 1, $\beta = 0.5$ and $\sigma = 0.1$. In our implementation, the programs are all written in Matlab. The stopping criteria that we used in both algorithms are $\|g(x_{i+1})\| \leq 10^{-6}$. The Euclidean norm is used in the convergence test to make these results comparable. The performance results will be shown in Fig. 1 and 2, respectively using the performance profile introduced by Dolan and More (2002). The performance profile seeks to find how well the solvers perform relative to the other solvers on a set of problems. In general $P(\tau)$ is the fraction of problems with performance ratio $\tau$ thus, a solver with high values of $P(\tau)$ or one that is located at the top right of the figure is preferable.

Figure 1 and 2 show that the CG-Broyden method has the best performance since it can solve 91% of the test problems while the Broyden method only solve 86%. Moreover, we can also say that the CG-Broyden method is the fastest solver on approximately 76% of the test problems for iteration and 79% of CPU-time. Therefore, the CG-Broyden method is better in solving the unconstrained optimization problems compare to the original Broyden method.

## CONCLUSION

We have presented a new hybrid method for solving unconstrained optimization problems. The numerical results for a small dimension of test problems show that the CG-Broyden method is efficient and robust in solving unconstrained optimization problems. The numerical results and figures from the programming are reported and analysed to show the characters of the proposed method.

## RECOMMENDATIONS

Our further interest is to try the CG-Broyden method with the coefficient of the conjugate gradient methods Fletcher and Reeves (1964), Hestenes and Steifel (1952) and Liu and Storey (1991) coefficient for $\beta_k$.

## REFERENCES

Andrei, N., 2008. An unconstrained optimization test functions collection. JAMO., 10: 147-161.

Armijo, L., 1966. Minimization of functions having lipschitz continuous first partial derivatives. Pac. J. Math., 16: 1-3.

Broyden, C.G., 1970. The convergence of a class of double-rank minimization algorithms 2: The new algorithm. IMA. J. Appl. Math., 6: 222-231.

Byrd, R.H. and J. Nocedal, 1989. A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. SIAM J. Numer. Anal., 26: 727-739.

Chong, E.K.P. and S.H. Zak, 2001. An Introduction to Optimization. John Wiley and Sons, New York, pp: 365-433.

Dolan, E.D. and J.J. More, 2002. Benchmarking optimization software with performance profiles. Math. Programming, 91: 201-213.

Fletcher, R. and C.M. Reeves, 1964. Function minimization by conjugate gradients. Comput. J., 7: 149-154.

Goldstein, A.A., 1965. On steepest descent. J. Soc. Indu. Appl. Math. Ser. A. Control, 3: 147-151.

Hestenes, M.R. and E. Steifel, 1952. Method of conjugate gradient for solving linear equations. J. Res. Nat. Bur. Stand., 49: 409-436.

Ibrahim, M.A.H., M. Mamat and W.J. Leong, 2014. The hybrid BFGS-CG method in solving unconstrained optimization problems. Abstract Appl. Anal., 2014: 1-6.

Li, D.H. and M. Fukushima, 2001. A modified BFGS method and its global convergence in nonconvex minimization. J. Comput. Appl. Math., 129: 15-35.

Liu, Y. and C. Storey, 1991. Efficient generalized conjugate gradient algorithms part 1: Theory. J. Comput. Applied Mathe., 69: 129-137.

Mamat, M., I. Mohd, L.W. June and Y. Dasril, 2009. Hybrid broyden method for unconstrained optimization. Intl. J. Numer. Methods Appl., 1: 121-130.

Michalewicz, Z. and S.J. Hartley, 1996. Genetic Algorithms Data Structures Evolution Programs. Springer, Berlin, Germany, USA.,.

More, J.J., B.S. Grabow and K.E. Hillstrom, 1981. Testing unconstrained optimization software. ACM Trans. Math. Software, 7: 17-41.

Shi, Z.J., 2006. Convergence of quasi-newton method with new inexact line search. J. Math. Anal. Appl., 315: 120-131.

Wolfe, P., 1969. Convergence conditions for ascent method. SIAM Rev., 11: 226-235.

Wolfe, P., 1971. Convergence conditions for ascent methods II: Some corrections. SIAM. Rev., 13: 185-188.

Xu, D.C., 2003. Global convergence of the broyden's class of quasi-newton methods with nonmonotone linesearch. Acta Math. Appl. Sinica, 19: 19-24.