Journal of Engineering and Applied Sciences 12 (18): 4627-4631, 2017
ISSN: 1816-949X

Hybrid Quasi-Newton and Conjugate Gradient Method for Solving Unconstrained Optimization Problems

[1]Nurul Aini, [1]Nurul Hajar, [1]Mustafa Mamat, [1]Norhaslinda Zull, [2]Mohd Rivaie
[1]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia
[2]Department of Computer Science and Mathematics, Universiti Teknologi MARA, Terengganu, Malaysia

Abstract: A hybrid of quasi-Newton and conjugate gradient method combines the search direction of the two methods to produce a new algorithm for solving unconstrained optimization functions. In this study, a modified hybrid Quasi-Newton method is presented where a new conjugate gradient coefficient is employed in the search direction. Based on the numerical results, the proposed hybrid method proved to be robust in comparison to the original quasi-Newton method and other hybrid methods.

Key words: Hybrid quasi-Newton, conjugate gradient, Armijo line search, hybrid methods, optimization, modified

INTRODUCTION

The general form of an unconstrained optimization problem is defined by:

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

where, f: $R^n \rightarrow R$ is a continuously differentiable function and its gradient at point x_k is denoted as g_k. The iterative method is often used to solve Eq. 1 and it is written as:

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

The variable $\alpha_k > 0$ denotes the stepsize, x_k is the kth iterative point and d_k is the search direction. The value of α_k can be obtained by two ways-exact and inexact line search. While the exact line search calculates the optimal stepsize, it is known to be very slow and ineffective when the initial point is far from the solution (Sun and Yuan, 2006). On the other hand, the inexact line search approximates the value of α_k according to some conditions which makes it faster and easier to implement. In this study, the Armijo line search is used to estimate the value of α_k due to its simplicity and easy application (Armijo, 1966). By using the Armijo conditions, the α_k is selected where for given s>0, $\beta \in (0, 1)$ and $\sigma \in (0, 1)$ we have $\alpha_k = \max\{s, s\beta, s\beta^2, ...\}$ such that:

$$f(x_k) - f(x_k + a_k d_k) \geq -\sigma \alpha_k g_k^T d_k, k = 0, 1, 2, ... \tag{3}$$

The Quasi-Newton method is well known to be highly efficient in solving small to medium scale optimization problems. Of the known different types of quasi-Newton method, it is generally accepted that the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is the most efficient one (Broyden *et al.*, 1973; Byrd and Nocedal, 1989). For this reason, the focus of this study will be on the BFGS method. Its search direction is written as:

$$d_k = -H_k g_k \tag{4}$$

where, H_k is the positive definite n×n approximate matrix of the Hessian of the objective function f at kth iteration while g_k is the gradient of f at point x_k. The update equation of the approximate Hessian matrix is as follows:

$$H_{k+1} = H_k + \left(1 + \frac{y_k^T H_k y_k}{s_k^T y_k}\right) \frac{s_k s_k^T}{s_k^T y_k} - \frac{s_k y_k^T H_k + H_k y_k s_k^T}{s_k^T y_k} \tag{5}$$

with $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$. While it is true that BFGS is robust when solving small scale problems, the same cannot be said for large scale ones. For problems involving large number of variables, the Conjugate Gradient (CG) algorithm is the best method to be used. The search direction of CG method is:

$$d_k = \begin{cases} -g_k, & k = 0 \\ -g_k + \beta_k d_{k-1}, & k \geq 1 \end{cases} \tag{6}$$

Corresponding Author: Nurul Aini, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia

where, β_k is the CG coefficient. Some examples of β_k are Fletcher-Reeves (FR) (Fletcher and Reeves, 1964), Conjugate Descent (CD) (Fletcher, 1987) and Rivaie-Mustafa-Ismail-Leong (RMIL) (Rivaie *et al*., 2012). Their equation are as follows:

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} \quad (7)$$

$$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}} \quad (8)$$

$$\beta_k^{RMIL} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T d_{k-1}} \quad (9)$$

In a bid to improve the original BFGS method, Ibrahim *et al*. (2014) combined the search direction of BFGS with the CG method and came up with a new d_k shown by:

$$d_k = \begin{cases} -H_k g_k, & k = 0 \\ -H_k g_k + \eta(-g_k + \beta_k d_{k-1}), & k \geq 1 \end{cases} \quad (10)$$

$$\beta_k = \frac{g_k^T g_{k-1}}{g_k^T d_{k-1}} \quad (11)$$

with $\eta>0$. This method, dubbed the BFGS-CG method was shown to out perform BFGS and some classical conjugate gradient methods in terms of number of iterations and CPU time. Following that, another hybrid BFGS and CG method was proposed by Ibrahim *et al*. (2014) which is referred as the HBFGS method. Its search direction is defined by:

$$d_k = \begin{cases} -H_k g_k, & k = 0 \\ -H_k g_k + \eta \beta_k d_{k-1}, & k \geq 1 \end{cases} \quad (12)$$

$$\beta_k = \frac{g_k^T g_k}{g_k^T d_{k-1}} \quad (13)$$

where, $\eta>0$. Based on the numerical results, the HBFGS method proved to be more efficient than the BFGS method. Both hybrid methods by Ibrahim *et al*. (2014) and Hery *et al*. (2014) were tested under Armijo line search.

Another known hybrid method combines quasi-Newton with the Steepest Descent (SD) method which is known for its global convergence property (Han and Newman, 2003). A hybrid of the quasi-Newton and Gauss-Seidel method aimed at solving the system of linear equations was presented by Ludwing (2007), Sofi *et al*. (2013, 2008), Mamat *et al*. (2009), Jaafar *et al*. (2013). On the other hand, Luo *et al*. (2008) proposed combining the quasi-Newton method with chaos optimization to solve the system of nonlinear equations.

MATERIALS AND METHODS

Hybrid quasi-Newton method: In this study, a modified hybrid Quasi-Newton method is proposed based on the method studied by Ibrahim *et al*. (2014). We apply a different CG coefficient in place of the one used by Ibrahim *et al*. (2014). The new β_k is known as β_k^{ARM} or ARM method and is formulated as:

$$\beta_k^{ARM} = -\frac{m_k \|g_k\|^2 - |g_k^T g_{k-1}|}{m_k g_{k-1}^T d_{k-1}}, \; m_k = \frac{\|d_{k-1} + g_k\|}{\|d_{k-1}\|} \quad (14)$$

Note that:

$$\beta_k^{ARM} = \frac{m_k \|g_k\|^2 - |g_k^T g_{k-1}|}{m_k(-g_{k-1}^T d_{k-1})} \leq \frac{m_k \|g_k\|^2}{m_k(-g_{k-1}^T d_{k-1})} = \beta_k^{CD} \quad (15)$$

From the Eq. 15, the ARM method can be simplified to the Conjugate Descent (CD) method, a type of classical CG method that has been reported to show good results (Du *et al*., 2001; Yuhong and Yaxiang, 1996). We call the resulting hybrid method the BFGS-ARM method and implement it in the following algorithm:

Algorithm 1 (Hybrid method):

Step 1: Given an initial point $x_0 \in R^n$, $x_0 \in R^{n\times n}$, set k = 0
Step 2: If the stopping criteria $\|g_k\| \leq 10^{-6}$ or k = 10000 is fulfilled, stop
Step 3: Compute the descent direction variable by (12) using (14) as β_k
Step 4: Compute a_k by Armijo line search (Eq. 3)
Step 5: Compute $x_{k+1} = x_k + \alpha_k d_k$
Step 6: Set $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$
Step 7: Compute H_{k+1} by (5)
Step 8: Set k = k+1 and go to Step 2

RESULTS AND DISCUSSION

Convergence analysis: A convergent algorithm should satisfy the sufficient descent and global convergence properties. In order to establish the convergence of the new CG method, we need the following assumptions of the objective function. Assuming that every d_k satisfies the descent condition, then:

$$g_k^T d_k < 0 \quad (16)$$

Table 1: A list of test problems

Test problems	Variables	Initial points
Three hump	2	(6, 6) (-19, 17) (61, 61)
Six hump	2	(6, 7) (18, 18) (46, 46)
Zettl	2	(6, 6) (16, 16) (64, 64)
Dixon and price	2, 4	(7, 7) (18, -19) (56, 56)
Raydan 1	2, 4	(7, …, 7), (12, …, 12) (22, …, 22)
Arwhead	2, 4, 10	(3, …, 3) (23, …, 23) (81, …, 81)
Generalized tridiagonal 1	2, 4, 10	(3, …, 3),(14, …, 14) (70, …, 70)
Extended quadratric penalty 2	2, 4, 10	(8.9, …, 8.9) (29, …, 29) (99, …, 99)
Extended beale	2, 4, 10, 100	(-1.3, …, -1.3) (2, …, 2) (-10, 10, …, -10, 10)
Tridia	2, 4, 10, 100, 500, 1000	(-7, …, -7) (15, …, 15) (63, …, 63)
Denschnb	2, 4, 10, 100, 500, 1000	(5, …, 5) (25, …, 25) (100, …, 100)
Extended rosenbrock	2, 4, 10, 100, 500, 1000	(10, …, 10) (18, …, 18) (55, …, 55)
Extended white and holst	2, 4, 10, 100, 500, 1000	(-4, 4, …, -4,4) (15, …, 15) (-43, …, -43)

for all $k \geq 0$. For the sufficient condition to hold:

$$g_k^T d_k \leq -C\|g_k\|^2 \tag{17}$$

To prove that the BFGS-ARM algorithm satisfies the sufficient descent property, the following assumption is needed.

Assumption 1: The objective function f is twice continuously differentiable. In some neighbourhood N of l, f (x) is continuously differentiable and its gradient is Lipschitz continuous then, there exists a constant $L>0$ such that $\|g(x)-g(y)\| \leq L\|x-y\|$ for all x, $y \in N$. The level set L is convex. There exists positive constants c1 and c2 satisfying:

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

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

Theorem 1: Suppose that Assumption 1 holds. Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by the algorithm for BFGS-ARM method while the stepsize is determined by using Armijo line search, then the sufficient descent condition $g_k^T d_k \leq -C\|g_k\|^2$ holds true for all $k \geq 0$ and $C>0$.

Theorem 2: Suppose that assumption 1 and theorem 1 hold. The sequences $\{g_k\}$ and $\{d_k\}$ are generated by the algorithm for BFGS-ARM method while the stepsize is determined by using Armijo line search. Then:

$$\lim_{k \to \infty} \|g_k\| = 0 \tag{18}$$

Numerical results: In this study, we discuss the numerical results of the tests conducted on the BFGS-ARM method in comparison to the original BFGS method, the BFGS-CG method (Ibrahim *et al*., 2014) and the HBFGS method (Ibrahim *et al*., 2014). Under Armijo line search, the efficiency of the four methods are studied by using them to solve 13 test problems taken from Ibrahim *et al.* (2014), Jamil and Yang (2013), Andrei (2008) at different number of variables from 2-1000. Following the suggestion by Hilstrom (1997) for each of the tests, the initial point begins from one close to the minimum point to one furthest away from the solution. This is to test the global convergence properties and the robustness of the proposed method. For the Armijo condition, we define s = 1, β = 0.5 and σ = 0.1. All of the calculations involved are made by Matlab (2012) subroutine programme via a portable PC with CPU processor intel (R) Core (TM) i5 and 4GB RAM memory.

The stopping criteria are set at $\|g_k\| \leq 10^{-6}$ and when the number of iterations exceed 1000. The results are recorded in terms of number of iterations, number of function evaluations and CPU time. They are then analysed by use of performance profile introduced by Dolan and More which is a mean to evaluate and compare the performance of the set of solvers S on a test set P (Dolan and More, 2002). The lists of functions used are as displayed in Table 1.

Assuming there are n_s solvers and n_p problems for each problem p and solver s, we define $t_{p,s}$ = computing time or number of iterations or function evaluations required to solve problem p by solver s.

As a baseline is required for comparisons, the performance of solver s on problem p is compared with the best performance by any solver on the problem by using the performance ratio:

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s}: s \in S\}}$$

Assuming that a parameter $r_M \geq r_{p,s}$ for all p, s is chosen and $r_{p,s} = r_M$ if and only if solver s does not solve problem p, an overall assessment of the performance of solver s can be obtained from:

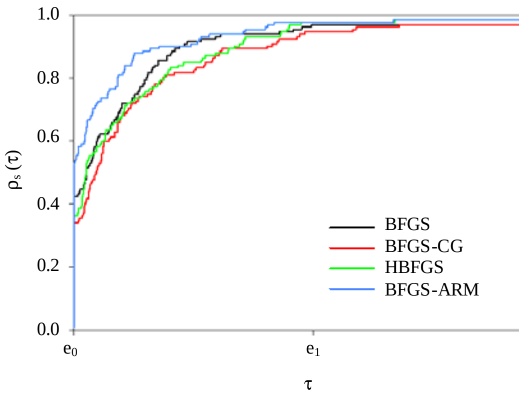

Fig. 1: Performance profile based on the number of iterations

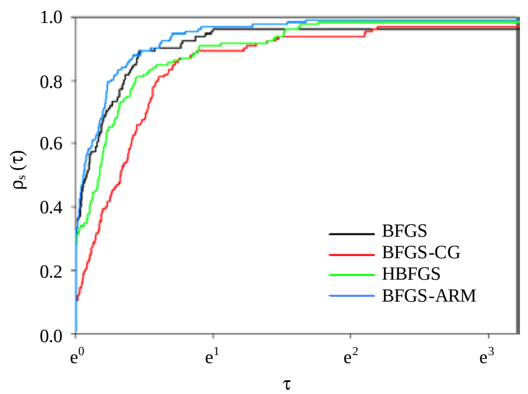

Fig. 2: Performance profile based on the CPU time

$$\rho_s(\tau) = \frac{1}{n_p} \text{size}\{p \in P: r_{p,s} \leq \tau\}$$

Where:

$\rho_s(\tau)$ = The probability for solver $s \in S$ that a performance ratio

$r_{p,s}$ = Within a factor $\tau \in R$ of the best possible ratio

Function ρ_s is the cumulative distribution function for the performance ratio. The performance profile ρ_s: R→[0, 1] for a solver is a nondecreasing, piecewise constant function which is continuous from the right. The value of $\rho_s(1)$ is the probability of a solver winning over the rest. Thus, it can also be said that the solver with highest value of $\rho_s(\tau)$ or located at the top right of the figure represents the best solver.

The following three figures show the performance of the four tested solvers based on number of iterations in Fig. 1, CPU time in Fig. 2 and number of function evaluations in Fig. 3. The left side of the figure determines the percentage of the test problems for which a solver is the fastest while the right side indicates the percentage of the test problems that are successfully solved by each solver.

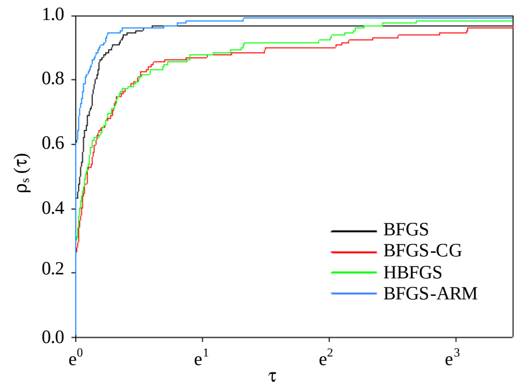

Fig. 3: Performance profile based on the number of function evaluations

Based on the results, both BFGS and BFGS-CG have the lowest percentage of test problems solved which is at 96.97% followed by HBFGS at 98.48%. This leaves the BFGS-ARM method with the highest number of problems solved at 99.24%.

CONCLUSION

This study proposes a modification on the hybrid BFGS and CG method by introducing an alternative choice of β_k referred here as the ARM method. The resulting algorithm is called the BFGS-ARM method. The efficiency of this solver is compared to the original BFGS method, the BFGS-CG method and the HBFGS method with the latter two both being hybrid quasi-newton and conjugate gradient method. Based on the results of the numerical tests, the new hybrid method is shown to be more efficient than the original BFGS and other hybrid quasi-Newton methods.

ACKNOWLEDGEMENTS

Researchers would like to thank the Ministry of Higher Education of Malaysia (MOHE) for the funding of this research under The Fundamental Research Grant Scheme (Grant No. 600-RMI/FRGS 5/3 (64/2015)) and MyPhD KPM.

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., J.E. Dennis and J.J. More, 1973. On the local and superlinear convergence of quasi-Newton methods. IMA. J. Appl. Math., 12: 223-245.

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.

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

Du, X.W., L.Q. Ye and C.X. Xu, 2001. Global convergence of a class of unconstrained optimal methods include the conjugate descent method. J. Eng. Math., 18: 120-122.

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

Fletcher, R., 1987. Practical Method of Optimization, Unconstrained Optimization. John Wiley & Sons, New York, USA., ISBN:9780471915478, Pages: 436.

Han, L. and M. Neumann, 2003. Combining quasi-Newton and steepest descent directions. Intl. J. Appl. Math., 12: 167-171.

Hery, M.A., M. Ibrahim and L.W. June, 2014. BFGS method: A new search direction. Sains Malaysiana, 43: 1591-1597.

Hilstrom, K.E., 1977. A simulation test approach to the evaluation of nonlinear optimization algorithms. ACM Trans. Mathe. Software, 3: 305-315.

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.

Jaafar, R., M. Mamat and I. Mohd, 2013. A new scaled hybrid modified BFGS algorithms for unconstrained optimization. Appl. Math. Sci., 7: 263-270.

Jamil, M. and X.S. Yang, 2013. A literature survey of benchmark functions for global optimisation problems. Intl. J. Math. Modell. Numer. Optimisation, 4: 150-194.

Ludwig, A., 2007. The Gauss-Seidel-quasi-Newton method: A hybrid algorithm for solving dynamic economic models. J. Econ. Dyn. Control, 31: 1610-1632.

Luo, Y.Z., G.J. Tang and L.N. Zhou, 2008. Hybrid approach for solving systems of nonlinear equations using chaos optimization and quasi-Newton method. Appl. Soft Comput., 8: 1068-1073.

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.

Rivaie, M., M. Mamat, L.W. June and I. Mohd, 2012. A new class of nonlinear conjugate gradient coefficients with global convergence properties. Appl. Math. Comput., 218: 11323-11332.

Sofi, A.Z.M., M. Mamat and I. Mohd, 2013. An improved BFGS search direction using exact line search for solving unconstrained optimization problems. Appl. Math. Sci., 7: 73-85.

Sofi, M.A.Z., M. Mamat, I. Mohd and Y. Dasril, 2008. An alternative hybrid search direction for unconstrained optimization. J. Interdiscip. Math., 11: 731-739.

Sun, W. and Y.X. Yuan, 2006. Optimization Theory and Methods: Nonlinear Programming. Springer-Verlag, New York, USA., ISBN-10: 0387249753, pp: 687.

Yuhong, D. and Y. Yaxiang, 1996. Convergence properties of the conjugate descent method. Adv. Math., 25: 552-562.