A SCALED CONJUGATE GRADIENT METHOD USING THE DFP UPDATE FOR UNCONSTRAINED OPTIMIZATION PROBLEMS

OSINUGA, I.A. and OLUBIYI, M.O.
DEPARTMENT OF MATHEMATICS,
FEDERAL UNIVERSITY OF AGRICULTURE, ABEOKUTA, OGUN STATE, NIGERIA.
Email: osinugaai@funaab.edu.ng, biyimoses01@gmail.com

ABSTRACT

The Conjugate Gradient (CG) and Quasi-Newton methods are famous methods for finding solution to optimization problems involving large variables such as problems in optimal inflation rate, minimal cost, maximal profit, minimal error, optimal design and many more. In this presentation, we propose a modification of the hybrid Davidon-Fletcher-Powell-Conjugate-Gradient (DFP-CG)methods developed by Wan Osman et.al (2017) by adapting a spectral-scaling memory less DFP-update. The numerical implementation of the proposed method on a some selected unconstrained optimization test problems by adopting the performance profile by Dolan et al. (2002) indicates that the newly suggested method is competitive, robust and in most instances more efficient when compared with some existing CG methods in the literature.

Keywords: DFP update, unconstrained optimization, Search Direction, Conjugate Gradient Method.

1. INTRODUCTION

Optimization comes from the word optimal which simply means "best". The word optimize simply means to make or take the best decision/choice among several available options under some conditions that must be satisfied. Optimization could be described as a methdod of determining the best solution to certain mathematically defined problems which are often models of physical reality Awatif et al. (2005). Mathematical Optimization is process that involves maximization (or minimizing) an objective (also known as cost) function by finding the desirable available values across set of choices at hand. i.e. mathematical optimization involves maximizing or minimizing mathematical functions of several variables, representing a stretch of choices available in the concerned situation. Common applications of mathematical optimization are found in optimal inflation rate, machine and deep learning, neural network, minimal cost, maximal profit, minimal error, data science and machine intelligence.

Optimization as a field of science is gaining wide range of adoption in the field of science as a result of its numerous ambit of application, as well as its usefulness in decision science and in the analysis of physical science systems. Al-Baali (1985) came up with the conclusion that optimization helps in determining the factors that gives the maximum (or minimum) value of a function, which often helps in making scientific decisions. The mathematical formulation for optimization problem is then given as

$$Optimize\ f(x)$$

$$x\ \varepsilon\ \Omega$$

subject to:
$$\begin{cases} D_i(x) = 0\,, & i\ \varepsilon\ P \\ \\ F_i\,(x) \geq 0, & i\ \varepsilon\ Q \end{cases} \quad (1)$$

Where optimize stands for minimize or maximize, and $f\colon R^n \rightarrow R,\ n > 1$, denotes the objective or cost function, which is continuously differentiable and $\Omega\ \varepsilon\ R^n$ is the feasible set containing admissible choices of x, P and Q are two finite set of indices. $D_i\ i\ \varepsilon\ P$ are the equality constraints and $F_i\ i\ \varepsilon\ Q$ are the inequality constraints.

The optimization methods based on line search utilizes the next iteration scheme:

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

Where x_k represent the current iterative point, x_{k+1} is the next iterative point, d_k is the search direction and $\propto_k$ denotes the step length.

2. REVIEW OF RELATED WORKS

In an attempt to solve large-scale optimization problems, it is more convenient, efficient and effective to use an iterative method compare to using a direct method which could be time consuming and computationally difficulty. The iterative scheme for unconstrained optimization problems is given by (2). Various iterative schemes has been suggested by numerous researcher using different approaches such as hybrid, scaled and parametric. Ahmed et al. (1989), Nocedal (2006), Dai et al. (2000) for more. The idea of combining Quasi-Newton and Conjugate Gradient methods was started by Buckley (1978). Luo *et al.* (2008) also combined Quasi-Newton and the Cauchy descent methods to arrive at the Quasi-Steepest Descent method. Jinkui (2014) proposed a new hybrid method which solves the system of non-linear equations by combining the Quasi-Newton method with the Chaos Optimization. Furthermore, Ibrahim *et al.(2014)* proposed a hybridization of Quasi-Newton method and the Conjugate method with the search direction given as a combination of the Broyden-Fletcher-Goldfarb-Shanno (BFGS) and conjugate gradient methods. Using the same idea, Wan-Osman et.al (2017) proposed a similar hybridization Quasi Newton and CG methods with the Davidon-Fletcher-Powell (DFP) method adopted. Other recent conjugate gradient methods can be found in Liu-Storey (LS) method, Liu *et al.(*1991), Rivaie-Mustafa-Ismail-Leong (RMIL) method Rivaie, et *al. (2012)*, Kamilu et al. (KMAR) method, Kamfa (2018), Saleh et al. (SM) method (2019), Sulaiman-Mustafa (SM1), Sulaiman, (2018), Usman et al. (2018) and Adeleke et al. (2018).

3. Motivation and the proposed Conjugate Method

We consider the quasi-Newton methods with the search direction (d_k) given by:

$$d_k = -H_k g_k \quad (3)$$

Here, H_k denotes the inverse Hessian approximation updated by the Broyden class. There exist several updating scheme for this update, some of which includes Symmetric Rank-one (SR1), Davidon-Fletcher-Powell (DFP), and Broyden-Fletcher-Goldfarb-Shanno (BFGS) updates.

If H_k is updated by DFP method, then,

$$H_{k+1} = H_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k H_k y_k^T}{y_k^T H_k y_k} \qquad (4)$$

with the secant equation $H_{k+1} y_k = s_k$ been satisfied

Memoryless quasi-newton approach is another way of solving an unconstrained optimization problem, where at each step of the iteration process, the Hessian approximation is updated using a multiple of an identity matrix and a spectral parameter θ_k. Thus the search direction is computed with no matrix computation nor its storage. In order enhance improve the computational performance of the classical DFP, the inverse Hessian is computed as a product of the identity matrix I and scaling parameter θ_k which is always positive, the restarted DFP updating scheme according to Arzuka *et al.* (2016) and Mamat *et al.* (2018) is given by:

$$H_{k+1} = I\theta_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{I\theta_k y_k I\theta_k y_k^T}{y_k^T I\theta_k y_k} \qquad (5)$$

Furthermore, Wan Osman *et al.*(2017) proposed a direction of descent titled "Hybrid-DFP-CG", this idea merges the attractive qualities of quasi-newton method and CG method. Wan Osman *et al.* (2017) came up with a descent direction d_k given by:

$$d_k = \begin{cases} -H_k g_k & k = 0 \\ -H_k g_k - gk + \beta_k d_{k-1}, & k \geq 1 \end{cases} \qquad (6)$$

Where $\beta_k = \frac{\eta g_k^T (g_k - g_{k-1})}{||g_{k-1}||^2}$, $\eta \, \varepsilon \, (0,1]$

i.e. $d_{k+1} = -H_{k+1} g_{k+1} - g_{k+1} + \beta_{k+1} d_k$ (7)

Motivated by the good performances of (5) and (6) above, we propose our search direction by incorporating the restarted DFP updating scheme (5) within the HDFP-CG (6) to have:

$$d_{k+1} = -\left(\theta_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{\theta_k y_k y_k^T}{y_k^T y_k}\right) g_{k+1} - g_{k+1} + \beta_k d_k \qquad (8)$$

and after some algebraic simplifications, we have that:

$$d_{k+1} = -(\theta_k + 1) g_{k+1} - \varphi_1 s_k + \varphi_2 y_k + \beta_k d_k \qquad (3.10)$$

where

$$\varphi_1 = \frac{s_k{}^T g_{k+1}}{s_k^T y_k} \quad , \varphi_2 = \frac{\theta_k y_k{}^T g_{k+1}}{y_k^T y_k} \ , \beta_k = \frac{g_k{}^T (g_k - g_{k-1})}{||g_{k-1}||^2} \quad \beta_k = \frac{g_k{}^T (g_k - g_{k-1})}{||g_{k-1}||^2}$$

Various scaling parameter θ_k can be considered, we prefer the following due to Mamat *et al.*(2008), given by $\theta_k = \frac{y_k^T y_k}{y_k^T s_k}$

Consider the following steps:

Step 1: Given any initial point $x_0 \varepsilon R^n$, determine $f(x_0)$ and $g(x_0)$, set $d_0 = -g_0$ and $k = 0$

Step 2: Stop if $||g_k|| \leq \varepsilon$, otherwise go to step 3

Step 3 Compute the step length $\propto_k$ using the Strong-Wolfe condition

Step 4: Compute $x_{k+1} = x_k + \propto_k d_k$

Step 5: Calculate d_k using (3.10)

Step 6: Put $k = k + 1$, then go to step 2.

4. NUMERICAL RESULTS AND DISCUSSION

The numerical result of the suggested algorithm against some well known conjugate gradient methods is reported. We test all these algorithms by using them to solve fifteen (5) test problems of dimensions 2, 80, 500, 1000,and 15600, which makes a total of 25 problems so as to evaluate the computational strength of the proposed method Scaled-Davidon-Fletcher-Powell (SDFP-CG) with Davidon-Fletcher-Powell (DFP), Wan Osman et al. (2017), Polyak, (1969), Andrei (2018), Zhang et al. (2007) and Arzuka et al.(2016) methods.

The parameter taken are σ = 0.9, δ=0.00001$ and the stop criterion as (i) $||g_k|| < 10^{-6}$ (ii) Number of iteration $(NI) < 1000$. All codes of the computer procedures were written in MATLAB R2018a with CPU 1.30GHz and 4.00GB RAM, Memory on HP650 and windows 7 operating system. We report the computational performances of the algorithms on some set of 50 unconstrained optimization test problems curled from Andrei[20]. The performance profile of Dolan E. *et al.*(2002) was adopted in the process of analyzing the computational strength of the proposed method Scaled-Davidon-Fletcher-Powell -(SDFP-CG) with Davidon-Fletcher-Powell - (DFP), Wan Osman et al. (2017)-HDFP-CG, Polyak (1969)-PRP, Andrei (2018)-AN, Zhang et al. (2007) - ZH and Arzuka et al.(2016) - AZ methods. The comparison is done by considering the number of iterations and the CPU time for each algorithm. Table 1 in the Appendices depicts the performances of the methods with respect to the number of iterations and the CPU time consumed. The proposed method (SDFP-CG), PRP, A, ZH And AZ solves 100% of the test problems while HDFP-CG method solves 84% of the test problems. From the results, it can also be seen that the proposed method (SDFP-CG), AZ, and AN are competitive and could be seen in their various numbers of iterations and CPU time. Although, DFP-CG methods seems not to be competitive with other algorithms. However, it solves lower dimensional problems better than the other algorithms. Furthermore, SDFP-CG outperforms all other algorithms in most cases especially when the dimension goes higher.

5. CONCLUSION

We have been able to propose a modification of hybrid DFP-CG method, by adopting the DFP updating scheme of the inverse Hessian approximation in the frame of a memory-less quasi-Newton Approach.

Furthermore, we establish the effectiveness and efficiency of the proposed method in solving some selected large-scale unconstrained optimization test problems when compared with some classical and hybrid methods.

REFERENCES

[1] Adeleke, O. J. and Osinuga, I. A. (2018): A five term hybrid Conjugate Gradient method with global convergence and descent properties for unconstrained optimization problem. Asian Journal of Sci. Res. 11:185-194.

[2] Al-Baali, M (1985): Descent property and global convergence of the fletcher Reeves method with exact line search. *IMAJ Numerical Analysis 5.* 121-124.

[3] Andrei, N. (2008): An Unconstrained Optimization Test Functions Collection. Advanced Modelling and Optimization.10(1), 147-161.

[4] Awatif, M.A. and Elsiddieg K. (2005): The Davidon Fletcher and Powell method tested on Quadratic functions, *International Journal of Basic Sciences and Applied Computing (IJBSAC), 2394-367X, volume 1, Issue 10.*

[5] Azurka, I. and Mohd, R. (2016): A scaled three term Conjugate Gradient Method for unconstrained optimization. *Journal of inequalities and Applications.*

[6] Buckley A.G. (1978):A combined Conjugate-Gradient Quasi-Newton Minimization Algorithm.Mathematical Progamming 15(1) 200-210.

[7] Dai, Y.H. and Yuan, Y.X. (2000): Non linear Conjugate Gradient Method. *Shangai Scientific and Technical Publishers,* Shangai

[8] Djamal, Touti Ahmed (1989): A study of Hybrid Conjugate Gradient Method. Phd Thesis, Loughborough University of Technology,United Kingdom.

[9] Dolan, E. and More, J. (2002). Benchmarking optimization software with performance profile. Mathematical Proramming, 91, 201213.

[10] Ibrahim, M. A. H., Mustafa M., Leong W. J., (2014). The Hybrid BFGS - CG Method in Solving Unconstrained Optimization Problems. Abstract and Applied Analysis, Volume 2014 Article ID 507102, 6pp.

[11] Jinkui, L. and Xuesha, W. (2014): New three-term Conjugate Gradient Method for solving unconstrained Optimization problems. ScienceAsia/ doi: 10.2306/science asia1513-1874.2014.40.295. 295-300.

[12] Kamfa, K (2018): Another improved three term PRP-CG method with global Convergent properties for solving unconstrained optimization problems. Malaysian Journal of Computing and Applied Mathematics. 1(1), 1-10.

[13] Liu Y., and Storey, C.(1991): Efficient generalized Conjugate Gradient Algorithm, Part 1. Journal of Optimization Theory and Application 69 129-137.

[14] Luo, Y.Z, Tang, G. and Zhou, L. (2008): {Hybrid approach for solving systems of nonlinear equations using chaos optimization and quasi-Newton method. Applied Soft Computing. 8(2), 1068-1073.

[15] Mamat, M. et.al (2018): Derivative free Davidon-Fletcher-Powell (DFP) for solving Symmetric Systems of non-linear equations. *IOP conference series, Material Science and Engineering.* 332-012030.

[16] Nocedal Jorge and Stephen J. Wright (2006): Numerical Optimization. *ISBN-10:0-387-3033-0, ISBN-13: 978-387-303-1.*

[17] Polyak, B.T. (1969): The Conjugate Gradient method in Extremal problems. Computational Mathematics and Mathematical Physics 9(4),94-112.

[18] Rivaie, M. and Mohd, I. (2012): A new class of non-linear Conjugate Gradient Coefficient with global convergence properties. Applied mathematics and Computation. 218(22), 11323-11332.

[19] Saleh, N.A, Mamat, M., Rivaie, M. and Sulaiman I. (2019). A combination of FR and HR coefficient in Conjugate Gradeint Method for unconstrained optimization. Malaysian Journal of Computing and Applied Mathematics, Vol 2(1), 42-49.

[20] Sulaiman, I.M. (2018): Solving fuzzy Non-linear equations with a new class of Conjugate Gradient method. Malaysian Journal of Computing and Applied Mathematics. 1(1), 11-19.

[21] Usman Y.A., Mamat, M. Mohamed, M.A., Firman, S. and Rivaie M. (2018): Modification of spectral Gradient method for unconstrained optimization. International Journal of Engineering and Technology, 7(3.28) 307-311.

[22] Wan Osman, W.F.H , M.A.H. Ibrahim and Mamat M. (2017): Hybrid DFP-CG method for solving unconstrained optimization problems. Journal of Physics, Conference series. 890-012033. IOP publishing. UK. Wan Osman, W.F.H , M.A.H. Ibrahim and Mamat M. (2017): Hybrid DFP-CG method for solving unconstrained optimization problems. Journal of Physics, Conference series. 890-012033. IOP publishing. UK.

[23] Zhang, L., Zhou, W. and Li, D.H (2007): Some descent three-term Conjugate Gradient methods and their global convergence. Optimization methods Software, 22(4),697-711.

APPENDIX 1

PROBLEM	DIMENSION	SDFP-CG	HDFP	PRP	ZH	AN	AZ
SINQUAD FUNCTION	2	0.042636/4	0.003817/14	0.047863/4	0.042676/4	0.041612/4	0.043123/4
	80	0.042643/5	0.057942/14	0.165632/1000	0.154767/1000	0.042983/5	0.042903/5
	500	0.045504/5	0.083721/14	0.546488/1000	0.523253/1000	0.046243/5	0.045506/5
	1000	0.050193/5	0.181617/14	0.048894/6	0.059143/1000	0.052211/5	0.943436/1000
	15600	0.141963/6	NAN	0.145502/6	0.158343/7	0.145436/6	15.334324/1000
HIMMELBG FUNCTION	2	0.044182/5	0.056163/27	0.474556/1000	0.409021/1000	0.044204/5	0.044217/5
	80	0.044464/5	0.056143/30	0.520648/1000	0.435109/1000	0.044983/5	0.044466/5
	500	0.046283/5	0.419828/30	0.918514/1000	0.809783/1000	0.047282/5	0.046284/5
	1000	0.049663/5	1.076584/32	1.403584/1000	1.317489/1000	0.049406/5	0.049681/5
	15600	0.125323/5	NAN	18.416172/1000	18.11461/1000	0.135562/5	0.128701/5
TRIDIA FUNCTION	2	0.411783/1000	0.034559/4	0.450142/1000	0.045244/5	0.021584/5	0.021844/5
	80	0.044984/5	0.025402/8	0.536231/1000	0.476592/1000	0.044723/5	0.044986/5
	500	0.047842/5	0.044132/9	1.011438/1000	0.980848/1000	0.047325/5	0.047901/5
	1000	0.050704/5	0.104889/9	1.723969/1000	1.70667/1000	0.050851/5	0.049993/5
	15600	0.124298/5	NAN	15.884321/1000	15.085847/1000	0.124543/5	0.130004/5
RAYDAN 1 FUNCTION	2	0.044252/5	0.02623/4	0.474206/1000	0.409467/1000	0.043944/5	0.044464/5
	80	0.049317/5	0.01523/3	0.534604/1000	0.673533/1000	0.044521/5	0.044738/5
	500	0.048438/5	0.05941/3	1.623386/1000	1.253386/1000	0.048751/5	0.048531/5
	876	0.049676/5	0.14103/3	3.601714/1000	3.601714/1000	0.049923/5	0.049923/5
	1000	0.490887/5	0.13625/3	5.381667/1000	5.381667/1000	0.050704/5	0.057036/5
	15600	0.126606/5	NAN	23.641612/1000	20.649092/1000	0.127142/5	0.136264/5
EXTENDED ROSENBROCK FUNCTION	2	0.03824/4	0.045213/4	0.038743/4	0.037977/4	0.037964/4	0.037703/4
	50	0.037949/4	0.031355/4	0.038483/4	0.038206/4	0.037965/4	0.037963/4
	80	0.038509/4	0.029384/4	0.038219/4	0.038227/4	0.038246/4	0.037962/4
	500	0.040042/4	0.088658/5	0.039503/4	0.040058/4	0.039536/4	0.040043/4
	1000	0.041602/4	0.230987/5	0.041878/4	0.041338/4	0.041338/4	0.041346/4
	15600	0.094491/4	14.876543/8	0.094916/4	0.096476/4	0.095164/4	0.096737/4

Table 1: ***Numerical Results showing the CPU time and the number of iterations of each method and the dimensions.***