International Journal of Engineering & Technology, 7 (2.14) (2018) 21-24

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

A modifications of conjugate gradient method for unconstrained optimization problems

Omar Alshorman[1], Mustafa Mamat[1*], Ahmad Alhawarat[2], Mohd Revaie[3]

[1]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, 21300 Kuala Nerus, Terengganu, Malaysia.
[2]College of science, Isra University, Amman, Jordan
[3]Department of Computer Science and Mathematics, Universiti Technology Mara, Terengganu, Malaysia
**Corresponding author E-mail: must@unisza.edu.my*

Abstract

The Conjugate Gradient (CG) methods play an important role in solving large-scale unconstrained optimization problems. Several studies have been recently devoted to improving and modifying these methods in relation to efficiency and robustness. In this paper, a new parameter of CG method has been proposed. The new parameter possesses global convergence properties under the Strong Wolfe-Powell (SWP) line search. The numerical results show that the proposed formula is more efficient and robust compared with Polak-Rribiere Ployak (PRP), Fletcher-Reeves (FR) and Wei, Yao, and Liu (WYL) parameters.

Keywords: *Conjugate Gradient Parameter; Inexact Line Search; Strong Wolfe-Powell Line Search; Global Convergence; Unconstrained Optimization.*

1. Introduction

The nonlinear Conjugate Gradient (CG) methods are designed to solve the following unconstrained optimization problems:

$$\min f(x) \quad , x \epsilon R^n \tag{1}$$

where $f: R^n \to R$ is continuously differentiable function, and its gradient is denoted by $g(x) = \nabla f(x)$.
The nonlinear CG method is an iterative method of the formula:

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

where x_k is the current iterate and α_k is known as the step size obtained using Strong Wolfe-Powell (SWP) line search, and d_k denotes the search direction defined by the formula:

$$d_k = \begin{cases} -g_k & , \text{ if } k = 0, \\ -g_k + \beta_k d_{k-1} & , \text{if } k \geq 1, \end{cases} \tag{3}$$

where $\beta_k \epsilon R$ is a scalar and $g_k = g(x_k)$. There are various familiar CG parameters such as Fletcher-Reeves (FR) 1964 [9], Hestenes-Stiefel (HS) 1952 [13], Polak-Ribiere-Polyak (PRP) 1969 [16, 15], Liu-Storey (LS) 1992 [14], Conjugate Decent (CD) 1967 [10], and Dai-Yuan (DY) 1999 [7]. Several authors studied the convergence behavior with some line search conditions, for example [1-5, 11, 12, 17, 18, 21]. The main formulas for β_k are given as:

$$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}} \tag{4}$$

$$\beta_K^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} \tag{5}$$

$$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T g_{k-1}} \tag{6}$$

$$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}} \tag{7}$$

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

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

2. The new formula and its global convergence

Alhawarat et.al [2] presented a new formula with new restart criteria. Here, we modify the formula to match the HS formula. The new modification is termed as OMHR based on the researchers' names (Omar, Mamat, Al-Hawarat & Rivaie). It is given as follows.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

$$\beta_k^{OMHR} = \begin{cases} \dfrac{\|g_k\|^2 - \mu_k |g_k^T g_{k-1}|}{d_{k-1}^T (g_k - g_{k-1})}, & \text{if } \|g_k\|^2 > \mu_k |g_k^T g_{k-1}| \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

where $\|.\|$ represents the Euclidean norm and μ_k is defined by:

$\mu_k = \frac{\|x_k - x_{k-1}\|}{\|y_{k-1}\|}$ where $y_{k-1} = g_k - g_{k-1}$.

We note $0 < \beta_k^{OMHR} < \beta_k^{HS}$.

$$\beta_k^{OMHR} \le 2 \frac{\|g_k\|^2}{d_{k-1}^T (g_k - g_{k-1})}$$

LEMMA 1. $\beta_k^{OMHR} \ge 0$ and

$d_{k-1}^T (g_k - g_{k-1}) > 0$Proven by and $\|g_k\|^2 > \mu_k |g_k^T g_{k-1}|$ then we have

$$\beta_k^{OMHR} = \frac{\|g_k\|^2 - \mu_k |g_k^T g_{k-1}|}{d_{k-1}^T (g_k - g_{k-1})} \le \frac{\|g_k\|^2 + |g_k^T g_{k-1}|}{d_{k-1}^T (g_k - g_{k-1})} \le 2 \frac{\|g_k\|^2}{d_{k-1}^T (g_k - g_{k-1})}$$

3. Numerical analyses of β_k^{OMHR}

Algorithm 1.

Step 1. Initialization. Given x_1, set $k = 1$.

Step 2. If the stopping criteria satisfied then stop.

Step 2. Compute d_k based on (3).

Step 4. Compute α_k based on some line search.

Step 5. Updating new point based on (2).

Step 6. Set $k = k + 1$ and go to Step 2.

$r_{p,s} =$According to Dolan and More [8], benchmark results are created by running a method or a solver (denoted by *S*) on a test problem (denoted by *p*) and taking the information of interest, such as CPU time, number of iterations, or function evaluation. By assuming n_s solvers and n_p problem, for each problem *p* where $p \epsilon P$ and solver *s* where $s \epsilon S$: Computing time (the number of iterations, CPU time or equivalent) required to solve problem p by solver s .To create a baseline for comparison, the performance of solver s on problem p are scaled by the best performance of any solver in *S* on the problem using the ratio $r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s} : s \epsilon S\}}$.

Suppose that a parameter $r_M \ge r_{p,s}$ for all p, s is chosen, and suppose that $r_{p,s} = r_M$ if and only if solver s does not solve problem p . Because we would like to obtain an overall assessment of the performance of a solver, we defined the measure as:

$\rho_s(t) = \frac{1}{n_p}$ size $\{ p \in P : r_{p,s} \le t$.

Thus, $\rho_s(t)$ is the probability for solver $s \in S$ that the performance ratio $r_{p,s}$ is within a factor $t \in R$ of the best possible ratio.

If we define the function p_s as the cumulative distribution function for the performance ratio, then the performance measure $p_s : R \to [0,1]$ for a solver is non-decreasing and piecewise continuous from the right. The value of $p_s(1)$ is the probability that the solver has the best performance of all of the solvers. In general, a solver with high values of $p(t)$, which would appear in the upper right corner of the figure, is preferable. To analyze the efficiency of the new method, we selected some of the standard test functions from CUTEr [12] and Andrei [6] in Table 1.

We performed a comparison with other CG methods, including PRP, FR, a newly introduced Wei, Yao and Liu [20] (WYL) and the new modification (OMHR). The stopping criteria $\|g_k\| \le 10^{-6}$ is used for all algorithms. Even PRP modification is very efficient. However, it does not have global convergence. For more, the reader can see Powell [19]. It is known that PR method is not as efficient as PRP or HS method since $g_{k-1}^T g_{k-1} = g_k^T g_k i.e \beta_k^{FR} = 1$. Figures1, 2, 3 and 4 plots the performance profile of the new OMHR method compared to PRP, FR, and WYL methods. The performance profiles presented in the figures include number of iteration, number function evaluation, number of gradient evaluation, and CPU time evaluation. The vertical axis of both figures shows the percentage of the test problems that were successfully solved by each method. The horizontal axis of the figures shows which percentage of the test problems is the fastest for the methods. Clearly, OMHR records the best performance compared to the other methods. Although PRP performs faster than OMHR, yet OMHR is able to solve 100% of the test problems compared to PRP (84%), FR (90%), and WYL (84.6 %). It was proven that OMHR is more efficient compared to PRP, FR, and WYL.

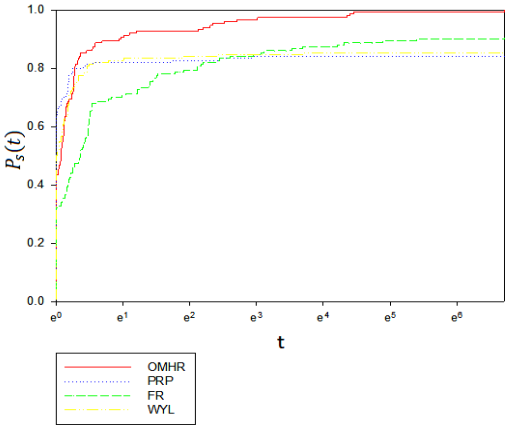


Fig.1: Performance profile based on the number of iteration.

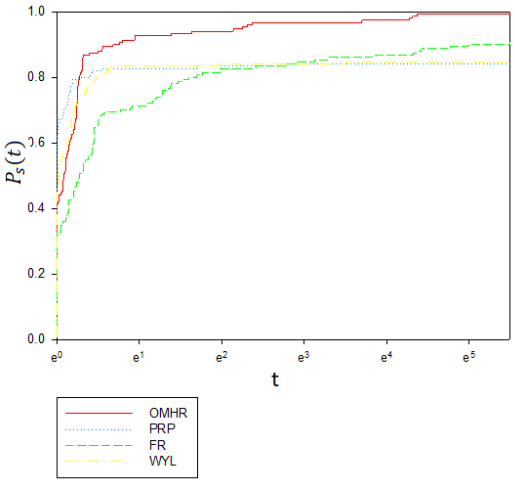


Fig.2: Performance profile based on the number function evaluation

1

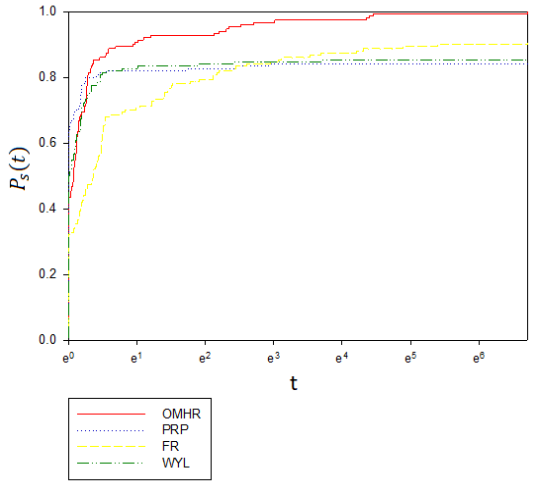


Fig.3: Performance profile based on the number gradient evaluation

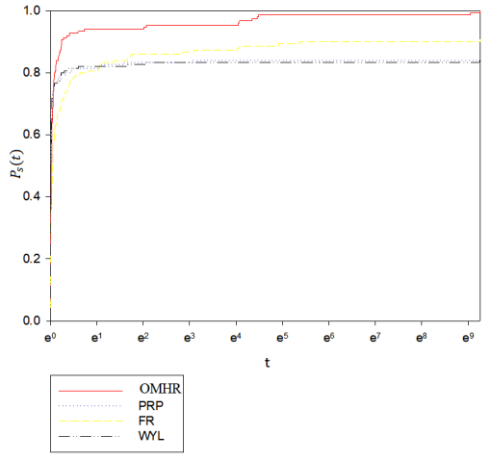


Fig.4: Performance profile based on the CPU time

Table 1: The Standard Test Functions

NO	Function	variables	Initial Points
1	Perturbed Quaratic function	50-100-500-1000-5000	(0.5, 0.5, …,0.5)
2	DENSCHNA (CUTE)	50-100-500-1000	(2, 2, …, 2)
3	Extended Tridiagonal 1 function	50-100-500-1000	(2, 2, …, 2)
4	Diagonal4 function	50-100-500-1000-5000	(1, 1, …,1)
5	Extended Himmelblau function	50-100-500-1000-5000	(1, 1, …,1)
6	Fletchcr function (CUTE)	50-100-500-1000-5000	(0.5, 0.5, …,0.5)
7	NONSCOMP function (CUTE)	50-100-500-1000	(3, 3, …, 3)
8	Extended DENSCHNB function	50-100-500-1000-5000	(1, 1, …,1)
9	Hager function17	2-10-50-100	(1, 1, …,1)
10	BIGGSB1 function (CUTE)	2-50-100-500-1000-5000	(0, 0, …, 0)
11	Shollow function	2-50-100-500-1000-5000	(-2, -2, …, -2)
12	Quadratic QF1 function	2-50-100-500-1000-1500	(0.5, 0.5, …,0.5)
13	Generlized Tridiagonal1 function	2-50-100-500	(2, 2, …, 2)
14	GenerlizedTridiagonal 2	2-100-200	(1, 1, …,1)
15	Quadratic QF1	2-50-100-500-1000-5000	(1, 1, …,1)
16	Extended quadratic penalty QP2	2-50-100-500-1000	(1, 1, …,1)
17	NONDIA (Shanno-78)	2-50-100-500-1000-5000	(-1,-1,…,-1
18	ARWHEAD	2-50-300-600-1200	(1, 1, …,1)
19	DQDRTIC	9-100-500-1000-5000	(3, 3, …, 3)
20	Extended Himmelblau Function	2-50-100-150	(1, 1, …,1)
21	Extended Block Diagonal BD1	2-50-100-500-1000-5000	(0.1, 0.1, …, 0.1)
22	Raydan 1 Function	2-45-50-100	(1, 1, …,1)
23	Extended Three Exponential Terms	2-50-100-500-1000	(0.1, 0.1, …, 0.1)
24	Generalized Tridiagonal-1 Function	2-50-100-500	(0.5, 0.5, …,0.5)
25	Power function	2-50-100-500-1000	(1, 1, …,1)
26	DIXON3DQ (CUTE)	2-50-100-500-1000	(-1, -1, …,-1)
27	SINQUAD (CUTE)	2-50-100-500-1000	(0.1, 0.1, …, 0.1)
28	A Quadratic Function QF2	2-50-100-500-1000	(0.5, 0.5, …,0.5)
29	Extended Trigonometric Function	2-50-100-500-1000	(0.2,0.2, …, 0.2)
30	EG2 (CUTE)	2-50-100-500	(1, 1, …,1)

4. Conclusions

In this paper, we proposed a new modification of CG method.The new method (OMHR) is nonnegative, and the numerical results, based on number of iterations and CPU times, showed that OMHR formula performs better than the other CG methods(PRP, FR, &WYL).

Acknowledgement

The authors would like to thank university Sultan Zainal Abidin (UNIZA).

References

[1] Al-Baali, M., "Descent property and global convergence of the Fletcher—Reeves method with inexact line search", *IMA Journal of Numerical Analysis*, Vol.5, No.1, (1985), pp.121-124.

[2] Alhawarat, A., Zabidin, S., Mustafa, M., & Mohd R., "An efficient modified Polak–Ribière–Polyak conjugate gradient method with global convergence properties", *Optimization Methods and Software*, Vol. 32, No.6, (2017), pp. 1299-1312.

[3] Alhawarat, A., Zabidin, S., "Modification of nonlinear conjugate gradient method with weak Wolfe-Powell line search", *Abstract and Applied Analysis, Hindawi Publishing Corporation*, Vol.2017, (2017).

[4] Alhawarat, A., Mustafa, M., Mohd R., & Zabidin, S., "An efficient hybrid conjugate gradient method with the strong Wolfe-Powell line search", *Mathematical Problems in Engineering* , Vol.2015, (2015).

[5] Alhawarat, A., Mustafa, M., Mohd R., & Ismail, M., "A new modification of nonlinear conjugate gradient coefficients with global convergence properties", *International Journal of Mathematical, Computational, Statistical, Natural and Physical Engineering*, Vol.8, No.1, (2014), pp. 54-60.

[6] Andrei, N., "An unconstrained optimization test functions collection", *Adv. Model. Optim*, Vol.10, No.1, (2008), pp. 147-161.

[7] Dai, Y., Yaxiang, Y., "A nonlinear conjugate gradient method with a strong global convergence property", *SIAM Journal on Optimization*, Vol.10, No.1, (1999), pp.177-182.

[8] Dolan, E. D., Jorge, J. M., "Benchmarking optimization software

with performance profiles", *Mathematical programming*, Vol.91, No.2, (2002), pp. 201-213.
[9] Fletcher, R., Colin, M. R., "Function minimization by conjugate gradients", *The Computer Journal*, Vol.7, No.2, (1964), PP.149-154.
[10] Fletcher, R., *Practical methods of optimization*, John Wiley & Sons, (2013).
[11] Gilbert, J., Jean, C., & Jorge, N., "Global convergence properties of conjugate gradient methods for optimization", *SIAM Journal on optimization*, Vol.2, No.1, (1992), pp.21-42.
[12] Gould, N. I. M., Dominique, O., & Philippe, L. T., "CUTEr and SifDec: A constrained and unconstrained testing environment, revisited", *ACM Transactions on Mathematical Software (TOMS)*, Vol.29, No.4, (2003), pp.373-394.
[13] Hestenes, M. R., Eduard, S., *Methods of conjugate gradients for solving linear systems*, NBS, (1952).
[14] Liu, Y., Storey, C., "Efficient generalized conjugate gradient algorithms, part 1: theory", *Journal of Optimization Theory and Applications*, Vol.69, No.1, (1991), pp.129-137.
[15] Polak, E., Gerard, R., "Note sur la convergence de méthodes de directions conjuguées." Revue françaised' informatiqueet de recherché pérationnelle", *Série rouge*, Vol.3, No.16, (1969), pp.35-43.
[16] Polyak, B. T., "The conjugate gradient method in extremal problems", *USSR Computational Mathematics and Mathematical Physics*, Vol.9, No.4, (1969), pp.94-112.
[17] Powell, M. J. D., "Restart procedures for the conjugate gradient method", *Mathematical programming*, Vol.12, No.1, (1977), pp.241-254.
[18] Salleh, Z., Ahmad, A., "An efficient modification of the Hestenes-Stiefel nonlinear conjugate gradient method with restart property", *Journal of Inequalities and Applications*, Vol.2016, No.1, (2016), pp.110.
[19] Wolfe, P., "Convergence conditions for ascent methods", *SIAM review*, Vol.11, No.2, (1969), pp.226-235.
[20] Wei, Z., Guoyin, L., & Liqun, Q., "New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems", *Applied Mathematics and Computation*, Vol.179, No.2, (2006), pp. 407-430.
[21] Zhang, Y., Hao, Z., & Chuanlin, Z., "Global convergence of a modified PRP conjugate gradient method", *Procedia Engineering*, Vol.31, (2012), pp.986-995.