



ISSN: 0067-2904

A Descent Modification of Conjugate Gradient Method for Optimization Models

Ibrahim Mohammed Sulaiman¹, Mustafa Mamat¹, Kamilu Kamfa², Muktar Danlami³

¹Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia

²Department of Mathematical Science, Faculty of Computing and Mathematics, Kano University of Science and Technology, Wudil, Nigeria.

³Department of Mathematics, Yusuf Maitama Sule University, Kano, Nigeria

Received: 21/8/2019

Accepted: 15/3/2020

Abstract

In this paper, we suggest a descent modification of the conjugate gradient method which converges globally provided that the exact minimization condition is satisfied. Preliminary numerical experiments on some benchmark problems show that the method is efficient and promising.

Keywords: Conjugate gradient parameter, exact minimization rule, Unconstrained Optimization, Sufficient descent property

Introduction

The conjugate gradient algorithms are among the most efficient algorithms because of their simplicity, convergence properties, and competence to solve large-scale unconstrained optimization problems. Consider the unconstrained optimization problem

$$\min f(x), x \in R^n \quad (1)$$

where $f: R^n \rightarrow R$ is smooth whose gradient $g(x)$ is available. The conjugate gradient (CG) methods are among the most preferred methods for solving problems (1). Starting with an initial point x_0 , the CG method generates a sequence of iterates $\{x_k\}$ through the following scheme [1].

$$x_{k+1} = x_k + \alpha_k d_k \quad k \geq 0 \quad (2)$$

where $\alpha_k \in R$ is the step size computed along the search direction d_k . The first direction of search is usually the negative of the gradient which is the steepest descent direction, i.e., $d_0 = -g_0$, while subsequent directions are recursively defined as follows.

$$d_k = -g_k + \beta_k d_{k-1} \quad (3)$$

in which β_k is known as the CG update parameter capable of reducing to linear CG method if the step size satisfies the exact minimization condition and (1) is a quadratic function that is strictly convex. The performance of these CG methods differs for general non-quadratic objective functions [2, 3]. Some of the efficient conjugate gradient coefficients are the Hestenes-Stiefel, (HS) [4], Polak-Ribiere-Polyak (PRP) [5, 6], and Liu and Storey (LS) [7] with their formula defined below.

$$\beta_k^{HS} = \frac{g_k^T y_k}{d_{k-1}^T y_k}, \quad \beta_k^{PRP} = \frac{g_k^T y_k}{\|g_{k-1}\|^2}, \quad \beta_k^{LS} = \frac{g_k^T y_k}{-d_{k-1}^T g_{k-1}}$$

where $y_k = g_k - g_{k-1}$ and $\|\cdot\|$ is known as the Euclidean norm.

The methods of HS, PRP, and LS have the same numerator and can perform a restart when the algorithm moves along d_{k-1} with a very small step size, that is, $x_{k+1} \approx x_k$ implying $g_k - g_{k-1} \approx 0$, and thus, produce effective numerical results. Their global convergence has been studied by numerous

*Email: sulaimanib@unisza.edu.my

researchers. For the convergence of various CG methods, it is usually required that α_k satisfies the exact minimization condition: $f(x_k + \alpha_k d_k) = \min f(x_k + \alpha d_k)$ $\alpha > 0$. (4)

On the other hand, other researchers require α_k satisfying the inexact line search such as the standard Wolfe condition:

$$f(x_k + \alpha_k d_k) \leq f(x_k) - \delta \alpha_k g_k^T d_k, \quad (5)$$

$$g(x_k + \alpha_k d_k)^T d_k > \sigma g_k^T d_k. \quad (6)$$

The convergence results of some unconstrained optimizations methods are obtained under the strong Wolfe line search (5) and

$$|g(x_k + \alpha_k d_k)^T d_k| \leq |\sigma g_k^T d_k|. \quad (7)$$

One interesting features of these methods is that, if α_k is the exact minimizer, then HS = PRP = LS. But, Powell [8] provided a counter-example which shows that there exist some nonconvex functions in which PRP does not converge, which also applies to the HS method. Despite the effective computational results of these methods, their convergence is yet to be established under some inexact line searches. Various modifications have received a growing interest around the globe.

Recently, Rivaie *et al.* [9] constructed a new denominator while retaining the numerator of the HS, PRP, and LS method, as follows;

$$\beta_k^{RMIL} = \frac{g_k^T y_k}{d_{k-1}^T (d_{k-1} - g_k)}. \quad (8)$$

The global convergence was established under exact line search. The RMIL method was extended by Rivaie *et al.* [1], as follows;

$$\beta_k^{RMIL*} = \frac{g_k^T y_k}{\|d_{k-1}\|^2}. \quad (9)$$

It is clear to see that $0 \leq \beta_k^{RMIL*} \leq \beta_k^{PRP}$. The RMIL* and PRP methods have similar features and thus possess the restart properties. Also, RMIL reduces to RMIL* if the exact minimization condition is satisfied. The global convergence of the RMIL AND RMIL* has been studied under exact line searches. Convergence analysis of recent modifications of the CG methods can be referred to form literature [10, 11, 12]. Also, application of the CG method to real-life problems can be referred to [13, 14].

New formula and its properties

In an attempt to enhance the performance of the CG methods while retaining the nice convergence properties, various researches have proposed many variants of the CG coefficient and established their global convergence proof. However, some of the recent modifications of the CG methods are too complicated and thus, their proofs are difficult to establish. Motivated by the descent properties and numerical efficiency of the RMIL and RMIL* methods, we proposed a simple variant of RMIL AND RMIL* as follows.

$$\beta_k^{MIMS} = \frac{g_k^T y_k}{d_{k-1}^T (d_{k-1} - g_k)} + \frac{g_k^T y_k}{2 \|d_{k-1}\|^2}. \quad (10)$$

where MIMS denotes the researchers name Mamat, Ibrahim Mohammed Sulaiman. The following algorithm describes the proposed MIMS method. This method inherits some nice properties of the RMIL and RMIL* method with better numerical performance. Another interesting feature of the proposed MIMS method is the ability to reduce to the standard RMIL* method provided that the exact minimization condition is satisfied. The algorithm of the proposed MIMS method is described as follows.

Algorithm 2.1:

Step 1. Given $x_0 \in R^n$, $\varepsilon \geq 0$, set $d_k = -g_k$ for $k = 0$. If $\|g_k\| \leq 10^{-6}$, then stop.

Step 2. Compute α_k by (4), (5 & 6), or (5 & 7).

Step 3. Let $x_{k+1} = x_k + \alpha_k d_k$ and check if $\|g_{k+1}\| \leq \varepsilon$, then stop.

Step 4. Compute β_k^{MIMS} by (10) and obtain the next d_{k+1} by (3).

Step 5. Update x_{k+1} by (2).

Step 6. Repeat Steps 2 to 4 with $k := k + 1$ until tolerance is satisfied.

For the convergence analysis, the following assumptions are frequently needed.

Assumption A. The function $f(x)$ is bounded below on the level set $\Omega = \{x \in R^n : f(x) \leq f(x_1)\}$.

Assumption B. In some neighbourhood \mathcal{C} of Ω , f is smooth and its gradient $\nabla f(x) = g$ is Lipschitz continuous, that is, $\forall x, y \in \mathcal{C}, \exists L > 0$ such that $\|g(x) - g(y)\| \leq L\|x - y\|$ (11)

Convergence Analysis

For the convergence analysis, we need to simplify the proposed method. From Rivaie et al., RMIL coefficient [9],

$$\beta_k^{RMIL} = \frac{g_k^T y_k}{d_{k-1}^T (d_{k-1} - g_k)} = \frac{\|g_k\|^2 - g_k^T g_{k-1}}{\|d_{k-1}\|^2 - d_{k-1}^T g_k}$$

Also, from RMIL*, it follows that;

$$\beta_k^{RMIL*} = \frac{g_k^T y_k}{\|d_{k-1}\|^2} = \frac{\|g_k\|^2 - g_k^T g_{k-1}}{\|d_{k-1}\|^2}$$

From (10), we have

$$\begin{aligned} \beta_k^{MIMS} &= \frac{\frac{g_k^T y_k}{d_{k-1}^T (d_{k-1} - g_k)} + \frac{g_k^T y_k}{\|d_{k-1}\|^2}}{2} = \frac{\frac{\|g_k\|^2 - g_k^T g_{k-1}}{\|d_{k-1}\|^2 - d_{k-1}^T g_k} + \frac{\|g_k\|^2 - g_k^T g_{k-1}}{\|d_{k-1}\|^2}}{2} \\ &\leq \frac{\|g_k\|^2}{\|d_{k-1}\|^2} \end{aligned} \tag{12}$$

Hence, we have

$$0 \leq \beta_k^{MIMS} \leq \frac{\|g_k\|^2}{\|d_{k-1}\|^2} \tag{13}$$

We begin by showing that the method is descent.

3.1 Sufficient descent condition

For this to hold, then $g_k^T d_k \leq -C\|g_k\|^2, \forall k \geq 0, C > 0$. (14)

Theorem 1 below would be used to show that (14) holds for the proposed MIMS method under (4).

Theorem 1. Consider a CG method with search direction (3) and CG coefficient (10), then condition (14) holds for all $k \geq 0$.

Proof. The proof of this theorem is by induction. If $k = 0$, then $g_0^T d_0 \leq -C\|g_0\|^2$. Hence, condition (14) holds true. We need to show that (14) holds for all $k \geq 1$. Multiplying through (3) by g_k^T will give;

$$g_k^T d_k = g_k^T (-g_k^T + \beta_k d_{k-1}) = -\|g_k\|^2 + \beta_k g_k^T d_{k-1} \tag{15}$$

However, $g_k^T d_{k-1} = 0$ under exact minimization rule [1, 2]. Hence,

$$g_k^T d_k = -\|g_k\|^2$$

Therefore, (14) holds true for $k + 1$ and this completes the proof. ■

3.2. Global convergence properties

The following lemma, which follows from the assumptions above, is useful in the convergence analysis of the CG methods.

Lemma 1. [15]. Suppose that Assumptions A and B hold true. Consider any CG methods of the form (2) and (3) where d_k satisfies

$$g_k^T d_k < 0 \tag{16}$$

and α_k is obtained using the exact minimization rules (4). Then,

$$\sum_{k=0}^{\infty} \frac{(g_{k-1}^T d_{k-1})^2}{\|d_{k-1}\|^2} < \infty \text{ or } \lim_{k \rightarrow \infty} \|g_{k-1}\| = 0. \tag{17}$$

Proof: The proof follows from [9]. By contradiction, we suppose that (17) is not true and that $\exists c > 0$ such that $\|g_{k-1}\| \geq c$.

From (3) and (10), we have

$$d_k + g_k = \beta_k^{MIMS} d_{k-1} \tag{18}$$

Squaring (18) will give

$$\|d_k\|^2 = (\beta_k^{MIMS})^2 \|d_{k-1}\|^2 - 2g_k^T d_k - \|g_k\|^2 \tag{19}$$

Dividing (16) by $(g_k^T d_k)^2$ gives

$$\begin{aligned} \frac{\|d_k\|^2}{(g_k^T d_k)^2} &= \frac{(\beta_k^{MIMS})^2 \|d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{2}{g_k^T d_k} - \frac{\|g_k\|^2}{(g_k^T d_k)^2} \\ &= \frac{(\beta_k^{MIMS})^2 \|d_{k-1}\|^2}{(g_k^T d_k)^2} - \left(\frac{1}{\|g_k\|} + \frac{\|g_k\|}{g_k^T d_k} \right) + \frac{1}{\|g_k\|^2} \\ &\leq \frac{(\beta_k^{MIMS})^2 \|d_{k-1}\|^2}{(g_k^T d_k)^2} + \frac{1}{\|g_k\|^2} \end{aligned}$$

From (13), we have

$$\left(\frac{\|g_k\|^2}{\|d_{k-1}\|^2}\right)^2 \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{1}{\|g_k\|^2} = \frac{\|g_k\|^2}{\|d_{k-1}\|^2 \|d_k\|^2} + \frac{1}{\|g_k\|^2} \leq \frac{1}{\|d_{k-1}\|^2} + \frac{1}{\|g_k\|^2} \tag{20}$$

By definition, $\frac{1}{\|d_0\|^2} = \frac{1}{\|g_0\|^2}$, thus, from (20), it follows that;

$$\frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} = \frac{1}{\|g_0\|^2} + \frac{1}{\|g_k\|^2}$$

which implies

$$\frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} \leq \sum_{i=0}^{\infty} \frac{1}{\|g_{i-1}\|^2} \tag{21}$$

$$\frac{(g_{k-1}^T d_{k-1})^2}{\|d_{k-1}\|^2} \geq \frac{c^2}{k}$$

Since $\|g_{k-1}\| \geq c$, then from (21), we have

$$\sum_{k=0}^{\infty} \frac{(g_{k-1}^T d_{k-1})^2}{\|d_{k-1}\|^2} = \infty$$

which contradicts the assertion in Lemma 1, and thus, the proof is completed. ■

Numerical experiment

In this section, we investigate numerical performance of the proposed MIMS conjugate gradient method on some unconstrained optimization benchmark problems considered from Andrei [16] and Molga & Smutnicki [17], as listed in Table 1. The performance result was compared with the existing CG methods of RMIL and RMIL*, based on iteration number and CPU time under exact minimization conditions defined in (4). For each benchmark problem, four different initial guesses with a varying dimension (n) are selected, ranging from points closer to the solution to points further away, as suggested [18]. The stopping condition was set as $\|g_{k+1}\| \leq \epsilon$, where $\epsilon = 10^4$. Meanwhile, the iteration is also terminated if the function evaluation exceeds 10000, or the number of iterations exceeds 1000. All test problems were coded on MATLAB version 7 (2015a) subroutine programming and solved on an Intel® corei5-2410M CPU @ 2.30 GHz processor, 4GB for RAM operating system. Also, the performance result was plotted using the Dolan and More [19] performance profile as shown in Figures-1 and 2, respectively.

Table 1- List of Test Functions

No	Function	(n)	Initial Guess
1	Treccani	2	(2,2), (9,9), (10,10), (13,13)
2	DQDRTIC	2	(2,2), (9,9), (10,10), (13,13)
3	Three Hump Camel	2	(0.5,0.5), (5,5), (10,10), (15,15)
4	Booth	2	(2,2), (9,9), (10,10), (13,13)
5	Ext DENCHNB	2,4	(5,5), (10,10), (20,20), (50,50)
6	Sphere	2,4,10,100	(2,2, ..., 2), (9,9, ..., 9), (10,10, ..., 10), (15,15, ..., 15)
7	Ext White and Holst	2,4,10,100,1000	(0,0, ..., 0), (2,2, ..., 2), (10,10, ..., 10), (13,13, ..., 13)
8	Gen Tridiagonal	2,4,10,100,1000	(0,0, ..., 0), (2,2, ..., 2), (6,6, ..., 6), (9,9, ..., 9)
9	Diagonal 4	2,4,10,100,1000	(2,2, ..., 2), (9,9, ..., 9), (10,10, ..., 10), (15,15, ..., 15)
10	Ext Tridiagonal 1	2,4,10,100,1000	(2,2, ..., 2), (9,9, ..., 9), (10,10, ..., 10), (15,15, ..., 15)
11	Ext Rosenbrock	2,4,10,100,1000	(0,0, ..., 0), (2,2, ..., 2), (10,10, ..., 10), (15,15, ..., 15)
12	Fletcher	2,4,10,100,1000	(2,2, ..., 2), (3,3, ..., 3), (9,9, ..., 9), (10,10, ..., 10)
13	NONSCOMP	2,4,10,100,1000	(2,2, ..., 2), (9,9, ..., 9), (10,10, ..., 10), (15,15, ..., 15)
14	Ext Quadratic QP2	2,4,10,100,1000	(2,2, ..., 2), (9,9, ..., 9), (10,10, ..., 10), (15,15, ..., 15)

15	Gen Quartic	2,4,10,100,1000	(2,2, ..., 2), (9,9, ..., 9), (10,10, ..., 10), (15,15, ..., 15)
----	-------------	-----------------	------------------------------------------------------------------

Figure-1 illustrates the performance comparison based on iteration number. It is obvious that the proposed MIMS method outperformed both methods of RMIL and RMIL*. Also, Figure- 2 shows that the MIMS method is preferable to the methods of RMIL and RMIL* based on CPU time.

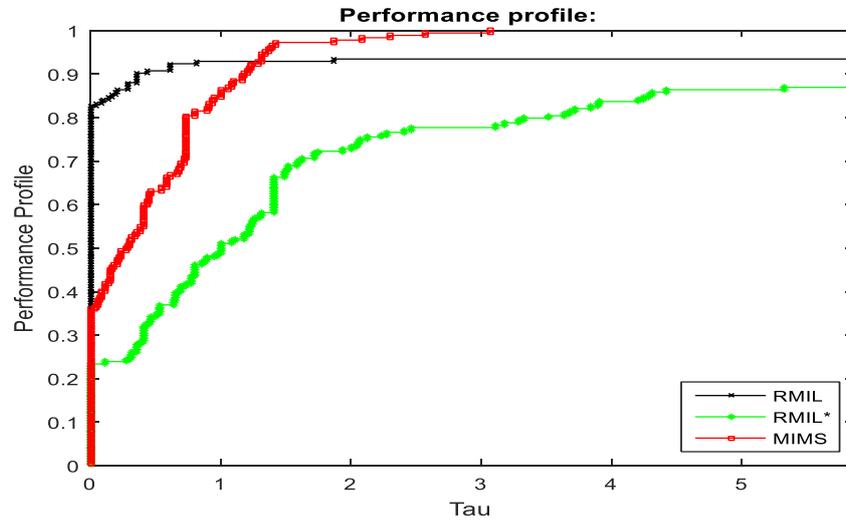


Figure 1 – Performance profile based on number of iterations

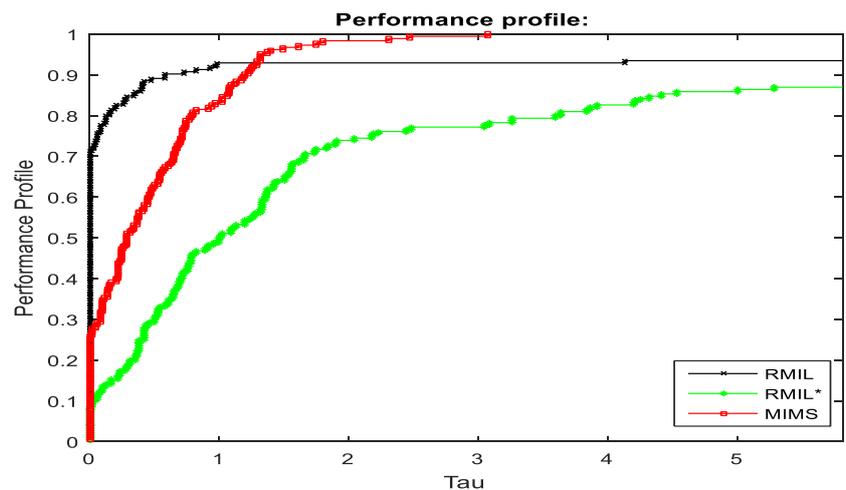


Figure 2 – Performance profile based on CPU time

As a final note, we can conclude that the proposed MIMS method has the best performance based on the performance profile illustrated in Fig 1 and Fig 2 above, since it can solve all of the test problems successfully.

Conclusions

In this paper, the proposed modification of RMIL methods guaranteed the descent condition and converged globally, provided that the exact minimization condition is satisfied. Also, the MIMS method inherited the restart mechanism of the RMIL* method with a better numerical performance. The proposed method was compared with RMIL and RMIL* conjugate gradient methods under exact line search. Numerical results illustrates the efficiency of the MIMS method. For further work, researchers interested in the area of conjugate gradient method can test this β_k^{MIMS} coefficient using the inexact line search.

Acknowledgements

The authors would like to extend their appreciation to the reviewers for their kind observations and suggestions, and also to Sultan Zainal Abidin University, Malaysia.

References

1. Rivaie, M., Abashar, A., Mamat, M. and Ismail, M. **2014**. The convergence properties of a new type of conjugate gradient methods, *Appl. Math. Sci.* **8**: 33-44.
2. Andrei, N. **2007**. Numerical comparison of conjugate gradient algorithms for unconstrained optimization. *Stud. Inform. Control.* **16**(4): 333-352.
3. Sulaiman, I.M. Mamat, M., Abashar, A., Rivaie, M., Mohamad, F.S. and Salleh, Z. **2015**. The global convergence properties of an improved conjugate gradient method. *Applied Mathematical Science*, **9**(38): 1857-1868.
4. Stiefel, E. and Hestenes, M.R. **1952**. Method of conjugate gradient for solving linear equation, *J. Res. Nat. Bur. Stand.*, **49**: 409-436.
5. Polak, E. and Ribière, G. **1969**. Note sur la convergence de directions conjuguée, *Rev. Francaise Informat Recherche Operationelle*, 3e Année. **16**: 35-43.
6. Polyak, B.T. **1969**. The conjugate gradient method in extreme problems, *USSR Comp. Math. Phys.*, **9**: 94-112.
7. Liu, Y and Storey, C. **1992**. Efficient generalized conjugate gradient algorithms. Part 1. *theory. J. Optim. Theory. Appl.* **10**: 129-137.
8. Powell, M.J.D. **1986**. Convergence properties of algorithm for nonlinear optimization, *SIAM Rev.*, **28**: 487-500.
9. Rivaie, M., Mamat, M., Leong, W.J. and Ismail, M. **2012**. A New conjugate gradient coefficient for large-scale nonlinear unconstrained optimization, *Int. Journal of Math. Analysis.* **6**(23): 1131-1146.
10. Sulaiman, I.M. Mamat, M., Abashar, A., Zabidin, S. **2015**. A Modified Nonlinear Conjugate Gradient Method for Unconstrained Optimization, *Applied Mathematical Sciences*, **9**(54): 2671 – 2682.
11. Abashar, A., Mamat, M., Rivaie, M., and Ismail, M. **2014**. Global convergence properties of a new class of conjugate gradient method for unconstrained optimization, *Applied Mathematics and Computation*, **8**(1): 3307-3319.
12. Mamat, M., Sulaiman, I.M., Maulana, M., Sukono, and Zakaria, Z.A. **2020**. An efficient spectral conjugate gradient parameter with descent condition for unconstrained optimization. *Journal of Advanced Research in Dynamical and Control Systems.* **12**(02): 2487-2493.
13. AliF, H. and Jawad, A. A. **2020**. Minimizing the Total Completion Time and Total Earliness Time Functions for a Machine Scheduling Problem Using Local Search Methods. *Iraqi Journal of Science*, 126-133. <https://doi.org/10.24996/ij.s.2020.SI.1.17>.
14. Sulaiman, I.M., Yakubu, A.U., and Mamat, M. 2020. Application of Spectral Conjugate Gradient Methods for Solving Unconstrained Optimization Problems. *An International Journal of Optimization and Control: Theories & Applications.* **10**(2):198-205.
15. Zoutendijk, G. **1970**. Nonlinear programming computational methods, in: Abadie J. (Ed.) *Integer and nonlinear programming*, North Holland, Amsterdam.
16. Andrei, N. **2008**. An unconstrained optimization test functions collection, *Advanced Modelling and optimization*, **10**(1): 147-161.
17. Molga, M. and Smutnicki, C. **2005**. Test Functions for optimization needs Available: <http://www.zsd.ict.pwr.wroc.pl/files/docs/functions.pdf>.
18. Hilstrom, K.E. **1977**. A simulation test approach to the evaluation of nonlinear optimization algorithms, *A.C.M. Trans. Maths. Softw.* **3**(4): 305-315.
19. Dolan, E. and Moré, J.J. **2002**. Benchmarking optimization software with performance profiles. *Mathematical Programming.* **91**(2): 201-213.