The Scientific Journal of King Faisal University

Basic and Applied Sciences

Global Convergence of Nonlinear Conjugate Gradient Coefficients with Inexact Line Search

Awad Abdelrahman [1], Osman O. O. Yousif [1], Mogtaba Mhammed [2] and Murtada K. Elbashir [3]

[1] Department of Mathematics, Faculty of Mathematical and Computer Sciences, University of Gezira, Wad Madani, Sudan
[2] Department of Mathematics, College of Science, Majmaah University, Zulfi, Saudi Arabia
[3] Department of Information Systems, College of Computer and Information Sciences, Jouf University, Sakaka, Saudi Arabia

LINK	RECEIVED	ACCEPTED	PUBLISHED ONLINE	ASSIGNED TO AN ISSUE
https://doi.org/10.37575/b/sci/210058	26/09/2021	26/11/2021	26/11/2021	01/12/2021
NO. OF WORDS	**NO. OF PAGES**	**YEAR**	**VOLUME**	**ISSUE**
3796	6	2021	22	2

ABSTRACT

Nonlinear conjugate gradient (CG) methods are significant for solving large-scale, unconstrained optimization problems, providing vital knowledge to determine the minimum point or optimize the objective functions. Many studies of modifications for nonlinear CG methods have been carried out to improve the performance of numerical computation and to establish global convergence properties. One of these studies is the modified CG method, which has been proposed by Rivaie *et al.* (2015). In this paper, we modify their work in such a way that one can obtain efficient numerical performance and global convergence properties. Due to the widespread use of the strong Wolfe line search in practice, our proposed modified method implemented its use. At the same time, to show the performance of the modified method in practice, a numerical experiment is performed.

KEYWORDS

Unconstrained optimization; conjugate gradient method; sufficient descent property; global convergence

CITATION

Abdelrahman, A., Yousif, O.O.O., Mhammed, M. and Elbashir, M.K. (2021). Global convergence of nonlinear conjugate gradient coefficients with inexact line search. *The Scientific Journal of King Faisal University: Basic and Applied Sciences*, **22**(2), 86–91. DOI: 10.37575/b/sci/210058

1. Introduction

Due to their global convergence and low memory requirements, conjugate gradient methods are widely used for solving unconstrained optimization problems. The unconstrained optimization problems can be formulated as follows:

$$\min_{x\in R^n} f(x), \qquad (1.1)$$

where $f: R^n \to R$ is nonlinear, continuously differentiable, and its gradient is denoted by $g(x)$, which should be available, when applied to solve the Problem (1.1), starting from an initial point $x_0 \in R^n$, and follows the iteration formula:

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0,1,2 \ldots, \qquad (1.2)$$

where $\alpha_k > 0$ is a step-size. The step-size is determined by a line search, and d_k is the search direction defined by:

$$d_k = \begin{cases} -g_k & \text{if } k = 0, \\ -g_k + \beta_k d_{k-1} & \text{if } k \geq 1, \end{cases} \qquad (1.3)$$

where $g_k = \nabla f(x_k)$ is the gradient vector of the function f at x_k and g_k^T is the transpose of g_k. The value β_k is a scalar known as the CG coefficient. The well-known classical CG methods formulas for β_k are:

$$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}}, \qquad (1.4)$$

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}, \qquad (1.5)$$

$$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T g_{k-1}}, \qquad (1.6)$$

$$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}, \qquad (1.7)$$

$$\beta_k^{LS} = -\frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T g_{k-1}}, \qquad (1.8)$$

$$\beta_k^{DY} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}}. \qquad (1.9)$$

To analyze the convergence, the CG methods are implemented under exact and inexact line searches. If the line search is exact, the step length α_k is obtained in the direction d_k by the rule:

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(x_k + \alpha d_k). \qquad (1.10)$$

In the inexact line search, α_k can be obtained using the strong Wolfe line search, in which the following conditions satisfy:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k, \qquad (1.11)$$

$$|g(x_k + \alpha_k d_k)^T d_k| \leq -\sigma g_k^T d_k, \qquad (1.12)$$

where $0 < \delta < \sigma < 1$.

To prove the global convergence of CG methods, the following sufficient descent condition is always required:

:

$$g_k^T d_k \leq -c\|g_k\|^2, \quad \text{for } k \geq 0, \text{ where } c > 0 \qquad (1.13)$$

The performance and behavior of different conjugate gradient methods for general non-quadratic functions with an inexact line search correspond to different choices for the important coefficient β_k. The global convergence properties of CG methods are essential properties for studying the coefficient β_k and obtaining good numerical performance. The CG methods were categorized by Andrei (2011) into three different methods: the classical CG method, the scaled CG method, and the hybrid and parameterized CG method. Al-Baali (1985), Touati-Ahmed and Storey (1990), and Gilbert and Nocedal (1992) used the inexact line search with a strong Wolfe condition in order to analyze the global convergence properties of FR(1.5) is known as Fletcher and Reeves (1964) and PRP(1.6) is known as Polak–Ribiere–Polyak (1969) methods. Al-Baali (1985) proved the global convergence of the FR method if the strong Wolfe line search is used and the parameter σ is restricted in $(0, \frac{1}{2})$. In addition, Guanghui *et al.* (1995) extended Al-Baali's (1985) result with the case that $\sigma = \frac{1}{2}$. The PRP and HS(1.4) is known as Hestenes and Steifel (1952) methods have excellent performance in practical computation, due to the inclusion of an approximate restart feature when jamming occurs, but their convergence properties are not perfect (Jiang *et al.*, 2018). Nevertheless, both perform better than FR in practical computations. The study of the PRP method has an efficient approach and has made significant improvements. Gilbert and Nocedal (1992) proceeded with classificatory analysis and concluded that the PRP method is

Corresponding Author: Awad Abdelrahman Abdalla Mohammed 00249962255448, awad.abdalla26@yahoo.com

globally convergent if β_k^{PRP} is constrained to be non-negative and α_k is determined by a line search step satisfying the sufficient descent condition $g_k^T d_k \leq -c\,\|g_k\|^2$.

Many studies have been carried out to establish global convergence and obtain competitive numerical results by proposing new methods or modifying existing methods.

Rivaie *et al.* (2012) have proposed a new coefficient denoted by β_k^{RMIL}, as follows:

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

which is globally convergent and has good performance when it is applied under exact line search. Dai (2016) made a simple modification to Rivaie *et al.* (2012), and, on this basis, Yousif (2020) established the sufficient descent conditions and global convergence using the strong Wolfe line search. Additionally, Rivaie *et al.* (2015) proposed a new class of nonlinear conjugate gradient coefficients with exact and inexact line searches and named this the RMIL+ method. The coefficient β_{k+1} in RMIL+ method is defined as follows:

$$\beta_{k+1}^{RMIL+} = \frac{g_{k+1}^T(g_{k+1} - g_k - d_k)}{\|d_k\|^2}, \qquad (1.14)$$

In this paper, in order to obtain better numerical performance than the methods in Rivaie *et al.* (2012), Rivaie *et al.* (2015), and Yousif (2020), we performed a modification on the method in Rivaie *et al.* (2015). In Section 2, we present a modified CG method and algorithm. The descent property and the global convergence under the strong Wolfe line search are described in Section 3. In Section 4, we present preliminary numerical results and discussion. Lastly, in Section 5, we present a conclusion.

2. Simple Modification of the RMIL+ Method

In this section, we modify the CG method in Rivaie *et al.* (2015), known as the RMIL+ method, to obtain better results. We refer to the modified method by MRMIL+, and its coefficient is defined as follows:

$$\beta_{k+1}^{MRMIL+} = \max\{0\,, \beta_{k+1}^{RMIL+}\}. \qquad (2.1)$$

Clearly, from (2.1), the suggested coefficient satisfies the following two inequalities:

$$\beta_{k+1}^{MRMIL+} \leq \frac{\|g_k\|^2}{\|d_k\|^2}\ , \text{if } \frac{g_{k+1}^T g_k + g_{k+1}^T d_k}{\|d_k\|^2} \geq 0\ . \qquad (2.2)$$

$$\beta_{k+1}^{MRMIL+} \leq m\frac{\|g_k\|^2}{\|d_k\|^2}, \text{if } \frac{g_{k+1}^T g_k + g_{k+1}^T d_k}{\|d_k\|^2} < 0\ , \text{where } m > 1. \qquad (2.3)$$

To distinguish between RMIL+ in Rivaie *et al.* (2015), and RMIL+ in Yousif (2020), we use MRMIL instead of RMIL+ as used in Rivaie *et al.* (2015).

By defining a new coefficient in (2.1), we can define the following new algorithm:

Algorithm 1

- **Step 1**: Initialization given x_0 , set $k = 0$.
- **Step 2**: Compute β_{k+1} based on (2.1).
- **Step 3**: Compute d_{k+1} based on (1.3). If $g_k = 0$, then stop.
- **Step 4**: Compute α_k based on a strong Wolfe line search (1.11), (1.12).
- **Step 5**: Updating new point based on (1.2).
- **Step 6**: Convergent test and terminal criteria.

If $f(x_{k+1}) \leq f(x_k)$ and $\|g_k\| \leq \epsilon$, then stop.

Otherwise go to Step 1 with $k = k + 1$.

3. Convergence Analysis

In this section, using the strong Wolfe line search, we will show the convergence properties of MRMIL +; essentially, the convergence properties of Algorithm 1. Since global convergence involves satisfying the sufficient descent condition, we first prove sufficient descent. The following basic assumption on the objective function f is always needed in the analysis of CG methods' global convergence properties, and Lemma 1 is needed to show the sufficient descent condition and global convergence properties of the MRMIL + method.

Assumption 1

(i) F(x) is bounded below on the level set on R^n and is continuously differentiable in a neighborhood N of the level set $\Gamma = \{x \in R^n |\ f(x) \leq f(x_0)\}$ at the initial point x_0, there exists a constant $B > 0$ such that:

$$\|x - y\| \leq B\,, \quad \forall\, x, y \in N \qquad (3.1)$$

(ii) The gradient $g(x) = \nabla f(x)$ is Lipschitz continuous in N, so a constant $L > 0$ exists, such that:

$$\|g(x) - g(y)\| \leq L\|x - y\| \text{ for any } x, y \in N. \qquad (3.2)$$

By using the Assumption 1, there exists a constant $\lambda \geq 0$ such that:

$$\|g(x)\| \leq \lambda \quad \forall\, x \in \Gamma. \qquad (3.3)$$

For example, if a function satisfying Assumption 1 is the Rosenbrock function (Andrei, 2008), which is defined as:

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2\ ,\ x \in \mathbb{R}^2,$$

with

$$\nabla f(x) = \begin{pmatrix} -400x_1(x_2 - x_1^2) + 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{pmatrix}.$$

So, for the initial point $x_0 = [-1 \quad 1]$, we get

$\nabla f(x_0) = \binom{4}{0}$, and $\|\nabla f(x_0)\| = \sqrt{(4)^2 + (0)^2} = \sqrt{16} = 4.$

Lemma 1: Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm 1. Then for $\sigma < \frac{1}{2}$ and $k \geq 0$, we have:

$$\frac{\|g_k\|}{\|d_k\|} \leq \frac{2}{4 - 3\tau}, \qquad (3.4)$$

$$\frac{\|g_k\|}{\|d_k\|} \leq \frac{m}{m - \tau}, \qquad (3.5)$$

where $\tau \in (\frac{2}{3}, 1)$, and $m > 1$.

Proof: The proof is by induction. For $k = 0$, it is obvious that $\frac{\|g_0\|}{\|d_0\|} = 1 \leq \frac{2}{4-3\tau}, \frac{\|g_0\|}{\|d_0\|} = 1 \leq \frac{m}{m-\tau}$. Next, assume that (3.4) and (3.5) are both true for $k \geq 1$. Then by rewriting equation (1.3) for $k + 1$ and then by using the dot product, we get:

$$d_{k+1}.\,d_{k+1} = (-g_{k+1} + \beta_{k+1}d_k).\,(-g_{k+1} + \beta_{k+1}d_k),$$

which leads to:

$$\|d_{k+1}\|^2 = \|g_{k+1}\|^2 - 2\beta_{k+1}g_{k+1}^T d_k + (\beta_{k+1}\,\|d_k\|)^2. \qquad (3.6)$$

Again rewriting (1.3) as:

$$d_{k+1} + g_{k+1} = \beta_{k+1}d_k\,, \qquad (3.7)$$

and then from the definition of the dot product, we obtain:

$$\|d_{k+1}\|^2 + \|g_{k+1}\|^2 + 2g_{k+1}^T d_{k+1} = (\beta_{k+1}\,\|d_k\|)^2. \qquad (3.8)$$

Substituting (3.8) in (3.6), led to:

$$\|d_{k+1}\|^2 + \|g_{k+1}\|^2 + 2g_{k+1}^T d_{k+1} = \|d_{k+1}\|^2 - \|g_{k+1}\|^2 + 2\beta_{k+1}g_{k+1}^T d_k\,,$$

$$2\|g_{k+1}\|^2 + 2g_{k+1}^T d_{k+1} = 2\beta_{k+1}g_{k+1}^T d_k\,,$$

$$\|g_{k+1}\|^2 + g_{k+1}^T d_{k+1} = \beta_{k+1}g_{k+1}^T d_k\,,$$

$$\|g_{k+1}\|^2 = -g_{k+1}^T d_{k+1} + \beta_{k+1}g_{k+1}^T d_k. \qquad (3.9)$$

Taking the absolute value of both sides of (3.9) and (1.12), we have:

$$\|g_{k+1}\|^2 = \left|-g_{k+1}^T d_{k+1} + \beta_{k+1} g_{k+1}^T d_k\right| \le \left|g_{k+1}^T d_{k+1}\right| + \sigma|\beta_{k+1}|\left|g_k^T d_k\right|. \quad (3.10)$$

Applying the Cauchy-Schwartz inequality, and using (2.2), we get:

$$\|g_{k+1}\|^2 \le \|g_{k+1}\|\|d_{k+1}\| + \sigma \frac{\|g_{k+1}\|^2}{\|d_k\|^2}\|g_k\|\|d_k\|,$$

$$\|g_{k+1}\|^2 \le \|g_{k+1}\|\|d_{k+1}\| + \sigma\|g_{k+1}\|^2 \frac{\|g_k\|}{\|d_k\|}.$$

Applying induction hypothesis of Lemma1, we obtain:

$$\|g_{k+1}\|^2 \le \|g_{k+1}\|\|d_{k+1}\| + \frac{2\sigma}{4-3\tau}\|g_{k+1}\|^2.$$

By letting $\gamma = \frac{2\sigma}{4-3\tau}$, then:

$$\|g_{k+1}\|^2 - \gamma\|g_{k+1}\|^2 \le \|g_{k+1}\|\|d_{k+1}\|,$$

$$\|g_{k+1}\|^2(1-\gamma) \le \|g_{k+1}\|\|d_{k+1}\|,$$

$$\frac{\|g_{k+1}\|}{\|d_{k+1}\|} \le \frac{1}{1-\gamma} \le \frac{2}{4-3\tau}. \quad (3.11)$$

From (3.10), and applying the Cauchy–Schwartz inequality, and from (2.3), we get:

$$\|g_{k+1}\|^2 \le \|g_{k+1}\|\|d_{k+1}\| + \sigma m \frac{\|g_{k+1}\|^2}{\|d_k\|^2}\|g_k\|\|d_k\|,$$

$$\|g_{k+1}\|^2 \le \|g_{k+1}\|\|d_{k+1}\| + \sigma m\|g_{k+1}\|^2 \frac{\|g_k\|}{\|d_k\|}.$$

Applying induction hypothesis, we obtain:

$$\|g_{k+1}\|^2 \le \|g_{k+1}\|\|d_{k+1}\| + \frac{\sigma m}{m-\tau}\|g_{k+1}\|^2.$$

By letting $\vartheta = \frac{\sigma m}{m-\tau}$, then:

$$\|g_{k+1}\|^2 - \vartheta\|g_{k+1}\|^2 \le \|g_{k+1}\|\|d_{k+1}\|,$$

$$\|g_{k+1}\|^2(1-\vartheta) \le \|g_{k+1}\|\|d_{k+1}\|,$$

$$\frac{\|g_{k+1}\|}{\|d_{k+1}\|} \le \frac{1}{1-\vartheta} \le \frac{m}{m-\tau}. \quad (3.12)$$

Therefore, from (3.11) and (3.12), Lemma 1 holds for $k+1$. The proof is completed.

3.1. Sufficient Descent Condition:

Theorem 1. Suppose Assumption 1 is true, then Algorithm 1 satisfies the sufficient descent property (1.13) with:

$$c = 1 - \frac{2\sigma}{4-3\tau} \quad \text{and} \quad c = 1 - \frac{\sigma m}{m-\tau}.$$

Proof: From (1.4) and (2.2), then:

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \beta_{k+1} g_{k+1}^T d_k \le -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\|d_k\|^2} g_{k+1}^T d_k. \quad (3.14)$$

from (1.12) and the Cauchy–Schwartz, we get:

$$g_{k+1}^T d_{k+1} \le -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\|d_k\|^2} g_{k+1}^T d_k \le -\|g_{k+1}\|^2 + \sigma\frac{\|g_{k+1}\|^2}{\|d_k\|^2}\|g_k\|\|d_k\| = -\|g_{k+1}\|^2 + \sigma\|g_{k+1}\|^2\frac{\|g_k\|}{\|d_k\|}.$$

From (3.4) in Lemma1, we obtain:

$$g_{k+1}^T d_{k+1} \le -\|g_{k+1}\|^2 + \frac{2\sigma}{4-3\tau}\|g_{k+1}\|^2 = -\|g_{k+1}\|^2\left(1 - \frac{2\sigma}{4-3\tau}\right) \quad (3.15)$$

In addition, from (3.14), (2.3), (1.12), and Cauchy–Schwartz inequality, we get:

$$g_{k+1}^T d_{k+1} \le -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\|d_k\|^2} g_{k+1}^T d_k \le -\|g_{k+1}\|^2 + \sigma m\frac{\|g_{k+1}\|^2}{\|d_k\|^2}\|g_k\|\|d_k\| = -\|g_{k+1}\|^2 + \sigma m\|g_{k+1}\|^2\frac{\|g_k\|}{\|d_k\|}.$$

From (3.5) in Lemma1, we obtain:

$$g_{k+1}^T d_{k+1} \le -\|g_{k+1}\|^2 + \frac{\sigma m}{m-\tau}\|g_{k+1}\|^2 = -\|g_{k+1}\|^2\left(1 - \frac{\sigma m}{m-\tau}\right). \quad (3.16)$$

Therefore, from (3.15), (3.16), we can deduce that (1.13) holds for $k \ge 0$. The proof is completed.

3.2 Global Convergence:

Lemma 2. Suppose Assumption 1 holds true. Consider any CG method of from (1.3), where d_{k+1} is a descent search direction and α_k satisfies the strong Wolfe line search. The following condition, known as the Zoutendijk condition, holds:

$$\sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty.$$

For the proof of this Lemma 2. Please refer to Zoutendijk (1970).

Theorem 2. Suppose that Assumption 1 holds true. Then Algorithm 1 is convergent, as follows:

$$\lim_{k\to\infty}\|g_k\| = 0 \quad \text{or} \quad \sum_{k=1}^{\infty}\frac{\|g_k\|^4}{\|d_k\|^2} < \infty.$$

Proof: We use a contradiction; that is, if Theorem 2 is not true, then there exists a constant $\epsilon > 0$, such that:

$$\|g_k\| \ge \epsilon. \quad (3.17)$$

We can rewrite (1.3) as:

$$d_{k+1} = -g_{k+1} + \beta_{k+1} d_k \quad (3.18)$$

and multiply both sides of (3.18) by d_{k+1}, thus obtaining:

$$\|d_{k+1}\|^2 = -g_{k+1}^T d_{k+1} + \beta_{k+1} d_{k+1}^T d_k. \quad (3.19)$$

Case 1, using (2.2), and divide both sides of (3.19) by $\|g_{k+1}\|^4$, we obtain:

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \le -\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^4} + \frac{\|g_{k+1}\|^2}{\|d_k\|^2}\frac{d_{k+1}^T d_k}{\|g_{k+1}\|^4},$$

applying the Cauchy–Schwartz, we get:

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \le -\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^4} + \frac{1}{\|d_k\|}\frac{\|d_{k+1}\|}{\|g_{k+1}\|^2},$$

$$\le -\frac{1}{2}\left(\frac{g_{k+1}}{\|g_{k+1}\|^2} + \frac{d_{k+1}}{\|g_{k+1}\|^2}\right)^2 \frac{1}{2}\frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^4} - \frac{1}{2}\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} + \frac{1}{\|d_k\|}\frac{\|d_{k+1}\|}{\|g_{k+1}\|^2},$$

$$\le -\frac{1}{2}\frac{1}{\|g_{k+1}\|^2} - \frac{1}{2}\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} + \frac{1}{\|d_k\|}\frac{\|d_{k+1}\|}{\|g_{k+1}\|^2},$$

$$\le -\frac{1}{2}\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} + \frac{1}{\|d_k\|}\frac{\|d_{k+1}\|}{\|g_{k+1}\|^2},$$

$$\frac{3}{2}\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \le \frac{1}{\|d_k\|}\frac{\|d_{k+1}\|}{\|g_{k+1}\|^2},$$

From (3.4) in Lemma 1, we get:

$$\frac{\|d_{k+1}\|}{\|g_{k+1}\|^2} \le \frac{2}{3}\frac{1}{\|d_k\|} \le \frac{2}{3}\frac{\mu}{\|g_k\|}, \text{ where } \mu = \frac{2}{4-3\tau}.$$

Squaring both sides of the above inequality, we obtain:

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \le \frac{4}{9}\frac{\mu^2}{\|g_k\|^2},$$

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \le \frac{4}{9}\frac{\mu^2}{\|g_k\|^2} \le \frac{4\mu^2}{9}\sum_{i=0}^{k}\frac{1}{\|g_k\|^2} \le \frac{4\mu^2}{9}\frac{k}{\epsilon^2},$$

$$\frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \ge \frac{9\epsilon^2}{4\mu^2 k}. \quad (3.20)$$

Therefore, from (3.20) and (3.17), it follows that:

$$\sum_{k=0}^{\infty}\frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} = \infty.$$

This contradicts the Zoutendijk condition in Lemma 2. Therefore, the proof is completed.

Case 2, using (2.3) and dividing both sides of (3.19) by $\|g_{k+1}\|^4$, we obtain:

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \le -\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^4} + m\frac{\|g_{k+1}\|^2}{\|d_k\|^2}\frac{d_{k+1}^T d_k}{\|g_{k+1}\|^4},$$

applying the Cauchy–Schwartz inequality, we get:

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \le -\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^4} + \frac{m}{\|d_k\|}\frac{\|d_{k+1}\|}{\|g_{k+1}\|^2},$$

$$\le -\frac{1}{2}\left(\frac{g_{k+1}}{\|g_{k+1}\|^2} + \frac{d_{k+1}}{\|g_{k+1}\|^2}\right)^2 - \frac{1}{2}\frac{\|g_{k+1}\|^2}{\|g_{k+1}\|^4} - \frac{1}{2}\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} + \frac{m}{\|d_k\|}\frac{\|d_{k+1}\|}{\|g_{k+1}\|^2},$$

$$\le -\frac{1}{2}\frac{1}{\|g_{k+1}\|^2} - \frac{1}{2}\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} + \frac{m}{\|d_k\|}\frac{\|d_{k+1}\|}{\|g_{k+1}\|^2},$$

$$\le -\frac{1}{2}\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} + \frac{m}{\|d_k\|}\frac{\|d_{k+1}\|}{\|g_{k+1}\|^2},$$

$$\frac{3}{2}\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \leq \frac{m}{\|d_k\|}\frac{\|d_{k+1}\|}{\|g_{k+1}\|^2},$$

From (3.5) in Lemma 1, we get:

$$\frac{\|d_{k+1}\|}{\|g_{k+1}\|^2} \leq \frac{2}{3}\frac{m}{\|d_k\|} \leq \frac{2}{3}\frac{m\rho}{\|g_k\|}, \text{ where } \rho = \frac{m}{m-\tau},$$

Squaring both sides of the above inequality, we obtain:

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \leq \frac{4}{9}\frac{m^2\rho^2}{\|g_k\|^2},$$

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \leq \frac{4}{9}\frac{m^2\rho^2}{\|g_k\|^2} \leq \frac{4m^2\rho^2}{9}\sum_{i=0}^{k}\frac{1}{\|g_k\|^2} \leq \frac{4m^2\rho^2}{9}\frac{k}{\epsilon^2},$$

$$\frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \geq \frac{9\epsilon^2}{4m^2\rho^2 k}. \qquad (3.21)$$

Therefore, from (3.21) and (3.17), it follows that:

$$\sum_{k=0}^{\infty}\frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} = \infty.$$

This contradicts the Zoutendijk condition in Lemma 2. Therefore, the proof is completed.

4. Numerical Results and Discussion

In this section, to show the efficiency of the proposed method in practice, we compare it with the RMIL, MRMIL, and RMIL+ methods. Most of the test problems used are from Andrei (2008). The condition $\|g_k\| \leq \epsilon$ where $\epsilon = 10^{-6}$ is used as a stopping criterion. All methods are run on a PC ACER (Intel® Core™ i3-3217u CPU @ 1.8 GHZ, with 4.00 GB RAM, Windows 10 Home Premium). The comparison is based on the number of iterations and CPU time. The strong Wolfe line search is used as the inexact line search. The performance results are shown in Figures 1 and 2, respectively, using the performance profile introduced by Dolan and More (2002). In Table 1, Dim is for the dimension of the test functions.

In the Dolan and More performance, we use the performance profile to introduce the notion of a means to evaluate and compare the performance of the set solvers S on a test set p. Assuming n_s solvers and n_p problems exist, for each problem p and solver s, they defined $t_{p,s}$ as computing time (the number of iterations or CPU time or others) required to solve problem p by solver s. They compared the performance on problem p by solver S with the best performance by any solver in this problem by using the performance ratio:

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

Suppose that a parameter $r_M \geq r_{p,s}$ for all p and S is chosen, and $r_{p,s} = r_M$ if and only if solver S does not solve problem p. The performance solver S of the given problems must be robust, but we would like to obtain all evaluation performance of the solver; thus, it was defined:

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

The $P(t)_s$ was probability for solver $s \in S$ that a performance ratio $r_{p,s}$ was within a factor $t \in R$ of efficient ratio. Then, function P_s was the cumulative distribution function for the performance ratio. The performance profile $P: R \to [0,1]$ as a solver as non-decreasing, piecewise, and continuous from the right. The value $P(1)_s$ is the probability that the solver will perform better than the rest of the solvers. In all, a solver with high values of $P(t)_s$ or at the top right of the figures is preferable or represents a robust solver.

Table 1: A list of test problems

No.	Problems	Dim	Initial points
1	Booth	2	(5, 5) (8, 8)
2	Three-hump camel	2	(-1, 2) (-1, 1)
3	Six-hump camel	2	(-1, 2) (5, 5)
4	Trecanni	2	(-5, -5) (10, 10)
5	Zettl	2	(-1, 2) (7, 7)
6	Leon	2	(0, 0) (-1, -1)
7	Extended. Maratos	4	(5,5,5,5) (10,10,10,10)
8	Arwhead	4	(-1,-1,-1,-1) (10,10,10,10)
9	Freudenstein & Roth	4	(-1, -1, -1, -1) (5, 5, 5, 5)
10	Generalized Quadratic	10	(2, 2, ..., 2) (10, 10, ..., 10)
11	Fletcher	10	(0, 0, ..., 0) (10, 10, ..., 10)
12	Generalized Tridiagonal 1	10	(2, 2, ..., 2) (-3, -3, ..., -3)
13	Hager	10	(1, 1, ..., 1) (-5, -5, ..., -5)
14	Ex. White & Holst	500	(-1.2, -1.2, ..., -1.2) (3, 3, ..., 3)
		1000	(2, 2, ..., 2) (-5, -5, ..., -5)
15	Extended Rosenbrock	1000	(3, 3, ..., 3) (10, 10, ..., 10)
		10000	(-1, -1, ..., -1) (-2, -2, ..., -2)
16	Extended Himmelblau	1000	(-1, -1, ..., -1) (20, 20, ..., 20)
		10000	(-1, -1, ..., -1) (20, 20, ..., 20)
17	Shallow	1000	(-5, -5, ..., -5) (5, 5, ..., 5)
18	Extended Beale	1000	(-1, -1, ..., -1) (2, 2,...,2)
		10000	(-1, -1, ..., -1) (0.2, 0.2, ..., 0.2)
19	Extended DENSCHNB	10 100	(-2, -2, ..., -2) (-10, -10, ..., -10) (1, 1, ..., 1) (-20, -20, ..., -20)
20	Extended Penalty	10 100	(5, 5, ..., 5) (-10, -10, ..., -10) (5, 5, ..., 5) (-10, -10, ..., -10)
21	Matyas	2	(10,10)
22	Colville	4	(-3,-3,-3,-3) (10,10,10,10)
23	Dixon and Price	100	(-2,...,-2) (10,...,10)
24	Sum squares	4 100	(-2,...,-2) (-2,...,-2)
25	Extended Wood	4	(10,...,10) (-1,...,-1)
26	Strait	4 100 1000	(-1,...,-1) (5,...,5) (5,...,5)
27	ARWHEAD	10	(-2,...,-2)
28	Generalized Rosenbrock	10	(-5,...,-5)
29	Quartic	100	(-4,...,-4)
30	DIXON3DQ	10 100	(-10,...,-10) (-7,...,-7)
31	NONSCOMP	10	(2,...,2)
32	POWER	100	(-5,...,-5)
33	Quadratic QF1 Extended	100	(3,...,3)
34	Quadratic Penalty QP2	10 1000	(10,...,10) (-3,...,-3)
35	Quadratic Penalty QP1	1000	(-3,...,-3)

Figure 1: Performance profile based on the number of iterations

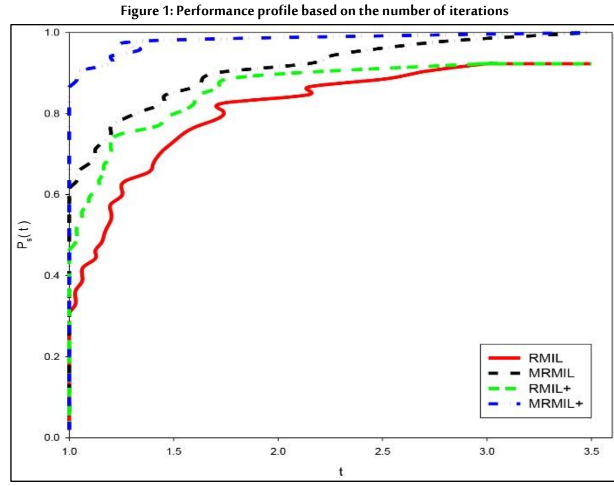
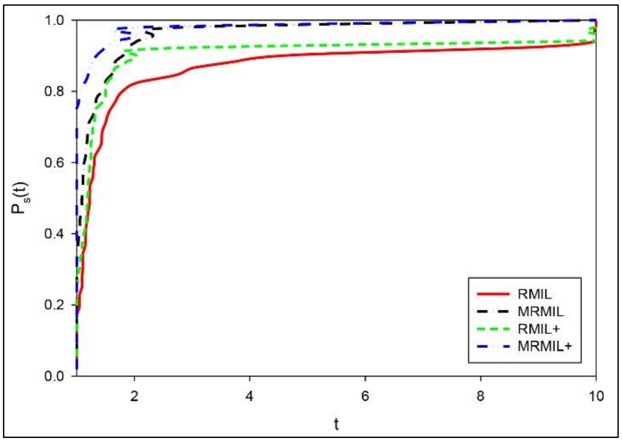


Figure 2: Performance profile based on the CPU time



Clearly from Figs. 1 and 2, the MRMIL+ curve lies above the RMIL, MRMIL, and RMIL+ curves. Therefore, the MRMIL+ method performs much better than the other three methods. Moreover, MRMIL and MRMIL+ methods solve all problems, whereas RMIL+ and RMIL solve about 83% of the problems. Therefore, since MRMIL+ has the top curve and it solves all problems, we can conclude that it is the best method.

5. Conclusion

In this paper, based on Rivaie *et al.* (2015), we have presented a modified version of the RMIL+ conjugate gradient method. The proof of the sufficient descent property and of the global convergence of our modified method when it is applied under the strong Wolfe line search has been established. Moreover, to show its efficiency in practical computation, we have compared it with the RMIL, MRMIL, and RMIL+ methods. The result of this comparison is that our modified method performs much better than the others.

Biographies

Awad Abdelrahman

Department of Mathematics, Faculty of Mathematical and Computer Sciences, University of Gezira, Wad Madani, Sudan, awad.abdalla26@yahoo.com, 00249962255448

Dr. Abdelrahman (Sudanese) is an assistant Professor at the University of Gezira. He has received his Ph.D. degree in mathematical sciences from the University of Malaysia Terengganu, Malaysia in 2017. Also, he is now a head of the documentation committee and a member of the courses development committee. He is interested in optimization and computational mathematics. He supervised undergraduate and M.Sc. students. He has participated in conferences in Malaysia and Sudan and served as a co-chair of the ICCCEEE18 conference, held in Khartoum in 2018. ORCID: 0000-0001-5467-5124.

Osman Omer Osman Yousif

Department of Mathematics, Faculty of Mathematical and Computer Sciences, University of Gezira, Wad Madani, Sudan, osman_om@hotmail.com, 00249119733535

Dr. Yousif (Sudanese) is an assistant professor at the University of Gezira, Sudan since 2017 after the approval of his Ph.D. degree from University of Malaysia Terengganu, Malaysia. Also, he is now the head of Mathematics Department. His research interests lie at the intersection of computational mathematics and optimization methods. He has supervised many graduation projects in these areas. Besides, he has participated in four conferences as a presenter in Malaysia and published four papers in ISI- and Scopus-indexed journals.

Mogtaba Mohammed

Department of Mathematics, College of Science, Majmaah University, Zulfi, Saudi Arabia, mogtaba.m@mu.edu.sa, 00966543244128

Dr Mohammed (Sudanese) is an associate professor. He has received a Ph.D. and a post doctorate scholarship in mathematical sciences from the University of Pretoria, South Africa. He is interested in mathematical modelling and computational mathematics. He has supervised several M.Sc. students and B.Sc. students. In addition, he has many papers published in ISI- and Scopus-indexed journals. He has been invited to give talks in several international conferences.

Murtada K. Elbashir

Department of information system, College of Computer and Information Sciences, Jouf University, Sakaka, Saudi Arabia, mkelfaki@ju.edu.sa, 00966557186485

Dr. Elbashir (Sudanese) is an associate professor. His Ph.D. degree is in computer science and technology from Central South University, Changsha, China. His current research interest includes machine learning and bioinformatics. He has more than 20 papers published in ISI- and Scopus-indexed journals. He has participated in different conferences in Sudan, China and South Africa. Also, he has been invited to give talks in several international conferences. He has supervised several M.Sc. and Ph.D. students.

References

Al-Baali, M. (1985). Descent property and global convergence of the Fletcher-Reeves method with inexact line search. *IMA Journal of Numerical Analysis*, **5**(1), 121–4.

Andrei, N. (2008). An unconstrained optimization test functions collection. *Adv. Model. Optim*, **10**(1), 147–61.

Andrei, N. (2011). Open problems in nonlinear conjugate gradient algorithms for unconstrained optimization. *Bulletin of the Malaysian Mathematical Sciences Society*, **34**(2), 19–330.

Dai, Y.H. and Yuan, Y. (1999). A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization*, **10**(1), 177–82.

Dai, Z. (2016). Comments on a new class of nonlinear conjugate gradient coefficients with global convergence properties. *Applied Mathematics and Computation*, **276**(n/a), 297–300.

Dolan, E.D. and Moré, J.J. (2002). Benchmarking optimization software with performance profiles. *Mathematical Programming*, **91**(2), 201–13.

Gilbert, J.C. and Nocedal, J. (1992). Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal on Optimization*, **2**(1), 21–42.

Guanghui, L., Jiye, H. and Hongxia, Y. (1995). Global convergence of the Fletcher-Reeves algorithm with inexact linesearch. *Applied Mathematics-A Journal of Chinese Universities*, **10**(1), 75–82.

Fletcher, R. (1987). *Practical Method of Optimization, Unconstrained Optimization*. New York: John Wiley and Sons.

Fletcher, R. and Reeves, C.M. (1964). Function minimization by conjugate gradients. *The computer journal*, **7**(2), 149–54.

Hestenes, M.R. and Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Standards*, **49**(6), 409–35.

Jiang, X. and Jian, J. (2019). Improved Fletcher–Reeves and Dai–Yuan conjugate gradient methods with the strong Wolfe line search. *Journal of Computational and Applied Mathematics*, **348**(n/a), 525–34.

Liu, Y. and Storey, C. (1991). Efficient generalized conjugate gradient algorithms, part 1: Theory. *Journal of Optimization Theory and Applications*, **69**(1), 129–37.

Rivaie, M., Mamat, M. and Abashar, A. (2015). A new class of nonlinear conjugate gradient coefficients with exact and inexact line searches. *Applied Mathematics and Computation*, **268**(n/a), 1152–63.

Rivaie, M., Mamat, M., June, L. W. and Mohd, I. (2012). A new class of nonlinear conjugate gradient coefficients with global convergence properties. *Applied Mathematics and Computation*, **218**(22), 11323–32.

Polak, E. and Ribiere, G. (1969). Note on the convergence of conjugate direction methods. *Mathematical Modeling and Numerical Analysis*, **3**(1), 35–43.

Powell, M.J. (1984). Nonconvex minimization calculations and the conjugate gradient method. In: M.J Powell (ed.) *Numerical Analysis*. Berlin, Germany: Springer.

Touati-Ahmed, D. and Storey, C. (1990). Efficient hybrid conjugate gradient techniques. *Journal of Optimization Theory and Applications*, **64**(2), 379–97.

Yousif, O.O.O. (2020). The convergence properties of RMIL+ conjugate gradient method under the strong Wolfe line search. *Applied Mathematics and Computation*, **367**(1), 124777.

Zoutendijk, G. (1970). Nonlinear programming, computational methods. In: J. Abadie (ed.) *Integer and Nnonlinear Programming*. Amsterdam, Switzerland: North-Holland.