

The convergence properties of some descent conjugate gradient algorithms for optimization models



I. M. Sulaiman^a, M. Mamat^{a,*}, A. E. Owoyemi^b, P. L. Ghazali^c, M. Rivaie^d, M. Malik^a

^aFaculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Besut Campus, 22200, Malaysia.

^bDepartment of General Studies, Federal College of Agricultural Produce Technology, Hoto GARA extension, Kano, Nigeria.

^cFaculty of Business and Management, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.

^dDepartment of Computer Science and Mathematics, Universiti Teknologi Mara, Terengganu, Malaysia.

Abstract

The three-term conjugate gradient (CG) algorithms are among the efficient variants of CG method for convex and non-convex functions. This is because most three-term algorithms are constructed using the classical CG method whose numerical performance has been tested and convergence proved. In this paper, we present a modification of RMIL+ CG method proposed by Dai [Z. Dai, Appl. Math. Comput., 267 (2016), 297–300] based on the convergence analysis of RMIL (2012) CG method. Interestingly, the modified method possesses sufficient descent condition and the global convergence prove was established using exact minimization condition. We further extended the results of the modified RMIL+ to construct a three-term CG algorithm and also show that the method satisfies the sufficient descent condition under the strong Wolfe line search. Preliminary numerical results are reported based on known benchmark problems which show that the proposed methods are efficient and promising compare to other CG methods.

Keywords: Conjugate gradient method, three-term CG algorithm, line searches, optimization models, convergence analysis.

2020 MSC: 65K05, 65K10.

©2021 All rights reserved.

1. Introduction

Consider the following optimization model

$$\min f(x), \quad x \in \mathbb{R}^n, \quad (1.1)$$

where the nonlinear function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is smooth whose gradient $g(x)$ is available. Problems of the form (1.1) often arise in areas of sciences, engineering, and social science [11, 28]. The CG algorithm is among the most efficient algorithm for obtaining the solution of (1.1) using the recursive formula

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 1, 2, 3, \dots, \quad (1.2)$$

where $\alpha_k > 0$ is the step-size obtained by certain line search process, and d_k is the direction of search

*Corresponding author

Email address: must@unisza.edu.my (M. Mamat)

doi: [10.22436/jmcs.022.03.02](https://doi.org/10.22436/jmcs.022.03.02)

Received: 2020-05-20 Revised: 2020-06-03 Accepted: 2020-06-15

defined by

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \quad d_0 = -g_0. \quad (1.3)$$

The scalar β_k refers to the CG parameter whose different forms characterizes different CG method, and $g_k = g(x_k)$. Some well-known formulas for β_k are:

$$\beta_k^{\text{HS}} = \frac{g_k^T(g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}} \quad \text{Hestenes – Stiefel [12]}, \quad (1.4)$$

$$\beta_k^{\text{FR}} = \frac{g_k^T g_k}{\|g_{k-1}\|^2} \quad \text{Fletcher – Reeves [10]}, \quad (1.5)$$

$$\beta_k^{\text{PRP}} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2} \quad \text{Polak – Ribiere [21, 22]}, \quad (1.6)$$

$$\beta_k^{\text{LS}} = \frac{g_k^T(g_k - g_{k-1})}{-d_{k-1}^T g_{k-1}} \quad \text{Liu – Storey [16]}, \quad (1.7)$$

$$\beta_k^{\text{CD}} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}} \quad \text{ConjugateDescent [9]}, \quad (1.8)$$

$$\beta_k^{\text{DY}} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}} \quad \text{Dai – Yuan [7]}. \quad (1.9)$$

The convergence properties of (1.4) - (1.9) have been studied by numerous researchers. From the numerical point of view, the PRP algorithm outperformed the algorithm of FR. Also, for convex objective function, PRP method would converge globally under exact minimization condition [21]. The exact line search is computed such that t_k satisfy:

$$f(x_k + t_k d_k) = \min_{t \geq 0} f(x_k + t d_k). \quad (1.10)$$

However, Powell [23] show that the PRP algorithm would not converge globally on several nonconvex functions.

Some convergence result also requires the step-size t_k to possess the weak Wolfe (WWP) or strong Wolfe (SWP) line search. The WWP line search is computed such that t_k satisfy.

$$f(x_k + t_k d_k) \leq f(x_k) + \delta t_k g_k^T d_k, \quad (1.11)$$

$$g(x_k + t_k d_k)^T d_k \geq \sigma g_k^T d_k. \quad (1.12)$$

The SWP line search is computed such that t_k satisfy (1.11) and

$$|g(x_k + t_k d_k)^T d_k| \leq -\sigma g_k^T d_k, \quad (1.13)$$

where $0 < \delta < \sigma < 1$. Al-Baali [1] show the FR method possess the descent condition:

$$g_k^T d_k \leq -C \|g_k\|^2, \quad C > 0, \quad (1.14)$$

and used the SWP line search to show that the algorithm converges globally for general function. For more convergence results of the CG method, please refer to [6, 7, 9, 15, 21, 25, 28, 31].

Recently, Rivaie et al. [24] construct an efficient variant of PRP method by defining a new denominator as follows:

$$\beta_k^{\text{RMIL}} = \frac{g_k^T(g_k - g_{k-1})}{\|d_{k-1}\|^2} \quad (1.15)$$

and show that (1.15) possess the descent property (1.14) and further prove the global convergence using exact minimization condition. However, Dai [5] showed that the convergence proof is not correct due to a wrong inequality used and presented some necessary corrections as follows:

$$\beta_k^{\text{RMIL}+} = \begin{cases} \frac{g_k^T(g_k - g_{k-1})}{\|d_{k-1}\|^2}, & \text{if } 0 \leq |g_k^T g_{k-1}| \leq \|g_k\|^2, \\ 0, & \text{otherwise,} \end{cases} \quad (1.16)$$

and further show that RMIL+ converges globally under exact line search with efficient numerical results. More recently, [20], extended the result of Dai [5] and prove the RMIL+ converge globally using the strong Wolfe conditions.

One of the efficient variants of Conjugate gradient algorithm is known as the three-term CG algorithm. This scheme often defines a new search direction d_k by the convex combination of an existing CG method with a search direction. Beale [4] was among the early researchers that studied the general three-term CG method. Using the β_k^{HS} method defined in (1.4), he constructed a new direction of search as follows:

$$d_k = -g_k + \beta_k d_{k-1} + \gamma_k d_t, \quad (1.17)$$

where d_t is the restart directions and $\gamma_k = 0$ for $k = t + 1$, and

$$\gamma_k = \frac{g_k^T y_t}{d_t^T y_t} \quad k > t + 1. \quad (1.18)$$

The performance of this method was improved using an efficient restart strategy developed by Powell [23] and McGuire [19]. The following conditions

$$g_k^T d_k \geq \varphi \|g_k\| \|d_k\| \quad (1.19)$$

and Powell-Beale condition

$$|g_{k-1}^T g_k| < 0.2 \|g_k\|^2 \quad (1.20)$$

are imposed on (1.17) so that it possesses the descent properties. Further research on the Three-term CG methods includes that of [29] who proposed a three term CG algorithm (TTPRP) using the coefficient of β_k^{PRP} as follows:

$$d_k = -g_k + \beta_k d_{k-1} + \theta_{k-1} y_{k-1},$$

where $\theta_k = -\frac{g_k^T d_{k-1}}{g_{k-1}^T g_{k-1}}$, and extended the result to β_k^{HS} (TTHS) [30] as follows:

$$d_k = \begin{cases} -g_k & \text{if } s_{k-1}^T y_k < \varepsilon_1 \|g_{k-1}\|^r s_{k-1}^T s_{k-1}, \\ -g_k + \beta_k^{HS} + \theta_{k-1} y_{k-1} & \text{otherwise,} \end{cases}$$

where $\theta_k = -\frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}}$ and $s_{k-1} = x_k - x_{k-1}$, $r \geq 0$, $\varepsilon_1 > 0$. An attractive feature of these methods is that:

$$g_k^T d_k = -\|g_k\|^2 \quad (1.21)$$

holds irrespective of the condition of t_k employed. The convergence analysis of TTPRP was established using a modified Armijo line search and that of TTHS under the standard Wolfe line search. For more studies and recent reference on three-term conjugate gradient methods, interested researchers may refer to [2, 3, 13, 14, 17, 18, 26, 29, 30].

This paper defined a variant of β_k^{RMIL+} method in such a way that the modification not only possesses the nice convergence properties of β_k^{RMIL+} , but also the efficient numerical effect of the method. Moreover, the modified method was extended to construct a new three-term CG algorithm whose convergence analysis was studied using the strong Wolfe conditions.

The other part of the article is structured as follows. In Section 2, we proposed the modified β_k^{RMIL+} and Three-term CG method and prove their convergence results. To illustrate the efficiency of the new modifications, some numerical experiments were conducted whose results are presented in Section 3. Finally, the conclusion is discussed in Section 4.

2. New formulas and algorithm

Motivated by the comments of Dai [5] on the convergence properties of RMIL method, this paper

defined a modified RMIL+ as follows:

$$\beta_k^{\text{SRMIL}^+} = \begin{cases} \frac{\|g_k\|^2 - g_k^T g_{k-1}}{\|d_{k-1}\|^2}, & \text{if } 0 \leq |g_k^T g_{k-1}| \leq \|g_k\|^2, \\ 0, & \text{otherwise,} \end{cases} \quad (2.1)$$

that is, we expanded the term $g_k^T(g_k - g_{k-1})$ in RMIL+ to $\|g_k\|^2 - g_k^T g_{k-1}$ and called the proposed method SRMIL+. It is obvious to know that under exact minimization condition, SRMIL+ will reduce to RMIL+ method. The algorithm for the proposed method is defined as follows.

Algorithm 2.1.

Stage 1. Given $x_0 \in \mathbb{R}^n$, $d_0 = -g_0$, set $k = 0$.

Stage 2. Compute t_k by (1.10).

Stage 3. Update the new point based on (1.2).

Stage 4. Compute β_k by (2.1) and update d_k by (1.3).

Stage 5. If $\|g_k\| \leq 10^{-6}$, then terminate. Else, go to stage 2 with $k = k + 1$.

In the convergence analysis of the CG algorithms, the following assumptions are often required [27].

Assumption A. $f(x)$ is bounded from below on the level set $\Omega = \{x \in \mathbb{R}^n / f(x) \leq f(x_0)\}$.

Assumption B. In some neighborhood N of Ω , f is smooth and $g(x)$ is Lipschitz continuous in N , such that, there exist $L > 0$ (constant) satisfying;

$$\|g(x) - g(y)\| \leq L \|x - y\| \quad \forall x, y \in N. \quad (2.2)$$

Some interesting properties of SRMIL+ method is presented as follows.

Lemma 2.2. The proposed $\beta_k^{\text{SRMIL}^+}$ satisfies

$$0 \leq \beta_k^{\text{SRMIL}^+} \leq \frac{\|g_k\|^2}{\|d_{k-1}\|^2}. \quad (2.3)$$

Proof. Since for $0 \leq |g_k^T g_{k-1}| \leq \|g_k\|^2$,

$$\beta_k^{\text{SRMIL}^+} = \frac{\|g_k\|^2 - g_k^T g_{k-1}}{\|d_{k-1}\|^2},$$

then, it is obvious that (2.3) holds. □

2.1. Convergence analysis

The convergence analysis of $\beta_k^{\text{SRMIL}^+}$ would be studied in this section. One of the main conditions that every CG algorithm should possess is the sufficient descent property defined in (1.14)

2.1.1. Sufficient descent condition

The following theorem would be used to show that $\beta_k^{\text{SRMIL}^+}$ possess the descent property.

Theorem 2.3. For any CG algorithm defined by (1.2) and (1.3), where β_k is given as (2.1), then, (1.14) holds for all $k \geq 0$.

Proof. Suppose $k = 0$, then, (1.14) becomes $g_0^T d_0 \leq -C \|g_0\|^2$. This implies that (1.14) holds true. Next, we need to show that (1.14) also holds true for $k \geq 1$.

From (1.4), we get

$$d_k = -g_k + \beta_k^{\text{SRMIL}^+} d_{k-1},$$

multiply through by g_k^T will give

$$\begin{aligned} g_k^T d_k &= -g_k^T g_k + \beta_k^{SRMIL+} g_k^T d_{k-1}, \\ &= \|g_k\|^2 + \beta_k^{SRMIL+} g_k^T d_{k-1}. \end{aligned} \quad (2.4)$$

However, for exact line search $g_k^T d_{k-1} = 0$ [13, 29]. Therefore, for $C = 1$, we have

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

which implies that (1.14) holds and thus, completes the proof. \square

2.1.2. Global convergence properties

The global convergence of β_k^{SRMIL+} would be discussed in this section.

Consider the assumptions A and B, then, the following lemma based on Zoutendijk [31] condition is very essential in the convergence studies of most CG algorithm.

Lemma 2.4. *Let Assumptions A and B holds true. For any CG iterative algorithm defined by (1.2), where d_k is defined by (1.3), and the step-size t_k satisfies (1.10). Then,*

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \quad (2.5)$$

The proof of this lemma follows from [31]. From (2.3) and by Lemma 2.2, we have the convergent theorem presented below.

Theorem 2.5. *Suppose Assumptions A and B holds true. For any CG method of the form (1.2) and (1.3), where the step-size t_k is computed by (1.10). Also, suppose Assumptions A and B and the (1.14) holds true. Then, either*

$$\lim_{k \rightarrow \infty} \|g_k\| = 0 \text{ or } \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty.$$

Proof. The prove of this theorem would be by contradiction. We suppose the theorem is not true, then, $\exists \psi > 0$ (ψ is a constant) such that

$$\|g_k\|^2 \geq \psi. \quad (2.6)$$

From (1.4), we have

$$d_k + g_k = \beta_k d_{k-1}, \quad (2.7)$$

squaring all side of (2.7) and dividing by $(g_k^T d_k)^2$ will give

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

which reduces to

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} = \frac{(\beta_k)^2 \|d_{k-1}\|^2}{(g_k^T d_k)^2} + \frac{1}{\|g_k\|^2}. \quad (2.8)$$

However, from (2.3), we have $\beta_k^{\text{SRMIL}^+} \leq \frac{\|g_k\|^2}{\|d_{k-1}\|^2}$. Substituting in (2.8), we obtain

$$\begin{aligned} \frac{\|d_k\|^2}{(g_k^T d_k)^2} &= \left(\frac{\|g_k\|^2}{\|d_{k-1}\|^2} \right)^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} + \frac{1}{\|g_k\|^2} \\ &= \frac{\|g_k\|^2}{\|d_k\|^2 \|d_{k-1}\|^2} + \frac{1}{\|g_k\|^2} \\ &\leq \frac{1}{\|d_{k-1}\|^2} + \frac{1}{\|g_k\|^2}. \end{aligned} \quad (2.9)$$

But $\frac{1}{\|d_0\|^2} = \frac{1}{\|g_0\|^2}$, hence, (2.9) reduces to

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} = \frac{1}{\|g_0\|^2} + \frac{1}{\|g_k\|^2},$$

which implies

$$\begin{aligned} \frac{\|d_k\|^2}{(g_k^T d_k)^2} &\leq \sum_{i=0}^k \frac{1}{\|g_i\|^2} \\ \frac{(g_k^T d_k)^2}{\|d_k\|^2} &\geq \frac{\psi^2}{k}. \end{aligned} \quad (2.10)$$

From (2.6) and (2.10), we have

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = \infty.$$

However, it contradicts our assertion in Lemma 2.2, and thus, completes the proof. \square

Theorem 2.6. Suppose Assumptions A and B holds. For any CG method defined using (1.2) and (1.3), where the step-size t_k is computed using (1.10). Also, suppose Assumptions A, B and (1.14) holds true. Then, either

$$\lim_{k \rightarrow \infty} \|g_k\| = 0 \text{ or } \sum_{k=0}^{\infty} \frac{(g_k)^4}{\|d_k\|^2} < \infty.$$

Proof. Squaring through (2.7) and applying (2.3) will give

$$\|d_k\|^2 = \left(\frac{\|g_k\|^2}{\|d_{k-1}\|^2} \right)^2 \|d_{k-1}\|^2 - 2g_k^T d_k - \|g_k\|^2.$$

After some simplification, we have

$$\|d_k\|^2 = \frac{\|g_k\|^4}{\|d_{k-1}\|^2} - 2g_k^T d_k - \|g_k\|^2. \quad (2.11)$$

Since, from Theorem 2.3, we showed that $\beta_k^{\text{SRMIL}^+}$ satisfied the descent condition, then, from (2.11), we have

$$\begin{aligned} \|d_k\|^2 &= \frac{\|g_k\|^4}{\|d_{k-1}\|^2} + 2C\|g_k\|^2 - \|g_k\|^2, \\ \|d_k\|^2 &= \frac{\|g_k\|^4}{\|d_{k-1}\|^2} - \|g_k\|^2(1 - 2C). \end{aligned} \quad (2.12)$$

Multiplying through (2.12) by $\frac{\|g_k\|^2}{\|d_k\|^2}$ gives

$$\|d_k\|^2 \frac{\|g_k\|^2}{\|d_k\|^2} = \left(\frac{\|g_k\|^4}{\|d_{k-1}\|^2} - \|g_k\|^2 (1 - 2C) \right) \frac{\|g_k\|^2}{\|d_k\|^2},$$

which after little simplification gives

$$\frac{\|d_k\|^2 \|g_k\|^2}{\|d_k\|^2} \leq \frac{\|g_k\|^4}{\|d_k\|^2}. \quad (2.13)$$

However, know that $\lim_{k \rightarrow \infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < 0$ based on Theorem 2.5 which means that if Theorem 2.6 is not true, then, $\lim_{k \rightarrow \infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = 0$ and from (2.13) $\infty \leq \frac{\|g_k\|^4}{\|d_k\|^2}$. This means that for sufficiently large k , the theorem will hold true. \square

Now similar to the construction of $\beta_k^{\text{SRMIL}+}$, we present the proposed three-term CG parameter as follows.

2.2. Three-term $\beta_k^{\text{SRMIL}+}$ method

Motivated by the structure of TTPRP [29], we extend the proposed $\beta_k^{\text{SRMIL}+}$ to construct a three-term CG parameter named $\beta_k^{\text{TTSRMIL}+}$.

An interesting feature of $\beta_k^{\text{TTSRMIL}+}$ is that the method would reduce to the proposed $\beta_k^{\text{SRMIL}+}$ method under exact minimization condition.

The proposed $\beta_k^{\text{TTSRMIL}+}$ is defined as expressed as follows:

$$d_k = \begin{cases} -g_0 & \text{if } k = 0, \\ -g_k + \beta_k d_{k-1} + \theta_k g_{k-1}, & \text{if } k \geq 1, \end{cases} \quad (2.14)$$

where $\beta_k = \beta_k^{\text{SRMIL}+}$ defined in (2.1), and

$$\theta_k = \frac{g_k^T d_{k-1}}{\|d_{k-1}\|^2}. \quad (2.15)$$

2.2.1. Sufficient descent condition

The following theorem would be used to show that $\beta_k^{\text{TTSRMIL}+}$ satisfies the descent condition under SWP line search.

Theorem 2.7. Consider the CG algorithm defined by (1.2) and (2.14), where β_k is given as (2.1) and t_k is generated by strong Wolfe line search (1.11) and (1.13). Then,

$$g_k^T d_k \leq -(1 - \sigma) \|g_k\|^2, \quad \forall k \geq 0, \quad (2.16)$$

Proof. To prove Theorem 2.7, we use mathematical induction. If $k = 0$, then $g_0^T d_0 = -\|g_0\|^2$. Thus, condition (2.16) holds true. Assume that (2.16) holds true for $k - 1$, this implies

$$g_{k-1}^T d_{k-1} < 0,$$

meaning that condition (1.13) is useful.

From (2.14), multiply by g_k^T will give

$$g_k^T d_k = -\|g_k\|^2 + (\beta_k)^2 \cdot g_k^T d_{k-1} + \theta_k \cdot g_k^T g_{k-1}.$$

Substituting (2.1) and (2.15), we have

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

which implies

$$\begin{aligned} g_k^T d_k &\leq -\|g_k\|^2 + \frac{\|g_k\|^2 \cdot g_k^T d_{k-1}}{\|d_{k-1}\|^2} \\ &\leq -\|g_k\|^2 + \sigma \|g_k\|^2 \\ &= -(1 - \sigma) \|g_k\|^2, \end{aligned}$$

which follows from (1.13), and thus, completes the prove. \square

3. Results and discussions

This section reports some preliminary numerical experiments based on Algorithm 2.1. We considered some unconstrained optimization benchmark problems for testing the proposed SRMIL+ and TTSMIL+ conjugate gradient methods and compared their performance RMIL algorithm, RMIL+ algorithm, and classical FR conjugate gradient method. The problems have been implemented using dimensions ranging from $2 \leq n \leq 10,000$ to illustrates the robustness of the proposed methods. The iterations are terminated when the stopping condition $\|g_k\| \leq 10^{-6}$ is satisfied.

1. SRMIL+ is a variant of RMIL+ satisfying the descent condition and its global convergence properties was established using exact line search in Section 2.
2. TTSMIL+ method is an extension of SRMIL+ possessing the descent condition, under SWP line search.
3. The RMIL+ is a modification of RMIL method satisfying the descent property and converge globally under exact and SWP line searches [5, 20, 24].
4. The FR method is among the most efficient and famous CG methods satisfying the descent condition and converge globally under both exact and SWP line searches [1, 31].

All codes were written on MATLAB (R2015a) and run on a CoreI5 PC with 2.50 GHz processor and 4 GB RAM using windows 7 operating systems. The comparison is based on number of iterations (NOI) and CPU time (in seconds). In some cases, the iteration terminates when the line search fails to find a positive step-size, and that point is denoted as failed point. The set of test problems, their initial points, and dimensions used to test the performance of both SRMIL+ and the TTSMIL+ are presented in Table 1 below. The dimension of the test problems is denoted as DIM.

The performance results are presented as follows:

1. The first part of the result would analyze the performance of SRMIL+ method and the results would be compared with that of the classical methods of RMIL [24], RMIL+ [5], and FR [31] method under exact minimization rule.
2. The second part of the results would analyze the performance of the TTSMIL+ and the results would be compared with that of the proposed SRMIL+, the FR [1] method, and RMIL+ [20] methods under SWP line search.

The performance results of SRMIL+ for number of iteration and CPU time are presented in Figures 1 and 2, respectively using the performance profile introduced by Dolan and More [8].

Using Dolan and More performance profile, we compare and evaluate the performance of SRMIL+, RMIL, RMIL+, and FR algorithms under exact line search. Suppose n_s solvers and n_p problems exists, for every problem p and solver s , Dolan and More defined

$\vartheta_{p,s}$ = Computation time (NO.IT. or CPU time) necessary to solve problem p by solver s .

Table 1: List of test functions.

No	Function	Dim(n)	Initial Points
1	Three Hump Camel	2	(0.5,0.5), (5,5), (10,10), (15,15)
2	Booth	2	(2,2), (9,9), (10,10), (13,13)
3	Zettl	2	(0,0), (0.5,0.5), (2,2), (10,10)
4	Treccani	2	(2,2), (9,9), (10,10), (13,13)
5	DQDRTIC	2	(2,2), (9,9), (10,10), (13,13)
6	STAIRCASE S1	2	(0,0), (10,10), (20,20), (50,50)
7	Trig double Bordered	2	(-1,-1), (10,10), (17,17), (50,50)
8	Raydan	2,4	(1,1), (9,9), (10,10), (13,13)
9	Ext DENCHNB	2,4	(5,5), (10,10), (20,20), (50,50)
10	Diagonal 1	2,4	(5,5), (10,10), (20,20), (50,50)
11	Power	2,4,10	(1,1,...,1), (9,9,...,9), (10,10,...,10), (13,13,...,13)
12	Hager	2,4,10	(1,1,...,1), (-1,-1,...,-1), (-10,-10,...,-10), (10,-10,...,-10)
13	Sum Square	2,4,10	(-3,-3,...,-3),(5,5,...,5),(13,13,...,13),(21,...,21)
14	Arwhead	2,4,10	(1,1,...,1), (9,9,...,9), (10,10,...,10), (13,13,...,13)
15	Ext Trigonometric	2,4,10,100	(0.2,0.2,...,0.2)(-1,-1,...,-1)(10,10,...,10),(13,13,...,13)
16	Shallow	2,4,10,100	(0,0,...,0), (2,2,...,2), (10,10,...,10), (13,13,...,13)
17	Sphere	2,4,10,100	(2,2,...,2), (9,9,...,9), (10,10,...,10), (15,15,...,15)
18	Ext Tridiagonal 1	2,4,10,100,1000	(2,2,...,2), (9,9,...,9), (10,10,...,10), (15,15,...,15)
19	Extended Beale	2,4,10,100,1000	(0.5,0.5,...,0.5),(-1,...,-1), (1,1,...,1), (10,10,...,10)
20	Fletcher	2,4,10,100,1000	(2,2,...,2), (3,3,...,3), (9,9,...,9), (10,10,...,10)
21	Ext White and Holst	2,4,10,100,1000	(0,0,...,0), (2,2,...,2), (10,10,...,10), (13,13,...,13)
22	Gen Tridiagonal	2,4,10,100,1000	(0,0,...,0), (2,2,...,2), (6,6,...,6), (9,9,...,9)
23	Diagonal 4	2,4,10,100,1000	(2,2,...,2), (9,9,...,9), (10,10,...,10), (15,15,...,15)
24	NONSCOMP	2,4,10,100,1000	(2,2,...,2), (9,9,...,9), (10,10,...,10), (15,15,...,15)
25	Ext Quadratic QP2	2,4,10,100,1000	(2,2,...,2), (9,9,...,9), (10,10,...,10), (15,15,...,15)
26	Quartic	2,4,10,100,1000	(2,2,...,2), (9,9,...,9), (10,10,...,10), (15,15,...,15)
27	Gen Quartic 1	2,4,10,100,1000,10000	(1,1,...,1), (0.5,0.5,...,0.5), (5,5,...,5), (10,10,...,10)
28	Three ExpTerms	2,4,10,100,1000,10000	(0,0,...,0), (10,...,10), (20,...,20), (50,...,50)
29	Ext Rosenbrock	2,4,10,100,1000,10000	(0,0,...,0), (2,2,...,2), (10,10,...,10), (15,15,...,15)
30	Ext Himmelblau	2,4,10,100,1000,10000	(-1,-1,...,-1),(1,1,...,1),(10,10,...,10),(20,20,...,20)

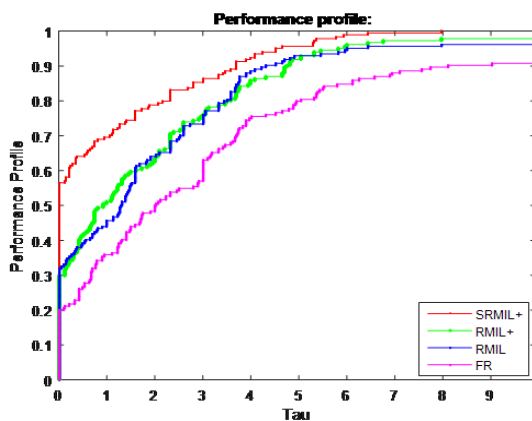


Figure 1: Performance Profile of SRMIL+ based on Number of Iterations.

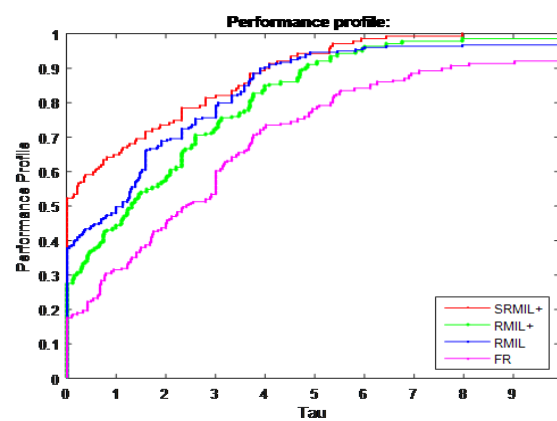


Figure 2: Performance Profile of SRMIL+ based on CPU Time.

For each algorithm, the performance profile plots the segment P of any given problem such that the method is in the neighborhood of a factor of τ of the fastest time. The uppermost curve indicates the method with the best performance. That is, the method that solved nearly all the test problems within

the shortest time.

From Fig 1 and Fig 2 we can see that the proposed SRMIL+ method performs much better than the FR method and slightly better than RMIL+ and RMIL both in terms of NOI Fig 1 and CPU time Fig 2.

The performance results of TTSRMIL+ are presented for both NOI in Fig 3 and CPU time in Fig 4, respectively.

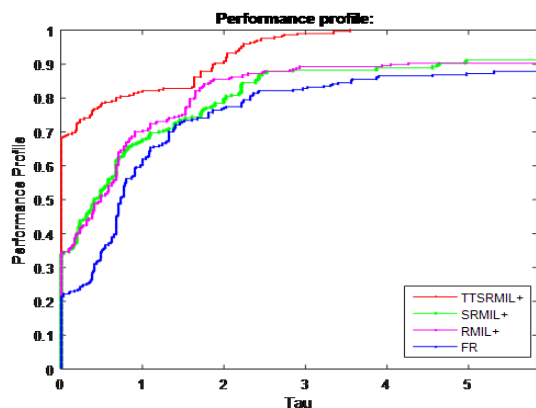


Figure 3: Performance Profile of TTSRMIL+ based on Number of Iteration.

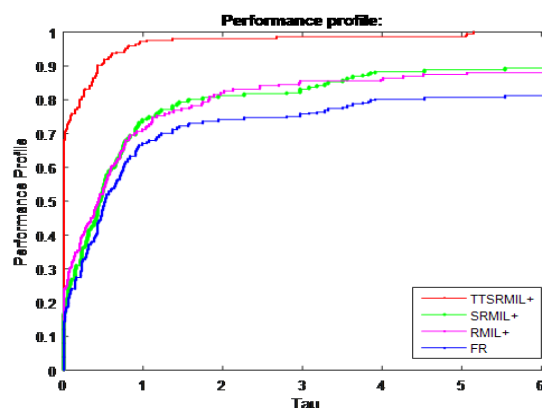


Figure 4: Performance Profile of TTSRMIL+ based on CPU Time.

For each figure, the left side measures the efficiency while the right side measures the robustness of each method.

The plot shows that propose TTSRMIL+ method performed better than the proposed SRMIL+ algorithm, the FR algorithm, and the RMIL+ algorithms both in terms of NOI Fig 3 and CPU time Fig 4.

The SRMIL+ method behaves more like the RMIL+ method. However, TTSRMIL+ has the least number of iteration and CPU time under strong Wolfe line search as can be seen on the left side of Fig 3 and Fig 4. More so, from Fig 1 and Fig 2, SRMIL+ lies above RMIL+ method, FR method, and RMIL method respectively, both under number of iterations and CPU time under exact line search. This show that the proposed SRMIL+ and TTSRMIL+ are efficient and promising. Thus, can be used in practical applications.

4. Conclusion

The CG algorithms plays an important part in unconstrained optimization, however, the three-term CG algorithms represent some stimulating numerical innovation extending the classical CG coefficient to produce efficient variants of CG methods.

This paper proposed a modification of RMIL+ method named SRMIL+ for optimization models. An interesting feature of the proposed algorithm is that would reduce to the classical RMIL+ under exact minimization condition. Also, SRMIL+ satisfies the descent condition and converge globally under exact minimization rule. The proposed SRMIL+ was further extended to construct a three-term method called TTSRMIL+. The idea of TTSRMIL+ was based on the structure of the TTPRP algorithm. The convergence analysis of TTSRMIL+ was studied under the SWP line search.

Numerical computations using some standard unconstrained optimization bench mark problems have been presented to illustrates the efficiency of the proposed methods. The results obtained show that SRMIL+ and TTSRMIL+ algorithms are efficient and faster than the methods of FR, RMIL, and RMIL+ under both exact and strong Wolfe line search. This has shown that the proposed methods can be used for practical applications and also, alternatives for solving unconstrained optimization problems

Numerous variants of the three-term CG algorithms can be explored using the proposed search directions computed as

$$d_0 = -g_0, \quad d_k = -g_k + \beta_k d_{k-1} + \gamma_k \theta_{k-1},$$

with β_k defined as SRMIL+. More so, the convergence of the proposed SRMIL+ was established under the exact line search, thus, more studies can be done on the convergence of SRMIL+ under different line searches.

Acknowledgment

The authors would like to thank the government of Malaysia for funding this research under grant no FRGS/1/2017/STGO6/Unisza/01/1.

References

- [1] M. Al-Baali, *Descent property and global convergence of the Fletcher-Reeves method with inexact line search*, IMA J. Numer. Anal., **5** (1985), 121–124. 1, 4, 3
- [2] A. Y. Al-Bayati, W. H. Sharif, *A new three-term non-linear conjugate gradient method for unconstrained optimization*, Can. J. Sci. Eng. Math., **1** (2010), 108–124. 1
- [3] N. Andrei, *A simple three-term conjugate gradient algorithm for unconstrained optimization*, J. Comput. Appl. Math., **241** (2013), 19–29. 1
- [4] E. M. L. Beale, *A derivative of conjugate gradients*, in: Numerical Methods for Nonlinear Optimization, Academic Press, London, **1972** (1972), 39–43. 1
- [5] Z. Dai, *Comments on a new class of nonlinear conjugate gradient coefficients with global convergence properties*, Appl. Math. Comput., **267** (2016), 297–300. 1, 1, 2, 3, 3
- [6] Y. H. Dai, Y. Yuan, *Convergence properties of the Fletcher-Reeves method*, IMA J. Numer. Anal., **16** (1996), 155–164. 1
- [7] Y. H. Dai, Y. Yuan, *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM J. Optim., **10** (2000), 177–182. 1, 1
- [8] E. D. Dolan, J. J. Moré, *Benchmarking optimization software with performance profiles*, Math. Program. **91** (2002), 201–213. 3
- [9] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, New York, (1989). 1, 1
- [10] R. Fletcher, C. M. Reeves, *Function minimization by conjugate gradients*, Comput. J., **7** (1964), 149–154. 1
- [11] P. L. Ghazali, I. Mohd, M. Mamat, W. M. A. W. Ahmad, *Mathematical Modelling in Family Takaful*, J. Appl. Sci., **11** (2011), 3381–3388. 1
- [12] M. R. Hestenes, E. L. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. National Bureau Standards, **49** (1952), 409–436. 1
- [13] K. Kamfa, M. Y. Waziri, I. M. Sulaiman, M. Mamat, S. S. Abas, *An Efficient Hybrid BFGS-CG Search Direction for Solving Unconstrained Optimization Problems*, J. Adv. Res. Dyn. Control Syst., **12** (2020), 1035–1041. 1, 2.1.1
- [14] J. K. Liu, Y. M. Feng, L. M. Zou, *Some three-term conjugate gradient methods with the inexact line search*, Calcolo, **55** (2018), 1–16. 1
- [15] G. Liu, J. Han, H. Yin, *Global convergence of the Fletcher-Reeves algorithm with an inexact line search*, Appl. Math. J. Chinese Univer., **10** (1995), 75–82. 1
- [16] Y. Liu, C. Storey, *Efficient generalized conjugate gradient algorithms*, J. Optim. Theory Appl., **69** (1991), 129–137. 1
- [17] J. K. Liu, Y. X. Zhao, X. L. Wu, *Some three-term conjugate gradient methods with the new direction structure*, Appl. Numer. Math., **150** (2020), 433–443. 1
- [18] M. Mamat, I. M. Sulaiman, M. Malik, F. Sukono, Z. A. Zakaria, *An Efficient Spectral Conjugate Gradient Parameter with Descent Condition for Unconstrained Optimization*, J. Adv. Res. Dyn. Control Syst., **12** (2020), 2487–2493. 1
- [19] M. F. McGuire, P. Wolfe, *Evaluating a restart procedure for conjugate gradients*, IBM Research Center, Yorktown Heights, (1973). 1
- [20] O. O. Y. Osman, *The convergence properties of RMIL+ conjugate gradient method under the strong Wolfe line search*, Appl. Math. Comput., **367** (2020), 18 pages. 1, 3, 3
- [21] E. Polak, G. Ribiere, *Note sur la convergence de directions conjuguées*, Rev. Francaise Informat Recherche Opertionelle, 3e annee, **16** (1969), 35–43. 1, 1, 1
- [22] B. T. Polyak, *The conjugate gradient method in extremal problems*, USSR Comput. Math. Math. Phys., **9** (1969), 94–112. 1
- [23] M. J. D. Powell, *Nonconvex Minimization Calculations and the Conjugate Gradient Method*, in: Numerical analysis, Springer-Verlag, Berlin, **1984** (1984), 122–141. 1, 1
- [24] M. Rivaie, M. Mamat, W. J. Leong, M. Ismail, *A new class of nonlinear conjugate gradient coefficients with global convergence properties*, Appl. Math. Comput., **218** (2012), 11323–11332. 1, 3, 3

- [25] I. M. Sulaiman, M. Mamat, A. Abashar, M. Rivaie, S. M. Fatma, Z. Salleh, *The global convergence properties of an improved conjugate gradient method*, Appl. Math. Sci., **9** (2015), 1857–1868. 1
- [26] I. M. Sulaiman, M. Mamat, A. O. Umar, K. Kamfa, E. N. Madi, *Some Three-Term Conjugate Gradient Algorithms with Descent Condition for Unconstrained Optimization Models*, J. Adv. Res. Dyn. Control Syst., **12** (2020), 2494–2501. 1
- [27] I. M. Sulaiman, S. Supian, M. Mamat, *New class of hybrid conjugate gradient coefficients with guaranteed descent and efficient line search*, IOP Conference Series: Materials Science and Engineering, **2019** (2019), 7 pages. 2
- [28] I. M. Sulaiman, U. A. Yakubu, M. Mamat, *Application of spectral conjugate gradient method for Solving unconstrained optimization problems*, Int. J. Optim. Control Theor. Appl., **10** (2020), 198–205. 1, 1
- [29] L. Zhang, W. Zhou, D. Li, *A descent modified Polak-Ribiere-Polyak conjugate gradient method and its global convergence*, IMA J. Numer. Anal., **26** (2006), 629–640. 1, 1, 2.1.1, 2.2
- [30] L. Zhang, W. Zhou, D. Li, *Some descent three-term conjugate gradient methods and their global convergence*, Optim. Methods Soft., **22** (2007), 697–711. 1, 1
- [31] G. Zoutendijk, *Nonlinear programming, computational methods*, in: Integer and Nonlinear Programming, North-Holland, Amsterdam, **1970** (1970), 37–86. 1, 2.1.2, 2.1.2, 4, 3