International Journal of Engineering & Technology, 7 (3.28) (2018) 329-334

International Journal of Engineering & Technology

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

Research paper

An Efficient Modified Hestenes-Steifel Conjugate Gradient Method with Global Convergence Properties Under Strong Wolfe Powell Line Search for Large–Scale Unconstrained Optimization

Talat Alkhouli[1,2], Mustafa Mamat[1]*, Mohd Rivaie[3], Sukono[4], Mohamad Afendee Mohamed[1], Puspa Liza Ghazali[5]

[1]Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, 22200, Terengganu, Malaysia
[2]Al-Balqa Applied University, Jordan
[3]Department of Computer Science and Mathematics, Univesiti Technology MARA, 23000 Terengganu, Malaysia
[4]Fakultas MIPA, Universitas Padjadjaran, Bandung Indonesia
[5]Faculty of Economy and Management Science, Universiti Sultan Zainal Abidin, 21030 Terengganu, Malaysia
**Corresponding author E-mail: must@unisza.edu.my*

Abstract

Conjugate gradient method has been widely used for solving unconstrained optimization and famously known for its low memory requirements and global convergence properties. Many works have been directed towards improving this method. In this paper, we propose a superior conjugate gradient coefficient β_k by revising the already proven Hestenes-Steifel formula. Theoretical proofs show that the new method fulfils sufficient descent condition if strong Wolfe-Powell inexact line search is used. Moreover, the numerical results showed that the proposed method has outclassed the exiting CG coefficients operating under standard test set. The numerical results also showed that the new formula for β_k performs significantly better than that of the original Hestenes-Steifel.

***Keywords**: unconstrained optimization; conjugate gradient; strong Wolfe–Powell line search; Hestenes-Steifel formula; global convergence.*

1. Introduction

Conjugate gradient (CG) method is a line search algorithm mostly known for its wide application in solving unconstrained optimization problems. Its low memory requirements and global convergence properties makes it one of the most preferred method in real life application such as in economics, engineering and physics. They are designed to solve problem of the form

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

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 generates a sequence by using the recurrence relation

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

where α_k is step size obtained by line search, x_k is the current iterate and d_k is the search direction defined by

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

where the scalar β_k is known as the conjugate gradient coefficient. Until now, many choices of β_k are available. Each of them shows different results when applied on unconstrained optimization functions.

In this paper, the strong Wolfe line search – an inexact line search - is used to calculate the step size in which the α_k should satisfy the two following conditions

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k \tag{4}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \leq \sigma |g_k^T d_k| \tag{5}$$

where $0 < \delta \leq \sigma < 1$.
Originally in 1952, Hestenes and Stiefel (HS) [1] proposed the initial conceptual of CG such that

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

In the following year, in 1964, another CG method proposed Fletcher and Reeves (FR) [2].

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}$$

For further improvement, in 1969, another CG method was invented by Polak and Ribiere (PRP) [3, 4].

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

Not long after, Fletcher introduced what is called Conjugate Descent (CD) [5] in 1987.

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

Afterward, in 1991, Liu and Storey (LS) [6] relied on the PRP method to propose robustly their own β_k which is

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

Dai and Yuan (DY) [7] exploited that in 1999 to extended the LS to become

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

The study of convergence performance for the aforementioned formulas with some line search conditions are widely available (see Zoutendijk [8], Powell [9-11], Wei [12], Dai [13], Al-Baali [19], Li and Feng [20] Dai and Yuan [25], Yuan and Sun [26]).
However, the question arise over the issues of global convergence of PRP, LS and HS methods which was never been proven under all stated line searches. Obviously, this is due to the fact that the descent in the value of the objective function on each iteration cannot be guaranteed (Hager and Zhang [27]). See also Yuan and Wei [28] and Andrei [29] for much detail.
During the last decade, many variations of the classical conjugate gradient methods were brought forward. Mamat et al. [30] with a bold step proposed a simple conjugate gradient parameter method and the succeeding algorithm known as Eigen Conjugate Gradient (ECG) was proved to possess the global convergence properties using the strong Wolfe line search.
In this paper, we streamline our works into working with the HS method. Based on manual calculations, the HS method is shown to outperform most if not all other simple conjugate gradient methods.
By modifying the HS method, Hager and Zhang [27] proposed a new variant called CG-DESCENT method. In CG-DESCENT method, the parameter β_k holds the following properties

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

The CG-DESCENT method is known to produce sufficient descent directions. In fact, Hager and Zhang [27] showed that the CG-DESCENT method is globally convergent under Wolfe line search. Dai [15] modified HS and suggested

$$\beta_k^* = \frac{\eta\ \|g_k\|^2}{g_k^T d_{k-1} - \eta\ g_{k-1}^T d_{k-1}} \quad , \quad \eta\ \in (\sigma, 1]$$

Zhang [16] extended the *HS* method and proposed the NHS method as follows.

$$\beta_k^{NHS} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|}\,|g_k^T g_{k-1}|}{\|g_{k-1}\|^2}$$

More recently, Rivaie et al. [14] proposed a new algorithm from the original *HS* method, and be known as Modified Hestenes-Steifel method (*MHS*). For this β_k, it comes with a re-formulated denominator while maintaining the numerator as that of Hestenes-Steifel formula given by.

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

Shapiee et al. [17] depict a variation conjugate gradient coefficient which relates also to *HS* formula and present.

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

In 2017, Hamoda et al. [18] follow the method of Wei et al. [27] and suggest the formula.

$$\beta_k^{MRM} = \frac{g_k^T\left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{\|g_{k-1}\|^2 + |g_k^T d_{k-1}|}$$

2. Proposed formula for β_κ

In recent years, much works have been directed towards developing a modified variant of conjugate gradient methods, of which as we mentioned earlier comes not only with strong convergence properties, but also achieved computational superiority over classical methods. As result to that, many variants of conjugate gradient algorithms were discovered. A survey by Andrei [21] has disclosed as many 40 nonlinear conjugate gradient algorithms for unconstrained optimization.
Wei et al. [22] presented a modified version of the PRP method famously known as the WYL method. From the idea of WYL, Zhang [23] proposed a new conjugate gradient method, to be known as NPRP, and he emphasized that the NPRP method satisfied the descent condition and the global convergence property under strong Wolfe line search. Moreover, Dai and Wen [24] further improved the NPRP method to another variant called DPRP method.
In this section, influenced by aforementioned concepts [14, 16, 22], we propose our β_k which identified as β_k^{TMR1}, where $TMR1$ indicates Tala't, Mustafa and Rivaie. The new β_k^{TMR1} is a modification of HS conjugate gradient method as follow:

$$\beta_k^{TMR1} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|}\,|g_k^T g_{k-1}|}{d_{k-1}^T\ (g_k - g_{k-1})} \qquad (6)$$

The representation algorithm is as in Algorithm 2.1.

Algorithm 2.1
S1: Initialization. Given $x_0 \in R^n, \varepsilon \geq 0,$ *set* $d_0 = -g_0$ *if* $\|g_0\| \leq \varepsilon$ then stop.
S2: Compute α_k by Strong Wolfe-Powell line search.
S3: Let $x_{k+1} = x_k + \alpha_k d_k$, $g_{k+1} = g(x_{k+1})$ *if* $\|g_{k+1}\| \leq \varepsilon$ then stop.
S4: Compute β_k by (6), and generate d_{k+1} by (2).
S5: Set $k = k + 1$ go to S2.

3. Analysis of Convergence Properties

We begin this section by examining the convergent properties of β_k^{TMR1}. To converge, an algorithm must fulfill an adequate descent condition and the global convergence properties.
The following assumptions are often required to establish the convergent properties of any newly proposed formula.

Assumption 1

$f(x)$ is function which is bounded below on the level set $\Omega = \{x \epsilon R^n : f(x) \leq f(x_0)\}$, that is a positive constant M exists such that $\|x\| \leq M$, $\forall x \in \Omega$.

Assumption 2

For some neighborhood N of Ω, with continuously differentiable objective function, its gradient $g(x)$ is Lipschitz continuous in N, i. e:

$$\exists\ L > 0 \ \ s.t \ \ \|g(x) - g(y)\| \leq L\|x - y\| \ \ \forall\, x, y \in N$$

The sufficient descent condition is the key point in the study of the CG method, its condition is given by

$$g_k^T d_k \leq -c\|g_k\|^2 , \quad \forall k \geq 0 \tag{7}$$

where $c \in (0,1)$

Theorem 3.1.1

Consider the two sequences $\{g_k\}$ and $\{d_k\}$ which were generated by the method of form (2), (3) and (6), with the value of α_k determined by the Strong Wolfe-Powell (SWP) line search defined following to (4) and (5), if $g_k \neq 0$, then the sufficient descent condition given in (7) holds.

Proof

We firstly need to streamline our new β_k^{TMR1}, so that our convergence proof will be dramatically easier. In (6), we can see that

$$\beta_k^{TMR1} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|}|g_k^T g_{k-1}|}{d_{k-1}^T (g_k - g_{k-1})}$$

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

Hence, we obtain

$$\beta_k^{TMR1} \leq \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \tag{8}$$

Using (5) and (8)

$$|\beta_{k+1}^{TMR1} g_{k+1}^T d_k| \leq \frac{\|g_{k+1}\|^2}{\|g_k\|^2}\sigma|g_k^T d_k| \tag{9}$$

In (3)

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} = -1 + \beta_{k+1}\frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2} \tag{10}$$

Using mathematical induction, the descent property of $\{d_k\}$ can be proved. Since $g_0^T d_0 = -\|g_0\|^2 < 0$, $if\ g_0 \neq 0$, let d_i, $i = 1,2,\dots,k$ are all descent directions, that is $g_i^T d_i < 0$.

In (9)

$$\frac{\|g_{k+1}\|^2}{\|g_k\|^2}\sigma g_k^T d_k \leq \beta_{k+1}^{TMR1} g_{k+1}^T d_k \leq -\frac{\|g_{k+1}\|^2}{\|g_k\|^2}\sigma g_k^T d_k$$

$$\frac{\sigma g_k^T d_k}{\|g_k\|^2} \leq \frac{\beta_{k+1}^{TMR1} g_{k+1}^T d_k}{\|g_{k+1}\|^2} \leq -\frac{\sigma g_k^T d_k}{\|g_k\|^2}$$

In (10)

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

By iterating this procedure, in addition to the fact that $g_0^T d_0 = -\|g_0\|^2$, we obtain

$$-\sum_{j=0}^{k}\sigma^j \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -2 + \sum_{j=0}^{k}\sigma^j \tag{11}$$

Since

$$\sum_{j=0}^{k}\sigma^j < \sum_{j=0}^{\infty}\sigma^j = \frac{1}{1-\sigma}$$

In (11) can be written as

$$-\frac{1}{1-\sigma} \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -2 + \frac{1}{1-\sigma} \tag{12}$$

By making the restriction $\sigma \in (0,0.5)$, we obtain $\frac{1}{1-\sigma} < 2$. Let $c = 2 - \frac{1}{1-\sigma}$, then $0 < c < 1$, and (12) is deriveable to

$$c - 2 \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -c$$

$$g_k^T d_k \leq -c\|g_k\|^2$$

This shows that (7) holds and hence the proof is complete.

Lemma 3.1.2

Assume that Assumptions 1 and 2 hold true for any iteration method given in (1), with d_k as a descent search direction and α_k satisfies the Strong Wolfe minimization rule. Consequently, the following Zoutendijk condition also holds

$$\sum_{k=1}^{\infty}\frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty$$

which is equivalent to

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

For much detail of this proof, refer to [8].

Theorem 3.1.3

Suppose that Assumptions 1 and 2 hold, $\{x_k\}$ generated by the Algorithm 2.1, where the step size is obtained under Strong Wolfe minimization rule. Then, Lemma 3.1.2 holds for all $k \geq 0$.

Proof

According to [34], by contradiction i.e., if Theorem 3.1.3 is not true, then $\exists\ c > 0\ s.t\ \|g_k\| \geq c$.

In (3)

$$d_{k+1} + g_{k+1} = \beta_{k+1} d_k$$

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

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

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

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

In (7)

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

$$\le \frac{c^2}{\|g_k\|^2} + \frac{2c}{\|g_{k+1}\|^3}$$

Since $\|g_k\| \ge c$, we know that $\|g_{k+1}\| \to \infty$ and $\|g_k\| \to \infty$. Therefore, it follows that

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

which implies that

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

However, this is contradictory to Lemma 3.1.2. Hence, the proof is concluded.

4. Results and discussion

In order to check the efficiency of $TMR1$, we compare $TMR1$ method with the classical methods FR, HS and MHS. Table 1 shows the computational performance of R2015a MATLAB program on a set of unconstrained optimization test problems. We select randomly 33 test functions which were collected by [33].
In this test, we choose $\varepsilon = 10^{-6}$ and stopping criteria is set to $\|g_k\| \le \varepsilon$ as [32] suggested. We selected three initial points, starting with a point close to the solution point moving on to a point distant away. This is necessary for the purpose of testing the global convergence properties of the new coefficient. The dimensions *n* of 33 problems are 2, 4, 10, 100 and 1000.
For selected cases, the calculations were blocked as a result of failing to find the positive step size by the line search. Thus, it was considered a failed. In addition, unsuccessful search can also be resulted when the number of iterations exceeding 10,000. We use two parameters, the iteration number and the CPU time from experimental result as the basis for our comparison. We employ the performance profile presented by [31] to obtain the results shown in Figure 1 and Figure 2.
All experimentations were carried out on a PC running on the Windows 7 OS, powered by Intel (R) processor with CoreTM i3-M350 (2.27GHz) having memory size of 4GB.

Table 1: Problem Functions

No.	Function	Dimension	Initial Point
1	SIX HUMP	2	(0.5,0.5), (8,8), (40,40)
2	THREE HUMP	2	(-1,1),(-2,2),(2,-2)
3	LEON	2	(2,2),(4,4),(8,8)
4	QUADRATIC QF1	2	(3,3),(5,5),(10,10)
5	MATYAS	2	(5,5),(10,10),(15,15)
6	DIAGONAL 2	2	(1,1),(5,5),(15,15)
7	BOOTH	2	(10,10),(25,25),(100,100)
8	RAYDAN	2	(3,3),(13,13),(22,22)
9	ZETTL	2	(5,5),(20,20),(50,50)
10	TRECANNI	2	(5,5),(10,10),(50,50)
11	NONDIA	2	(10,10),(20,20),(35,35)
12	HAGER	2	(7,7),(15,15),(20,20)
13	EXTENDED MARATOS	2	(10,10),(60,60),(120,120)
14	EXTENDED PENALTY	2	(40,40),(80,80),(100,100)
15	GENERLIZED TRIDIAGONAL1	2	(3,3),(21,21),(90,90)
16	QUADRATIC QF2	2	(4,4),(40,40),(80,80)
17	CLOVILLE	4	(2,..,2),(4,..,4),(10,..,10)
18	EXTENDED WOOD	4	(5,..,5),(20,..,20),(30,..,30)
19	DIXON & PRICE	2 4	(6,6),(18,18),(60,60) (6,..,6),(18,..,18),(60,..,60)
20	ARWHEAD	2 10	(8,8),(24,24),(32,32) (8,..,8),(24,..,24),(32,..,32)
21	GENERRALIZED QUARTIC	2 10	(7,7),(70,70),(140,140) (7,..,7),(70,..,70),(140,..,140)
22	FLETCHCR	2 10,100,1000	(12,12),(15,15),(35,35) (12,..,12),(15,..,15),(35,..,35)
23	ROSENBROCK	2 10,100,1000	(3,3),(15,15),(75,75) (3,..,3),(15,..,15),(75,..,75)
24	SHALLOW	2 10,100,1000	(2,2),(12,12),(200,200) (2,..,2),(12,..,12),(200,..,200)
25	WHITE & HOLST	2 10,100,1000	(3,3),(6,6),(10,10) (3,..,3),(6,..,6),(10,..,10)
26	EXTENDED BEALE	2 10,100,1000	(-4,-4),(-1,-1),(4,4) (-4,..,-4),(-1,..,-1),(4,..,4)
27	PERTURBED QUADRATIC	2 10,100,1000	(1,1),(5,5),(10,10) (1,..,1),(5,..,5),(10,..,10)
28	EXTENDED TRIDIAGONAL1	2 10,100,1000	(25,25),(50,50),(75,75) (25,..,25),(50,..,50),(75,..,75)
29	DIAGONAL 4	2 10,100,1000	(1,1),(20,20),(40,40) (1,..,1),(20,..,20),(40,..,40)
30	SUM SQUARES	2 10,100,1000	(1,1),(5,5),(10,10) (1,..,1),(5,..,5),(10,..,10)
31	EXTENDED DENSCHNB	2 10,100,1000	(5,5),(30,30),(50,50) (5,..,5),(30,..,30),(50,..,50)
32	EXTENDED HIMMELBLAU	2 10,100,1000	(10,10),(50,50),(125,125) (10,.,10),(50,.,50),(125,.,125)
33	EXTENDED BLOCK DIAGONAL BD1	2 10,100,1000	(1,1),(5,5),(10,10) (1,..,1),(5,..,5),(10,..,10)

The studies in [31] on performance of the set solvers S given a test set P, serves as a basis for evaluating our method. For each problem p and solver s, suppose there exist n_s solvers and n_p problems such that

$t_{p,s}$ = time taken to solve problem p by solver s.

We compared the performance of solver *s* attending the problem p for the finest performance against other solver, based on the performance ratio parameter

$$r_{p,s} = \frac{1}{min\{t_{p,s} : s \in S\}}$$

Consider that a parameter $r_M \ge r_{p,s}\ \forall\, p,s$ is chosen such that $r_M = r_{p,s}$ if and only if the problem p is unsolvable. The performance of solver s on any given problem can be of lots of attractions, however, because we would like to gain an overall assessment of solver's performance, we further define

$$p_s(t) = \frac{1}{n_p}\{p \in P : r_{p,s} \leq t\}$$

Thus, $p_s(t)$ is the probability for solver $s \in S$ that a performance ratio $r_{p,s}$ was within a factor $t \in R$ of the best possible ratio. For the performance ratio, a function p_s represents the cumulative distribution function. The performance profile $p_s: R \rightarrow [0,1]$ for a solver is a non-decreasing, piecewise, and continuous from the right. The $p_s(1)$ defines the probability that the solver is expected to outclass the rest of other solvers. Obviously, having a high value of $P(\tau)$, represented by being at the top right of the figure shows the superiority of the best solver.

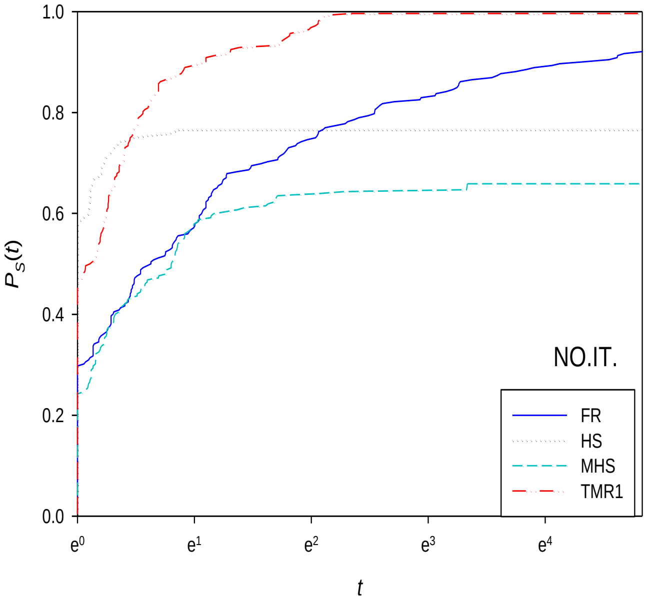

Fig. 1: Number of iterations

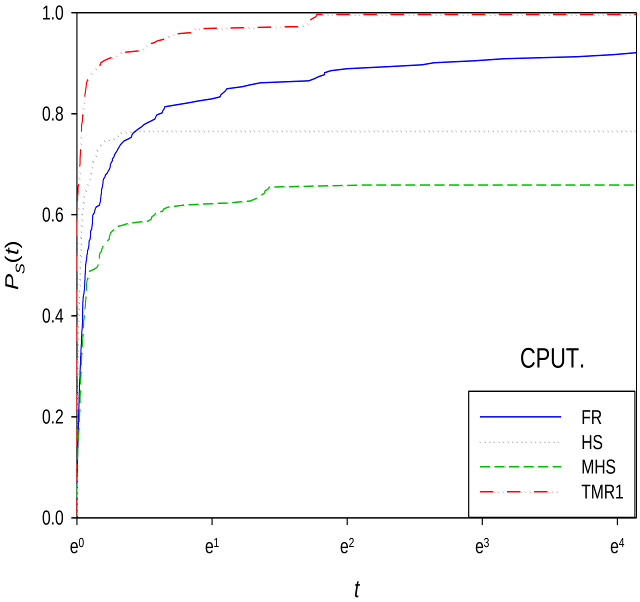

Fig. 2: Processor time

Figure 1 shows the measurement for the performance profile based on the number of iteration, whereas, Figure 2 shows its CPU time counterpart. Observe that the plots from Figure 1 and Figure 2 are nearly identical in shape. The red-colored curve indicates that the TMR1 method perform significantly well than the FR, HS and MHS methods.

From results shown in Figure 1 and Figure 2, it is obvious that TMR1 outperforms the other methods, by being able to solve all test problems and in turn reaches 100% success rate. Comparing with FR, HS and MHS that only reach 92%, 77% and 66% respectively in solving the given test problems. Hence, our new method shows its superiority and competitiveness against the other existing methods for solving unconstrained optimization problem.

5. Conclusion

This study introduces a new conjugate gradient method for unconstrained optimization problem which was shown to suffice descent condition and achieving global convergence. The whole experiment was executed under the used of strong Wolfe-Powell line search. For most cases, experimental results conclude that the $TMR1$ method is by far more efficient than others. Hence, the method we proposed not only possesses excellent global convergence but in addition is much superior than the HS conjugate gradient method.

Acknowledgement

This project is funded by the Malaysian Fundamental Research Grant Scheme (FRGS/1/2017/STG06/Unisza/01/1) and Research Management, Innovation & Commercialization Center, Universiti Sultan Zainal Abidin.

References

[1] Stiefel, E. (1952). Methods of conjugate gradient for solving linear equation. J. Res. Nat. Bur. Standards, 49, 409-436.

[2] Fletcher, R., & Reeves, C. M. (1964). Function minimization by conjugate gradients. Computer Journal, 7(2), 149-154.

[3] Polyak, B. T. (1969). The conjugate gradient method in extremal problems. USSR Computational Mathematics and Mathematical Physics, 9(4), 94-112.

[4] Polak, E., & Ribiere, G. (1969). Note sur la convergence de méthodes de directions conjuguées. Revue française d'informatique et de recherche opérationnelle. Série Rouge, 3(16), 35-43.

[5] Fletcher, R. (2013). Practical methods of optimization. John Wiley and Sons.

[6] Liu, Y., & Storey, C. (1991). Efficient generalized conjugate gradient algorithms, part 1: theory. Journal of Optimization Theory and Applications, 69(1), 129-137.

[7] Dai, Y. H., & Yuan, Y. (1999). A nonlinear conjugate gradient method with a strong global convergence property. SIAM Journal on Optimization, 10(1), 177-182.

[8] Zoutendijk, G. (1970). Nonlinear programming, computational methods. Integer and Nonlinear Programming, 1970, 37-86.

[9] Powell, M. J. D. (1977). Restart procedures for the conjugate gradient method. Mathematical Programming, 12(1), 241-254.

[10] Powell, M. J. (1984). Nonconvex minimization calculations and the conjugate gradient method. In D. F. Griffiths (Ed.), Numerical Analysis. Berlin: Springer, pp. 122-141.

[11] Powell, M. J. (1986). Convergence properties of algorithms for nonlinear optimization. Siam Review, 28(4), 487-500.

[12] Wei, Z., Li, G., & Qi, L. (2006). New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems. Applied Mathematics and Computation, 179(2), 407-430.

[13] Dai, Z. F. (2011). Two modified HS type conjugate gradient methods for unconstrained optimization problems. Nonlinear Analysis: Theory, Methods and Applications, 74(3), 927-936.

[14] Rivaie, M., Mamat, M., Mohd, I., & Fauzi, M. (2010). Modified Hestenes-Steifel conjugate gradient coefficient for unconstrained optimization. Journal of Interdisciplinary Mathematics, 13(3), 241-251.

[15] Dai, Z. F. (2011). Two modified HS type conjugate gradient methods for unconstrained optimization problems. Nonlinear Analysis: Theory, Methods and Applications, 74(3), 927-936.

[16] Zhang, L. (2009). New versions of the Hestenes-Stiefel nonlinear conjugate gradient method based on the secant condition for optimization. Computational and Applied Mathematics, 28(1), 111-133

[17] Shapiee, N., Rivaie, M., & Mamat, M. (2016). A new classical conjugate gradient coefficient with exact line search. AIP Conference Proceedings, 1739(1), 1-8.

[18] Hamoda, M., Rivaie, M., Mamat, M., & Salleh, Z. (2015). A new nonlinear conjugate gradient coefficient for unconstrained optimization. Applied Mathematical Sciences, 9, 1813-1822.

[19] 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), 121-124.

[20] Li, M., & Feng, H. (2011). A sufficient descent LS conjugate gradient method for unconstrained optimization problems. Applied Mathematics and Computation, 218(5), 1577-1586.

[21] Andrei, N. (2011). 40 conjugate gradients algorithms for unconstrained optimization. Bull. Malay. Math. Sci. Soc., 34, 319-330.

[22] Wei, Z., Yao, S., & Liu, L. (2006). The convergence properties of some new conjugate gradient methods. Applied Mathematics and Computation, 183(2), 1341-1350.

[23] Zhang, Y., Zheng, H., & Zhang, C. (2012). Global convergence of a modified PRP conjugate gradient method. Procedia Engineering, 31, 986-995.

[24] Dai, Z., & Wen, F. (2012). Another improved Wei–Yao–Liu nonlinear conjugate gradient method with sufficient descent property. Applied Mathematics and Computation, 218(14), 7421-7430.

[25] Dai, Y. H., & Yuan, Y. (2000). Nonlinear conjugate gradient methods. Shanghai Science and Technology Publisher, Shanghai.

[26] Yuan, Y., & Sun, W. (1999). Theory and methods of optimization. Science Press of China.

[27] Hager, W. W., & Zhang, H. (2005). A new conjugate gradient method with guaranteed descent and an efficient line search. SIAM Journal on Optimization, 16(1), 170-192.

[28] Yuan, G., & Wei, Z. (2009). New line search methods for unconstrained optimization. Journal of the Korean Statistical Society, 38(1), 29-39.

[29] Andrei, N. (2009). Accelerated conjugate gradient algorithm with finite difference Hessian/vector product approximation for unconstrained optimization. Journal of Computational and Applied Mathematics, 230(2), 570-582.

[30] Mamat, M., Rivaie, M., Mohd, I., & Fauzi, M. (2010). A new conjugate gradient coefficient for unconstrained optimization. Int. J. Contemp. Math. Sciences, 5(29), 1429-1437.

[31] Dolan, E. D., & Moré, J. J. (2002). Benchmarking optimization software with performance profiles. Mathematical Programming, 91(2), 201-213.

[32] Hillstrom, K. E. (1977). A simulation test approach to the evaluation of nonlinear optimization algorithms. ACM Transactions on Mathematical Software, 3(4), 305-315.

[33] Andrei, N. (2008). An unconstrained optimization test functions collection. Advanced Modeling and Optimization, 10(1), 147-161.

[34] Abashar, A., Mamat, M., Rivaie, M., Mohd, I., & Omer, O. (2014). The proof of sufficient descent condition for a new type of conjugate gradient methods. AIP Conference Proceedings, 1602(1), 296-303.