

A Robust Modification of Hestenes–Stiefel Conjugate Gradient Method with Strong Wolfe Line Search

¹Awad Abdelrahman, ¹Osman O.O. Yousif and ²Khalid I. Osman

¹Department of Mathematics, Faculty of Mathematical and Computer Sciences, University of Gezira, Wad Madani, Sudan

²Department of Mathematics, Faculty of Sciences and Art, King Khalid University, Sarat Abedah, Saudi Arabia

Article history

Received: 09-02-2020

Revised: 25-03-2020

Accepted: 13-05-2020

Corresponding Author:

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

Abstract: Nonlinear Conjugate Gradient (CG) methods are extensively used for solving large-scale unconstrained optimization problems. Numerous of studies constructed scales and modifications have been conducted recently to improve (CG) methods. In this paper, a simple modified by its conjugate gradient method was proposed. In addition to, established global convergence property and sufficient descent condition, under Strong Wolfe line search. Numerical result shows that the proposed formula is competitive when compared to other well-known (CG) parameters.

Keywords: Unconstrained Optimization, Conjugate Gradient Method, Sufficient Descent Property, Global Convergence

Introduction

The nonlinear Conjugate Gradient methods (CG) are utilized to find the minimum value of function for unconstrained optimization problems. In general, the method has the following form:

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

where, $f: R^n \rightarrow R$ is continuous differentiable nonlinear function and which gradient denotes by $g(x)$. The CG methods are given by an iterative method of the form:

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

where, x_k is the k th iterative point, the $\alpha_k > 0$ is a step-length and d_k is the called conjugate gradient search direction with:

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

where, β_k is a scalar, the step-length $\alpha_k > 0$ is obtained by attainment a one dimensional search, known as the ‘line searches’. The most common line searches are exact and inexact line searches. The inexact line search has many types of methods known as Armijo (Fletcher, 1997), Wolfe (1969), Goldstein (1965) and Strong Wolfe (Dai and Yuan, 1999; Hilstrom, 1977). In this paper, utilized strong Wolfe line search for computed α_k and defined as follows:

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + \delta \alpha_k g_k^T d_k, \\ |g(x_k + \alpha_k d_k)^T d_k| &\leq \sigma |g_k^T d_k|, \end{aligned} \quad (1.4)$$

where, $0 < \delta < \sigma < 1$ are two constants.

There are at least six well-known formulas for β_k , which are given below:

$$\beta_k^{HS} = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}} \quad (\text{Hestenes and Stiefel, 1952}),$$

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} \quad (\text{Fletcher and Reeves, 1964}),$$

$$\beta_k^{PRP} = \frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T g_{k-1}} \quad (\text{Polak and Ribiere, 1969}),$$

$$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}} \quad (\text{Conjugate descent (Fletcher, 1997)}),$$

$$\beta_k^{LS} = -\frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T g_{k-1}} \quad (\text{Liu and Storey, 1992}),$$

$$\beta_k^{DY} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}} \quad (\text{Dai and Yuan, 1999}).$$

These parameters β_k (Hestenes and Stiefel, 1952; Fletcher and Reeves, 1964; Polak and Ribiere, 1969; Hestenes and Stiefel, 1952; Liu and Storey, 1992; Dai and Yuan, 1999) are equivalent, when f is a strong convex quadratic function with an exact line search. If f is non-quadratic functional, each choice for the β_k leads to very different performance and convergence of the

corresponding algorithms (Andrei, 2011). The behavior of convergence of the β_k 's formulas with some line search conditions has been established by many authors for many years ago (Rivaie *et al.*, 2012; Dolan and More, 2002; Goldstein, 1965; Powell, 1977; Powell, 1984; Dai and Yuan, 1999; Andrei, 2011; Hilstrom, 1977; Fletcher, 1997; Fletcher and Reeves, 1964; Liu and Storey, 1992; Hestenes and Stiefel, 1952; Andrei, 2008; Zhang, 2009; Zoutendijk, 1970). Until recently, they are seeking for convenient β_k 's that are efficient in a numerical performance, possessed global convergence and sufficient decent condition. The numerical performance of the FR conjugate gradient method has often been much slower than that of the PRP conjugate gradient method and fewer cases it is faster than PRP. The global convergence of the FR method with exact line search was fulfilled by Zoutendijk (1970) and also Al-Baali (1985) established the FR method is globally convergent under strong Wolfe Condition when $\sigma < 0.5$, overtake Liu and Storey (1992) expanded that result to $\sigma \leq 0.5$. The CD and DY methods established global convergence under Strong Wolfe line search (Dai and Yuan, 1999; Hilstrom, 1977) and they have the same numerical performance as FR under exact line search. The PRP conjugate gradient method has a good numerical performance (Powell, 1977), but does not have good convergence property (Powell, 1984) and so as HS method. The global convergence of the PRP method for convex objective function under exact line search was established by Polak and Ribiere (1969). Powell (1986) proposed a counterexample and showed the non-convergent sequence of the PRP method for non-convex function, which also applied to the HS method. Different modifications of the HS method building a nice basis for an upgrade the performance and convergence property. For good news that, the convergence of the standard HS method with various inexact line search does not yet established. One important notice is that, when $\|x_k - x_{k-1}\|$ is, being small, both the nominator and denominator of β_k^{HS} become small so that β_k^{HS} might be unbounded (Dai, 2010). Qi *et al.* (1999) established the global convergence of a modified HS method, where β_k takes the form:

$$\beta_k^{OHL} = \max \left\{ 0, \min \left\{ \beta_k^{HS}, \frac{1}{\|g_k\|} \right\} \right\}.$$

Perry (1977) observed that the search direction in the HS method can be written as:

$$d_k = -P_k g_k,$$

where:

$$P_k = I - \frac{d_k y_{k-1}^T}{d_k^T y_k}.$$

Noting that $P_k^T y_{k-1} = 0$, P_k is an affine transformation that transforms R^n into the null space of $P_k^T y_{k-1}$.

This paper is organized as follows. In Section 2, is presented the underlying idea of modification and present the modification of the HS conjugate gradient method and algorithm. In Section 3, we established sufficient descent property and global convergence property with the strong Wolfe line search, if the parameter $\sigma = 0.1$. In Section 4, preliminary numerical results can be introduced.

The Modified Method

The nonlinear conjugate gradient method for unconstrained optimizations is uncomplicated and has abated memory requirement properties and is very effective for large-scale optimization problems, which the HS method is one of the most efficient methods. However, the standard HS method has nice numerical performance but fails in the convergence of non-convex functions under the inexact line search technique. Therefore, include a simple medication to overcome this deficiency, a modified HS formula is defined by:

$$\beta_k^{MN*} = \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_{k-1}\|} \|g_k g_{k-1}\|}{d_{k-1}^T \left(d_{k-1} - \frac{\|g_k\|}{\|g_{k-1}\|} g_k \right)}, \quad (2.1)$$

and $\|\cdot\|$ stands for the Euclidean norm of vectors.

The formula (2.1) has the following properties:

$$\beta_k^{MN*} \begin{cases} \leq \frac{\|g_k\|^2}{\|d_{k-1}\|^2}, & \text{if } g_k d_{k-1} \leq 0, \\ \leq \frac{\|g_k\|^2}{\|w d_{k-1}\|^2}, & \text{if } g_k d_{k-1} > 0, \end{cases} \text{, where } w = 1 - \frac{\|g_{k-1}\|}{\|d_{k-1}\|} \beta_k^{FR}. \quad (2.2)$$

Lemma 2.1

Consider any method (1.2) and (1.3) be generated by Algorithm 2.1, let $\delta > 0$ and $\mu > 0$ and $\beta_k = \beta_k^{MN*}$ is given as (2.1), then:

$$\mu - \delta \leq \|d_{k-1}\| \leq \delta + \mu$$

Proof

We can write (1.3), as:

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

then, by triangle inequality the above equation become:

$$|\beta_k| \|d_{k-1}\| = \|d_k + g_k\| \leq \|g_k\| + \|d_k\|,$$

and $|\beta_k| > 0$, dividing both sides of above inequality by $|\beta_k|$:

$$\|d_{k-1}\| \leq \frac{\|g_k\|}{|\beta_k|} + \frac{\|d_k\|}{|\beta_k|}. \quad (2.3)$$

is obtain, also $\frac{\|g_k\|}{|\beta_k|} > 0, \frac{\|d_k\|}{|\beta_k|} > 0$ and let $\delta = \frac{\|g_k\|}{|\beta_k|}, \mu = \frac{\|d_k\|}{|\beta_k|}$,

then from (2.3):

$$\|d_{k-1}\| \leq \delta + \mu. \quad (2.4)$$

Also from (1.3) and by triangle inequality, is obtain:

$$\begin{aligned} \|d_k\| &= -\|g_k + \beta_k d_{k-1}\| \leq \|g_k\| + |\beta_k| \|d_{k-1}\|, \\ \|d_k\| - \|g_k\| &\leq |\beta_k| \|d_{k-1}\|, \\ \frac{\|d_k\|}{|\beta_k|} - \frac{\|g_k\|}{|\beta_k|} &\leq \|d_{k-1}\|, \end{aligned}$$

then, $\frac{\|g_k\|}{|\beta_k|} > 0, \frac{\|d_k\|}{|\beta_k|} > 0$ and let $\delta = \frac{\|g_k\|}{|\beta_k|}, \mu = \frac{\|d_k\|}{|\beta_k|}$,

substitution in the above inequality, is obtain:

$$\mu - \delta \leq \|d_{k-1}\|. \quad (2.5)$$

The proof is complete.

Lemma 2.2

Consider any method (1.2) and (1.3) be generated by Algorithm 2.1, let and $\beta_k = \beta_k^{MN^*}$ is given as (2.1), then:

$$\begin{aligned} \frac{\|d_k\|}{\|g_k\|} &\leq 1 + \frac{M}{\mu - \delta}, \text{ if } g_k d_{k-1} \leq 0, \text{ and} \\ \frac{\|d_k\|}{\|g_k\|} &\leq 1 + \frac{M}{w(\mu - \delta)}, \text{ if } g_k d_{k-1} > 0 \end{aligned} \quad (2.6)$$

Proof

From (2.2), if $g_k d_{k-1} \leq 0$, (1.3), $\beta_k = \beta_k^{MN^*}$, (2.5) and triangle inequality, we have:

$$\begin{aligned} d_k &= -g_k + \beta_k d_{k-1}, \\ 0 &\leq \|d_k\| \leq \|g_k\| + |\beta_k| \|d_{k-1}\|, \\ 0 &\leq \|d_k\| \leq \|g_k\| + \frac{\|g_k\|^2}{\|d_{k-1}\|^2} \|d_{k-1}\| = \|g_k\| + \frac{\|g_k\|^2}{\mu - \delta} \end{aligned}$$

dividing above inequality by $\|g_k\|$ and from (3.2), we obtain:

$$0 \leq \frac{\|d_k\|}{\|g_k\|} \leq 1 + \frac{M}{\mu - \delta}. \quad (2.7)$$

From (2.2), if $g_k d_{k-1} > 0$, (1.3), $\beta_k = \beta_k^{MN^*}$, (2.5) and triangle inequality, we have:

$$\begin{aligned} 0 &\leq \|d_k\| \leq \|g_k\| + |\beta_k| \|d_{k-1}\|, \\ 0 &\leq \|d_k\| \leq \|g_k\| + \frac{\|g_k\|^2}{w\|d_{k-1}\|^2} \|d_{k-1}\| = \|g_k\| + \frac{\|g_k\|^2}{w(\mu - \delta)}, \end{aligned}$$

dividing above inequality by $\|g_k\|$ and from (3.2), we obtain:

$$0 \leq \frac{\|d_k\|}{\|g_k\|} \leq 1 + \frac{M}{w(\mu - \delta)}. \quad (2.8)$$

The proof is complete.

The Algorithm 2.1 of the $\beta_k^{MN^*}$ method is presented.

Algorithm 2.1

Step 1: Initialization. Given $x_0 \in R^n, d_0 = -g_0, k = 0$, if $g_0 = 0$ then stop.

Step 2: Compute β_k based on (2.1).

Step 3: Compute d_k based on (1.3), if $\|g_k\| \leq \varepsilon$, then stop; otherwise go to the next step.

Step 4: Compute $\alpha_k > 0$ based on (1.4)

Step 5: Updating new point based on (1.2), if $\|g_k\| \leq \varepsilon$, then stop

Step 6: Set $k = k + 1$ and go to step 3.

Convergence Analysis

In this section, we will analyze and study the convergence properties of $\beta_k^{MN^*}$.

Sufficient Descent Conditions

Before giving the sufficient descent conditions, are needed the following assumptions.

Assumption A

- (i) The $f(x)$ is bounded below on the level set R^n and continuous, differentiable in a neighborhood N of the level set $\ell(x_0) = \{x \in R^n | f(x) \leq f(x_0)\}$ at the initial point x_0
- (ii) The gradient $g(x) = \nabla f(x)$ is Lipschitz continuous in N , so a constant $L > 0$ exists, such that:

$$\|g(x) - g(y)\| \leq \|Lx - y\|, \text{ for any } x, y \in N. \quad (3.1)$$

By using the Assumption A and there exists $M > 0$ such that:

$$\|g(x)\| \leq M \quad \forall x \in \ell. \quad (3.2)$$

Theorem 3.1

Consider any method (1.2) and (1.3) be generated by Algorithm 2.1, Assumption A holds and let the step-length α_k be determined by the strong Wolfe line search (1.4), where $\beta_k = \beta_k^{MN^*}$ is given as (2.1), then for all $k \geq 0$:

$$g_k^T d_k \leq -c \|g_k\|^2, \text{ holds} \quad (3.3)$$

Proof

If $k = 0$, then $g_0^T d_0 = -c \|g_0\|^2$, hence, condition (3.3) holds true. We also need to show that for $k \geq 1$, condition (3.3), that is will also hold true.

If $g_k^T d_{k-1} \leq 0$, from (1.3) multiply by g_k and second inequality of (1.4), we get:

$$g_k^T d_k = -\|g_k\|^2 + \beta_k^{MN^*} g_k^T d_{k-1} \leq -\|g_k\|^2 + |\beta_k^{MN^*}| \sigma \|g_{k-1}\| \|d_{k-1}\|. \quad (3.4)$$

From (1.4), (2.2), (3.4), we get:

$$g_k^T d_k \leq -\|g_k\|^2 + \frac{\|g_k\|^2}{\|d_{k-1}\|^2} \sigma \|g_{k-1}\| \|d_{k-1}\| = -\|g_k\|^2 \left(1 - \frac{\sigma \|g_{k-1}\|}{\|d_{k-1}\|} \right). \quad (3.5)$$

Substitution (2.5) in to (3.5) and (3.2), is obtained:

$$g_k^T d_k \leq -\|g_k\|^2 \left(1 - \frac{\sigma M}{\mu - \delta} \right). \quad (3.6)$$

If $g_k^T d_{k-1} > 0$, from (1.3) multiply by g_k and from (1.4), is obtained:

$$g_k^T d_k = -\|g_k\|^2 + \beta_k^{MN^*} g_k^T d_{k-1} \leq -\|g_k\|^2 + \frac{\|g_k\|^2}{w \|d_{k-1}\|^2} \sigma \|g_{k-1}\| \|d_{k-1}\| = -\|g_k\|^2 \left(1 - \frac{\sigma \|g_{k-1}\|}{w \|d_{k-1}\|} \right). \quad (3.7)$$

Substitution (2.5) in to (3.7) and (3.2), is obtained:

$$g_k^T d_k \leq -\|g_k\|^2 \left(1 - \frac{\sigma M}{w(\mu - \delta)} \right). \quad (3.8)$$

from (3.6), (3.8), is obtained (3.3). And this completes the proof of Theorem.

Global Convergence Properties

The following lemma, called the Zoutendijk condition, is usually used to prove global convergence of CG method. It was given by Zoutendijk (1970).

Lemma 3.2

Suppose that x_0 is an initial point for which Assumption A holds. Consider any method in the form (1.2) and (1.3), where d_k is a descent direction and α_k satisfies the strong Wolf condition line search (1.4). Then we have:

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \quad (3.9)$$

Proof

We get from the second inequality of (1.4):

$$\sigma g_{k-1}^T d_{k-1} \leq g_k^T d_{k-1} \leq -\sigma g_{k-1}^T d_{k-1}. \quad (3.10)$$

This together with the Lipchitz condition (3.1), implies:

$$(\sigma - 1) g_{k-1}^T d_{k-1} \leq (g_k - g_{k-1})^T d_{k-1} \leq L \alpha_{k-1} \|d_{k-1}\|^2,$$

then:

$$\alpha_{k-1} \geq \frac{(\sigma - 1) g_{k-1}^T d_{k-1}}{L \|d_{k-1}\|^2} = \frac{(1 - \sigma) |g_{k-1}^T d_{k-1}|}{L \|d_{k-1}\|^2}. \quad (3.11)$$

Based on the Assumption A, we get from the first inequality of (1.4) that:

$$-\sum_{k=0}^{\infty} \alpha_{k-1} g_{k-1}^T d_{k-1} \leq +\infty. \quad (3.12)$$

From (3.12) and (3.11), the Zoutendijk condition (3.9) holds.

Lemma 3.2 is equivalent to the following inequality:

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \quad (3.13)$$

Now, the global convergence Theorem of the formula $\beta_k^{MN^*}$ is established.

Theorem 3.2

Suppose that the Assumption A holds. Consider any CG method in the form of (1.2) and (1.3), where α_k is obtained by the strong Wolfe line search (1.4). Also, the descent condition holds. Then:

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \quad (3.14)$$

Proof

To prove Theorem 3.2, contradiction is used. That is, if Theorem 3.2 is not true, then a constant $\epsilon > 0$ exists, such that:

$$\|g_k\| \geq \epsilon \tag{3.15}$$

Rewriting (1.3) as:

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

and squaring both sides of the equation, we obtain:

$$\|d_k\|^2 = -\|g_k\|^2 - 2g_k^T d_k + (\beta_k^{MN^*})^2 \|d_{k-1}\|^2. \tag{3.16}$$

From (2.2), if $g_k^T d_{k-1} \leq 0$ and an Equation (3.16), we get:

$$\begin{aligned} \|d_k\|^2 &\leq -\|g_k\|^2 - 2g_k^T d_k + \left(\frac{\|g_k\|^2}{\|d_{k-1}\|^2}\right)^2 \|d_{k-1}\|^2, \\ \|d_k\|^2 &\leq -\|g_k\|^2 - 2g_k^T d_k + \frac{\|g_k\|^4}{\|d_{k-1}\|^2}. \end{aligned} \tag{3.17}$$

Dividing both sides of (3.17) by $\|g_k\|^4$ and from (2.5), Cauchy-inequality, is obtained:

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leq \frac{-1}{\|g_k\|^2} - \frac{2g_k^T d_k}{\|g_k\|^4} + \frac{1}{\|d_{k-1}\|^2} \leq \frac{-1}{\|g_k\|^2} + \frac{2\|d_k\|}{\|g_k\|^3} + \frac{1}{(\mu - \delta)^2},$$

and from (2.7), is obtained:

$$\begin{aligned} \frac{\|d_k\|^2}{\|g_k\|^4} &\leq \frac{-1}{\|g_k\|^2} + \frac{1}{\|g_k\|^2} \left(1 + \frac{M}{\mu - \delta}\right) + \frac{1}{(\mu - \delta)^2}, \\ \frac{\|d_k\|^2}{\|g_k\|^4} &\leq \frac{1}{\|g_{k+1}\|^2} \left(\frac{M + \mu - \delta}{\mu - \delta}\right) + \frac{1}{(\mu - \delta)^2}, \end{aligned}$$

from above inequality, we take $\vartheta = \frac{M + \mu - \delta}{\mu - \delta}$, $\gamma = (\mu - \delta)^2$, then is obtained:

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leq \frac{1}{\|g_{k+1}\|^2} \vartheta + \frac{1}{\gamma} = \frac{1}{\epsilon^2} \vartheta + \frac{1}{\gamma} = \frac{\epsilon^2 + \vartheta \gamma}{\gamma \epsilon^2},$$

suppose that the (3.14) does not hold. Then, there exists $\epsilon > 0$, such that (3.15) holds for all $k \geq 0$. Then the above inequality become:

$$\frac{\|d_k\|^2}{\|g_k\|^4} \geq \frac{\epsilon^2 \gamma}{\epsilon^2 + \vartheta \gamma}. \tag{3.18}$$

Therefore, from (3.18) and (3.13), it follows that:

$$\begin{aligned} \sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} &\geq \sum_{k=0}^{\infty} \frac{\epsilon^2 \gamma}{\epsilon^2 + \vartheta \gamma} = \frac{\epsilon^2 \gamma}{\epsilon^2 + \vartheta \gamma} \sum_{k=0}^{\infty} 1, \text{ therefore,} \\ \sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} &\geq \infty. \end{aligned} \tag{3.19}$$

From (2.2), if $g_k^T d_{k-1} > 0$ and an equation (3.16), we get:

$$\begin{aligned} \|d_k\|^2 &\leq -\|g_k\|^2 - 2g_k^T d_k + \left(\frac{\|g_k\|^2}{w\|d_{k-1}\|}\right)^2 \|d_{k-1}\|^2, \\ \|d_k\|^2 &\leq -\|g_k\|^2 - 2g_k^T d_k + \frac{\|g_k\|^4}{w^2 \|d_{k-1}\|^2}, \end{aligned}$$

dividing both sides of the above inequality by $\|g_k\|^4$ and from (2.5), Cauchy-inequality, is obtained:

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leq \frac{-1}{\|g_k\|^2} - \frac{2g_k^T d_k}{\|g_k\|^4} + \frac{1}{\|d_{k-1}\|^2} \leq \frac{-1}{\|g_k\|^2} + \frac{2\|d_k\|}{\|g_k\|^3} + \frac{1}{w^2 (\mu - \delta)^2},$$

then from above inequality and (2.8), is obtained:

$$\begin{aligned} \frac{\|d_k\|^2}{\|g_k\|^4} &\leq \frac{-1}{\|g_k\|^2} + \frac{1}{\|g_k\|^2} \left(1 + \frac{M}{w(\mu - \delta)}\right) + \frac{1}{w^2 (\mu - \delta)^2}, \\ \frac{\|d_k\|^2}{\|g_k\|^4} &\leq \frac{1}{\|g_k\|^2} \left(\frac{M + w(\mu - \delta)}{w(\mu - \delta)}\right) + \frac{1}{w^2 (\mu - \delta)^2}, \end{aligned}$$

let $\tau = \frac{M + w(\mu - \delta)}{w(\mu - \delta)}$ and $q = w^2 (\mu - \delta)^2$, then, is obtained:

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leq \frac{\tau}{\|g_k\|^2} + \frac{1}{q},$$

suppose that the (3.14) does not hold. Then, there exists $\epsilon > 0$, such that (3.15) holds for all $k \geq 0$. Then the above inequality becomes:

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leq \frac{\tau}{\epsilon^2} + \frac{1}{q} = \frac{q\tau + \epsilon^2}{q\epsilon^2}. \tag{3.20}$$

Therefore, from (3.20) and (3.13), it follows that:

$$\begin{aligned} \sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} &\geq \sum_{k=0}^{\infty} \frac{q\epsilon^2}{q\tau + \epsilon^2} = \frac{q\epsilon^2}{q\tau + \epsilon^2} \sum_{k=0}^{\infty} 1, \text{ therefore,} \\ \sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} &\geq \infty. \end{aligned} \tag{3.21}$$

Then from (3.19), (3.21) contradicts (3.13) and this shows that (3.14) holds. The proof is complete.

Numerical Result

In this section, most of the problems from Andrei (2008); have been used to test and analyze the efficiency of $\beta_k^{MN^*}$ compared to FR, HS and CD. Stopping criteria is set to $\|g_k\| \leq \epsilon$ where $\epsilon = 10^{-6}$. As suggested by Hilstrom (1977); for each of test problem, four or five initial points

starting are used. All runs are performed on a PC ACER (Intel® Core™ i3-3217u CPU @ 1.8 GHZ, with 4.00 GB RAMS, Windows 10 Home Premium). Numerical results are compared based on the number of iterations and CPU time. Every problem mentioned in Table 1 is solved using Matlab12 subroutine programming. The strong Wolfe line search is being used as to give the inexact value of the step-size. The performance results are shows in Fig. 1 and 2 respectively, using a performance profile introduced by Dolan and More (2002).

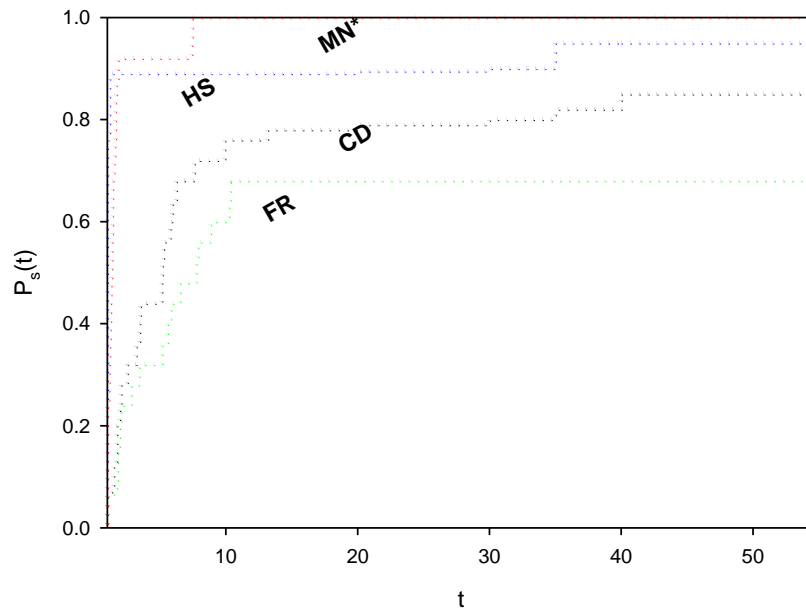


Fig. 1: Performance profile based on the number of iterations

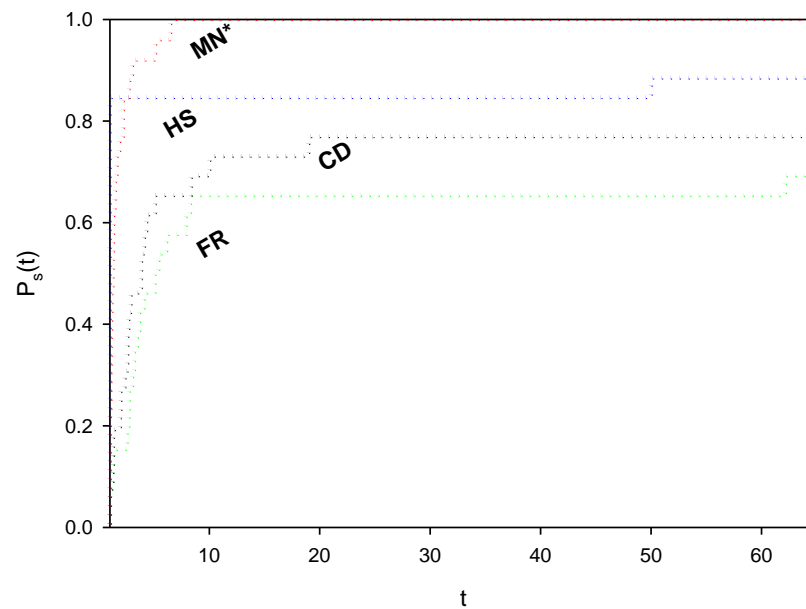


Fig. 2: Performance profile based on the CPU time

Table 1: A list of problem functions

No	Functions	N	Initial points
1	Zettl	2	(-10,-10),(-3,-3),(8,8),(20,20),(30,30)
2	Six-hump camel back	2	(-7,-7),(-2,-2),(3,3),(10,10)
3	Three-hump Camel back	2	(-3,-3),(2,2),(8,8),(13,13)
4	Trecanni	2	(-10,-10),(-5,-5),(7,7),(20,20),(30,30)
5	Hager	(2,4,10,100)	(-5,...,-5),(-1,...,-1),(3,...,3),(7,...,7)
6	Raydan1	(2,4,10,100)	(-3,...,-3),(2,...,2),(5,...,5),(10,...,10)
7	Shallow	(6,10,100,500,1000)	(8,...,8),(20,...,20),(40,...,40),(100,...,100)
8	Extended Tridiagonal2	(4,10)	(-6,...,-6),(-1,...,-1),(2,...,2),(5,...,5)
9	Extended Maratos	(2,4,10,100,500,1000)	(-9,...,-9),(-5,...,-5),(2,...,2),(6,...,6),(10,...,10)
10	Extended Tridiagonal1	(4,6,10,100,500,1000)	(2,...,2),(8,...,8),(14,...,14),(28,...,28)
11	Himmelblau	(4,10,100,500,1000)	(5,...,5),(11,...,11),(17,...,17),(31,...,31)
12	Generalized Quartic	(2,10,100,500,1000)	(-5,...,-5),(2,...,2),(7,...,7),(11,...,11)
13	Extended Rosenbrock	(4,10,100,500,1000)	(3,...,3),(6,...,6),(11,...,11),(23,...,23)
14	Extended Denschnb	(4,10,10,100)	(-5,...,-5),(-1,...,-1),(3,...,3),(10,...,10)
15	Arwhead	(2,4,100,500,1000)	(10,...,10),(50,...,50),(100,...,100),(200,...,200)
16	Freudenstein and Roth	(4,10,100,500,1000,5000)	(-6,...,-6),(-3,...,-3),(1,...,1),(7,...,7)
17	Flethcr	(4,10,100,500)	(-7,...,-7),(-2,...,-2),(5,...,5),(17,...,17)
18	Extended White &Holst	(4,10,100,500,1000)	(-8,...,-8),(-3,...,-3),(2,...,2),(10,...,10)
19	powell	(4,100,500,1000)	(-5,...,-5),(-1,...,-1),(3,...,3),(10,...,10)
20	Extended Penalty	(2,4,6,10,100,500,1000,5000)	(-3,...,-3),(-1,...,-1),(2,...,2),(8,...,8)
21	Levy	(2,10,100,500)	(-8,...,-8),(-2,...,-2),(3,...,3),(7,...,7)
22	Styblinski-Tang	(4,10,100,500)	(-5,...,-5),(-1,...,-1),(2,...,2),(4,...,4)
23	Leon	(2,6,10,100,1000)	(-1,...,-1),(0,...,0),(0.5,...,0.5),(1.2,...,1.2)
24	McCormick	(2,6,10,50)	(-1,...,-1),(1,...,1),(2,...,2),(4,...,4)
25	Chichinadze	(4,10,100,500)	(-10,...,-10),(-2,...,-2),(5,...,5),(20,...,20)
26	Schaffer	(2,10,100,500)	(-50,...,-50),(-3,...,-3),(10,...,10),(30,...,30)

The performance profile is used 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 exists, for each problem p and solver s , they defined $t_{p,s}$ = computing time (the number of iterations or CPU time or others) required to solve problems p by solver s . They compared the performance of problem p by solver s with the best performance by any solver on this problem 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 are chosen and $r_{p,s} = r_M$ if and only if solver s does not solve problem p . The performance solver s of given problems have to be the best, but we would like to obtain all evaluation performance of the solver, then 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 $P: R \rightarrow [0,1]$ for solver was a non-decreasing, piecewise and continuous from the right. The value $P(1)_s$ is the probability that the solver will win over the rest of the solvers. In all, a solver with high values of $P(t)_s$ or at the top right of the figures are preferable or represent the best solver.

Figures 1 and 2 show that MN* method is an efficient numerical performance, it can solve all the test problem and reach 100%. The numerical performance of the well-known conjugate gradient coefficients FR, HS and CD is divided into two groups; first group which is related to HS is given better in numerical performance than the second one which is related to FR and CD. The first group numerical performance seems to be less efficient than MN* method, it solves only 85% of the problems, but it is better than the second group numerical performance, which is solved only 80%. The numerical performance of MN* method seems to be robust and efficient when the first group and the second group are compared numerical performances. Hence, the formula MN* is robust and efficient.

Acknowledgement

The authors are grateful to the editor and the anonymous reviewers for their valuable comments and suggestions, which have substantially improved this paper.

Authors' contributions

All authors equally contributed in this work.

Ethics

The authors declare that this article is original, contains unpublished material, and no ethical issues involved.

References

- Al-Baali, M., 1985. Descent property and global convergence of the Fletcher-Reeves method with inexact line search. *IMA J. Numerical Anal.*, 5: 121-124. DOI: 10.1093/imanum/5.1.121
- Andrei, N., 2008. An unconstrained optimization test functions collection. *Adv. Modell. Optim.*, 10: 147-161.
- Andrei, N., 2011. A modified Polak–Ribière–Polyak conjugate gradient algorithm for unconstrained optimization. *J. Math. Programm. Operat. Res.*, 60: 1457-1471.
DOI: 10.1080/02331931003653187
- Dai, Y.H. and Y. Yuan, 1999. A nonlinear conjugate gradient method with a strong global convergent property. *SIAM J. Optim.*, 10: 177-182.
DOI: 10.1137/S1052623497318992
- Dai, Y.H., 2010. Nonlinear conjugate gradient methods. Chinese Academy of Sciences, Zhong Guan Cun Donglu, Beijing.
- Dolan, E. and J.J. Moré, 2002. Benchmarking optimization software with performance profile. *Math. Prog.*, 91: 201-213.
DOI: 10.1007/s101070100263
- Fletcher, R. and C. Reeves, 1964. Function minimization by conjugate gradients. *Comput. J.*, 7: 149-154.
DOI: 10.1093/comjnl/7.2.149
- Fletcher, R., 1997. *Practical Method of Optimization*. 2nd Edn., Wiley, New York.
- Goldstein, A.A., 1965. On steepest descent. *SIAM J. Control*, 3: 147-151. DOI: 10.1137/0303013
- Hestenes, M.R. and E. Stiefel, 1952. Method of conjugate gradient for solving linear equations. *J. Res. Nat. Bur. Stand.*, 49: 409-436.
DOI: 10.6028/jres.049.044
- Hillstom, K.E., 1977. A simulation test approach to the evaluation of nonlinear optimization algorithms. *ACM. Trans. Math. Softw.*, 3: 305-315.
DOI: 10.1145/355759.355760
- Liu, Y. and C. Storey, 1992. Efficient generalized conjugate gradient algorithms, part 1: theory. *J. Optim. Theory Applic.*, 69: 129-137.
DOI: 10.1007/BF00940464
- Perry, J.M., 1977. A class of conjugate gradient algorithms with a two-step variable-metric memory. Center for Mathematical Studies in Economics and Management Sciences, Northwestern University.
- Polak, E. and G. Ribiere, 1969. Note sur la convergence de methodes de directions conjuguees. *R.I.R.O.*, 3: 35-43. DOI: 10.1051/m2an/196903R100351
- Powell, M.J.D., 1977. Restart procedure for the conjugate gradient method. *Math. Prog.*, 2: 241-254.
DOI: 10.1007/BF01593790
- Powell, M.J.D., 1984. Non-Convex Minimization Calculation and the Conjugate Gradient Method. In: *Numerical Analysis*, Griffiths, D.F. (Ed.), Springer-Verlag, Berlin, pp: 122-141.
- Powell, M.J.D., 1986. Convergence properties of algorithms for nonlinear optimizations. *SIAM Rev.*, 28: 487-500. DOI: 10.1137/1028154
- Qi, H.D., J.Y. Han and G.H. Liu, 1999. A modified Hestenes-Stiefel conjugate gradient algorithm. *Chinese Annals Math., Series A*, 17: 177-184.
- Rivaie, M., M. Mamat, L. Wah June and M. Ismail, 2012. A new class of nonlinear conjugate gradient coefficients with global convergence properties. *Applied Math. Comput.*, 218: 11323-1133.
DOI: 10.1016/j.amc.2012.05.030
- Wolfe, P., 1969. Convergence conditions for ascent method. *SIAM Rev.*, 11: 226-235.
DOI: 10.1137/1011036
- Zhang, L., 2009. An improved Wei-Yao-Liu nonlinear conjugate gradient method for optimization computation. *Applied Math. Comput.*, 215: 2269-2274. DOI: 10.1016/j.amc.2009.08.016
- Zoutendijk, G., 1970. Nonlinear Programming Computational Methods. In: *Integer and Nonlinear Programming*, Abadie, J. (Ed.), North Holland, Amsterdam, pp: 37-86.