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

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

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:

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

(1.1)

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

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

(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 direction with:

$$d_k = \begin{cases} -g_k & \text{if } k = 0 \\ -g_k + \beta_k d_{k-1} & \text{if } k \geq 1. \end{cases}$$

(1.3)

where, $\beta_k$ 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 $\alpha_k$ and defined as follows:

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

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

(1.4)

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

There are at least six well-known formulas for $\beta_k$, which are given below:

$$\beta_{k}^{BS} = \frac{\langle g_k, g_{k-1} \rangle}{\langle g_{k-1}, g_{k-1} \rangle} \quad \text{(Hestenes and Stiefel, 1952),}$$

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

$$\beta_{k}^{PR} = \frac{\langle g_k, g_{k-1} \rangle}{\langle g_{k-1}, g_{k-1} \rangle} \quad \text{(Polak and Ribiere, 1969),}$$

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

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

$$\beta_{k}^{DY} = \frac{\langle g_k, g_{k-1} \rangle}{\langle g_{k-1}, g_{k-1} \rangle} \quad \text{(Dai and Yuan, 1999).}$$

These parameters $\beta_k$ (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 $\beta_k$ leads to very different performance and convergence of the
corresponding algorithms (Andrei, 2011). The behavior of convergence of the $\beta_k$’s formulas with some line search conditions has been established by many authors for many years ago (Rivai 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 $\beta_k$’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 $\sigma>0.5$. 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 $\beta_{k}^{HS}$ become small so that $\beta_{k}^{HS}$ might be unbounded (Dai, 2010). Qi et al. (1999) established the global convergence of a modified HS method, where $\beta_k$ takes the form:

$$\beta_{k}^{Q} = \max \left\{ 0, \min \left( \beta_{k}^{MS}, \frac{1}{\|g_k\|} \right) \right\}.$$  

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

$$d_k = -P_k g_k,$$

where:

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

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:

$$\beta_{k}^{MS} = \frac{\|g_k\|^2}{\|d_{k-1}\|^2 - \|g_k\|^2 \|g_{k-1}\|},$$

(2.1)

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

The formula (2.1) has the following properties:

$$\beta_{k}^{MS} \begin{cases} \frac{\|g_k\|^2}{\|d_{k-1}\|^2 - \|g_k\|^2 \|g_{k-1}\|}, & \text{if } g_k d_{k-1} \leq 0, \\ \frac{\|g_k\|^2}{\|wd_{k-1}\|}, & \text{if } g_k d_{k-1} > 0, \end{cases}$$

(2.2)

where $w = 1 - \frac{\|g_{k-1}\|}{\|d_{k-1}\|} \beta_{k-1}^{FR}$.

Lemma 2.1

Consider any method (1.2) and (1.3) be generated by Algorithm 2.1, let $0 < \sigma < 1$, $\mu > 0$ and $\beta_k = \beta_k^{MS}$ 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:

\[ \| \beta_i \| \| d_{i-1} \| = \| d_i + g_i \| \leq \| g_i \| + \| d_{i-1} \|, \]

and \[ | \beta_i | > 0, \]

dividing both sides of above inequality by \[ | \beta_i | \:

\[ \| d_{i-1} \| \leq \frac{\| g_i \|}{| \beta_i |} + \frac{\| d_{i-1} \|}{| \beta_i |}, \]

(2.3)
is obtained, also \[ \frac{\| g_i \|}{| \beta_i |} > 0, \frac{\| d_{i-1} \|}{| \beta_i |} > 0 \] and let \[ \delta = \frac{\| g_i \|}{| \beta_i |}, \mu = \frac{\| d_{i-1} \|}{| \beta_i |}, \]

then from (2.3):

\[ \| d_{i-1} \| \leq \delta + \mu. \]

(2.4)

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

\[ \| d_i \| = |g_i + \beta_i d_{i-1}| \leq |g_i| + |\beta_i| \| d_{i-1} \|, \]

\[ \| d_i \| - |g_i| \leq |\beta_i| \| d_{i-1} \|, \]

\[ \frac{\| d_i \|}{|\beta_i|} - |g_i| \leq |d_{i-1}|, \]

\[ \frac{\| d_i \|}{|\beta_i|} \leq |d_{i-1}|. \]

(2.5)

Then, \[ |g_i| > 0, \frac{\| d_i \|}{|\beta_i|} > 0 \] and let \[ \delta = \frac{\| g_i \|}{| \beta_i |}, \mu = \frac{\| d_{i-1} \|}{| \beta_i |}, \]

substitution in the above inequality, is obtained:

\[ \mu - \delta \leq \| d_{i-1} \|. \]

The proof is complete.

**Lemma 2.2**

Consider any method (1.2) and (1.3) be generated by Algorithm 2.1, let and \[ \beta_i = \beta_i^{\text{inv}} \] is given as (2.1), then:

\[ \frac{\| d_i \|}{\| g_i \|} \leq 1 + \frac{M}{\mu - \delta}, \quad \text{if } g_i d_{i-1} \leq 0, \text{ and} \]

\[ \frac{\| d_i \|}{\| g_i \|} \leq 1 + \frac{M}{w(\mu - \delta)}, \quad \text{if } g_i d_{i-1} > 0 \]

(2.6)

**Proof**

From (2.2), if \[ g_k d_{k-1} \leq 0, \] (1.3), \[ \beta_k = \beta_k^{\text{inv}} \], (2.5) and triangle inequality, we have:

\[ d_i = -g_i + \beta_i d_{i-1}, \]

\[ 0 \leq \| d_i \| \leq |g_i| + |\beta_i| \| d_{i-1} \|, \]

\[ 0 \leq \| d_i \| \leq |g_i| + \frac{|g_i|^2}{\| d_{i-1} \|^2} \| d_{i-1} \| = |g_i| + \frac{|g_i|^2}{\mu - \delta}, \]

(2.7)

dividing above inequality by \[ \| g_i \| \] and from (3.2), we obtain:

\[ 0 \leq \frac{\| d_i \|}{\| g_i \|} \leq 1 + \frac{M}{\mu - \delta}. \]

From (2.2), if \[ g_k d_{k-1} > 0, \] (1.3), \[ \beta_k = \beta_k^{\text{inv}} \], (2.5) and triangle inequality, we have:

\[ 0 \leq \| d_i \| \leq |g_i| + |\beta_i| \| d_{i-1} \|, \]

\[ 0 \leq \| d_i \| \leq |g_i| + \frac{|g_i|^2}{w \| d_{i-1} \|^2} \| d_{i-1} \| = |g_i| + \frac{|g_i|^2}{w(\mu - \delta)}. \]

(2.8)

dividing above inequality by \[ \| g_i \| \] and from (3.2), we obtain:

\[ 0 \leq \frac{\| d_i \|}{\| g_i \|} \leq 1 + \frac{M}{w(\mu - \delta)}. \]

The proof is complete.

The Algorithm 2.1 of the \[ \beta_i^{\text{inv}} \] method is presented.

**Algorithm 2.1**

**Step 1:** Initialization. Given \( x_0 \in \mathbb{R}^n \), \( d_0 = -g_0 \), \( k = 0 \), if \( g_0 = 0 \) then stop.

**Step 2:** Compute \( \beta_k \) 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 \( \beta_i^{\text{inv}} \).

**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 \( \mathbb{R}^n \) and continuous, differentiable in a neighborhood \( N \) of the level set \( \ell(x_0) = \{ x \in \mathbb{R}^n \mid f(x) \leq f(x_0) \} \) 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:

\[ \| g(x) - g(y) \| \leq L \| x - y \|, \text{for any } x, y \in N. \]

(3.1)

By using the Assumption A and there exists \( M > 0 \) such that:
|g(x)| \leq M \quad \forall x \in \ell.  
\eqref{eq: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 \( \alpha_k \) be determined by the strong Wolfe line search \( (1.4) \), where \( \beta = \beta_S^{\text{adv}} \) is given as \( (2.1) \), then for all \( k \geq 0 \):

\[
g_i^T d_i \leq -\| g_i \|^2. 
\eqref{eq:3.3}

**Proof**

If \( k = 0 \), then \( g_0^T d_0 = -\| g_0 \|^2 \), 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 \( g_i^T d_{i-1} \leq 0 \), from \( (1.3) \) multiply by \( g_i \) and second inequality of \( (1.4) \), we get:

\[
g_i^T d_i = -\| g_i \|^2 + \beta_S^{\text{adv}} g_i^T d_{i-1} \leq -\| g_i \|^2 + \beta_S^{\text{adv}} \| g_{i-1} \| d_{i-1}.
\eqref{eq:3.4}

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

\[
g_i^T d_i \leq -\| g_i \|^2 + \| g_i \|^2 + \| g_i \|^2 - \| g_{i-1} \| d_{i-1} \| d_{i-1} \| d_{i-1} \|
\eqref{eq:3.5}

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

\[
g_i^T d_i \leq -\| g_i \|^2 \left( 1 - \frac{\sigma M}{\mu + \delta} \right).
\eqref{eq:3.6}

If \( g_i^T d_{i-1} > 0 \), from \( (1.3) \) multiply by \( g_i \) and from \( (1.4) \), is obtained:

\[
g_i^T d_i = -\| g_i \|^2 + \beta_S^{\text{adv}} g_i^T d_{i-1} \leq -\| g_i \|^2 + \| g_i \|^2 - \| g_{i-1} \| d_{i-1} \| d_{i-1} \| d_{i-1} \|
\eqref{eq:3.7}

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

\[
g_i^T d_i \leq -\| g_i \|^2 \left( 1 - \frac{\sigma M}{w(\mu + \delta)} \right).
\eqref{eq: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 \( d_i \) is a descent direction and \( \alpha_k \) satisfies the strong Wolfe condition line search \( (1.4) \). Then we have:

\[
\sum_{i=0}^{\infty} \| d_i \|^2 < \infty.
\eqref{eq:3.9}

**Proof**

We get from the second inequality of \( (1.4) \):

\[
\sigma g_i^T d_{i-1} \leq g_i^T d_{i-1} \leq -\sigma g_i^T d_{i-1}.
\eqref{eq:3.10}

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

\[
(\sigma - 1) g_i^T d_{i-1} \leq (g_i - g_{i-1})^T d_{i-1} \leq L \alpha_{i-1} \| d_{i-1} \|^2,
\eqref{eq:3.11}

then:

\[
\alpha_{i-1} \geq \frac{(\sigma - 1) g_i^T d_{i-1}}{L \| d_{i-1} \|^2} = \frac{1 - \sigma}{L} \left| g_i^T d_{i-1} \right|
\eqref{eq:3.12}

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

\[
-\sum_{i=0}^{\infty} \alpha_{i-1} g_i^T d_{i-1} \leq \infty.
\eqref{eq:3.13}

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

Lemma 3.2 is equivalent to the following inequality:

\[
\sum_{i=0}^{\infty} \| g_i \|^2 < \infty.
\eqref{eq:3.14}

Now, the global convergence Theorem of the formula \( \beta_S^{\text{adv}} \) is established.

**Theorem 3.2**

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

\[
\liminf_{k=0} \| g_i \| = 0.
\eqref{eq:3.15}
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:

\[
|g_k| \geq \epsilon \quad (3.15)
\]

Rewriting (1.3) as:

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

and squaring both sides of the equation, we obtain:

\[
\|d_k\|^2 = \|g_k\|^2 - 2g_k^T d_k + (\beta_k^T \beta_k)^{\frac{1}{2}} \|d_{k-1}\|^2. \quad (3.16)
\]

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

\[
\|d_k\|^2 \leq \|g_k\|^2 - 2g_k^T d_k + \left( \frac{\|g_k\|^2}{\|d_{k-1}\|^2} \right) \|d_{k-1}\|^2.
\]

Dividing both sides of (3.17) by \( \|g_k\|^2 \) and from (2.5), Cauchy-inequality, is obtained:

\[
\frac{\|d_k\|^2}{\|g_k\|^2} \leq -1 + \frac{2g_k^T d_k}{\|g_k\|^2} + \frac{1}{\|d_{k-1}\|^2} \leq -1 + \frac{2\|d_k\|^2}{\|g_k\|^2} + \frac{1}{\|d_{k-1}\|^2} - \frac{\|g_k\|^2}{\|d_{k-1}\|^2} - \frac{2\|d_k\|^2}{\|g_k\|^2} + \frac{1}{\|d_{k-1}\|^2} - \frac{\|g_k\|^2}{\|d_{k-1}\|^2}.
\]

and from (2.7), is obtained:

\[
\frac{\|d_k\|^2}{\|g_k\|^2} \leq -1 + \frac{1}{\|g_k\|^2} \left( \frac{M}{\mu - \delta} \right) + \frac{1}{\|d_{k-1}\|^2},
\]

\[
\frac{\|d_k\|^2}{\|g_k\|^2} \leq 1 + \frac{1}{\|g_k\|^2} \left( \frac{M + \mu - \delta}{\mu - \delta} \right) + \frac{1}{\|d_{k-1}\|^2},
\]

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

\[
\frac{\|d_k\|^2}{\|g_k\|^2} \leq \frac{1}{\|g_k\|^2} \left( \frac{M + \mu - \delta}{\mu - \delta} \right) + \frac{1}{\|d_{k-1}\|^2},
\]

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 becomes:

\[
\frac{\|d_k\|^2}{\|g_k\|^2} \geq \frac{\epsilon^2}{\epsilon + \gamma^2}. \quad (3.18)
\]

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

\[
\sum_{i=0}^{\infty} \frac{\|d_i\|^2}{\|g_i\|^2} \geq \sum_{i=0}^{\infty} \frac{\epsilon^2}{\epsilon + \gamma^2} = \frac{\epsilon^2}{\epsilon + \gamma^2} \sum_{i=0}^{\infty} 1, \quad \therefore, \sum_{i=0}^{\infty} \frac{\|d_i\|^2}{\|g_i\|^2} \geq \infty.
\]

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

\[
\|d_k\|^2 \leq \|g_k\|^2 - 2g_k^T d_k + \left( \frac{\|g_k\|^2}{\|d_{k-1}\|^2} \right) \|d_{k-1}\|^2.
\]

dividing both sides of the above inequality by \( \|g_k\|^2 \) and from (2.5), Cauchy-inequality, is obtained:

\[
\|d_k\|^2 \leq \frac{1}{\|g_k\|^2} \left[ \frac{\|g_k\|^2}{\|d_{k-1}\|^2} \right] \|d_{k-1}\|^2 - \frac{2\|d_k\|^2}{\|g_k\|^2} + \frac{1}{\|d_{k-1}\|^2} - \frac{\|g_k\|^2}{\|d_{k-1}\|^2} - \frac{2\|d_k\|^2}{\|g_k\|^2} + \frac{1}{\|d_{k-1}\|^2} - \frac{\|g_k\|^2}{\|d_{k-1}\|^2}.
\]

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

\[
\|d_k\|^2 \leq \frac{1}{\|g_k\|^2} \|g_k\|^2 \left( 1 + \frac{M}{w(\mu - \delta)} \right) + \frac{1}{w^2(\mu - \delta)}.
\]

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

\[
\|d_k\|^2 \leq \frac{\tau}{\|g_k\|^2} \left( 1 + \frac{1}{q} \right).
\]

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:

\[
\|d_k\|^2 \leq \frac{\epsilon^2}{\epsilon + \gamma^2} \|g_k\|^2 + \frac{\epsilon^2}{\epsilon + \gamma^2} \|g_k\|^2.
\]

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

\[
\sum_{i=0}^{\infty} \|d_i\|^2 \geq \sum_{i=0}^{\infty} \frac{\epsilon^2}{\epsilon + \gamma^2} = \frac{\epsilon^2}{\epsilon + \gamma^2} \sum_{i=0}^{\infty} 1, \quad \therefore, \sum_{i=0}^{\infty} \|d_i\|^2 \geq \infty.
\]
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 $\beta^{MN}$ 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 shown 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](image1.png)

![Fig. 2: Performance profile based on the CPU time](image2.png)

59
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 \( r_{p,s} \) = 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:

\[
\frac{r_{p,s}}{\min\{r_{p,s}: s \in S\}}.
\]

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:

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

The \( P(t) \) 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 \( P_t \) 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 \( P(1) \), is the probability that the solver will win over the rest of the solvers. In all, a solver with high values of \( P(t) \), 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


