Journal of Mathematical Research & Exposition
Mar., 2010, Vol. 30, No. 2, pp. 297–308
DOI:10.3770/j.issn:1000-341X.2010.02.013
Http://jmre.dlut.edu.cn

A Modified Hestenes-Stiefel Conjugate Gradient Method and Its Convergence

Zeng Xin WEI*, **Hai Dong HUANG**, **Yan Rong TAO**
College of Mathematics and Information Science, Guangxi University, Guangxi 530004, P. R. China

Abstract It is well-known that the direction generated by Hestenes-Stiefel (HS) conjugate gradient method may not be a descent direction for the objective function. In this paper, we take a little modification to the HS method, then the generated direction always satisfies the sufficient descent condition. An advantage of the modified Hestenes-Stiefel (MHS) method is that the scalar β_k^{HS*} keeps nonnegative under the weak Wolfe-Powell line search. The global convergence result of the MHS method is established under some mild conditions. Preliminary numerical results show that the MHS method is a little more efficient than PRP and HS methods.

Keywords conjugate gradient method; sufficient descent condition; line search; global convergence.

Document code A
MR(2000) Subject Classification 90C30
Chinese Library Classification O221.1

1. Introduction

Consider the unconstrained optimization problem $\min\{f(x)|x \in \Re^n\}$, where $f : \Re^n \to \Re$ is a continuously differentiable function whose gradient is denoted by g. The conjugate gradient method can be described by the iterative scheme:

$$x_{k+1} = x_k + t_k d_k, \tag{1.1}$$

where the positive step-size t_k is obtained by some line search, and the direction d_k is generated by the rule:

$$d_k = \begin{cases} -g_k, & \text{if } \ k = 1, \\ -g_k + \beta_k d_{k-1}, & \text{if } \ k \geq 2, \end{cases} \tag{1.2}$$

where g_k denotes $g(x_k)$, β_k is a scalar. Some well-known formulas for β_k are given as follows:

$$\beta_k^{\rm FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}{}^{[7]}; \ \beta_k^{\rm HS} = \frac{g_k^{\rm T}(g_k - g_{k-1})}{(g_k - g_{k-1})^{\rm T} d_{k-1}}{}^{[11]}; \ \beta_k^{\rm PRP} = \frac{g_k^{\rm T}(g_k - g_{k-1})}{\|g_{k-1}\|^2}{}^{[18,19]},$$

where $\|\cdot\|$ denotes the ℓ_2-norm. Their corresponding conjugate gradient methods will be abbreviated as FR, HS and PRP methods. Although the above mentioned conjugated gradient methods

Received December 22, 2007; Accepted May 21, 2008
Supported by the National Natural Science Foundation of China (Grant No. 10761001).
* Corresponding author
E-mail address: zxwei@gxu.edu.cn (Z. X. WEI); haidongh@126.com (H. D. HUANG); nameyanrongtao@163.com (Y. R. TAO)

are identical when f is a strongly convex quadratic function and the line search is exact, the behavior of these methods differs markedly when f is a general function and the line search is inexact. Usually, two major inexact line searches are the weak Wolfe-Powell (WWP) line search, i.e.,

$$f(x_k + t_k d_k) - f(x_k) \leq \delta t_k g_k^{\mathrm{T}} d_k, \tag{1.3}$$

and

$$g(x_k + t_k d_k)^{\mathrm{T}} d_k \geq \sigma g_k^{\mathrm{T}} d_k, \tag{1.4}$$

and the strong Wolfe-Powell (SWP) line search, namely, (1.3) and

$$|g(x_k + t_k d_k)^{\mathrm{T}} d_k| \leq \sigma |g_k^{\mathrm{T}} d_k|, \tag{1.5}$$

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

For the FR method, if a bad direction and tinny step from x_{k-1} to x_k are generated, the next direction and the next step are also likely to be poor unless a restart along the gradient direction is performed. In spite of such a defect, Zoutendijk [27] proved that the FR method with exact line search is globally convergent. Al-Baali [1] extended this result to inexact line searches. On the other hand, the PRP and HS methods perform similarly in terms of theoretical property. Both two methods perform better than the FR method in numerical computation, because these two methods essentially perform a restart if a bad direction occurs. Nevertheless, Powell [17] showed that PRP and HS methods can cycle infinitely without approaching a solution, which implies that they are not globally convergent. Gibert and Nocedal [9] firstly proved that the PRP conjugate gradient method converges globally when the sufficiently decreasing condition, namely,

$$g_k^{\mathrm{T}} d_k < -c\|g_k\|^2, \tag{1.6}$$

where c is a positive constant, and the so-called Property $(*)$ [9] are satisfied. Dai and Yuan [4–6] gave further study of the convergence of the PRP method when β_k is defined by $\beta_k = \max\{0, \beta_k^{PRP}\}$. The global convergence of related PRP methods was also studied in Wei [23–25] and Li [14]. In order to match the requirement for the convergence theory, some modified PRP methods [10, 13, 16] add some strong assumptions or use complicated line searches. Recently, Hager and Zhang [12] developed a new conjugate gradient method which always generates descent directions and reduces to the HS method if the exact line search is adopted.

Birgin and Martinez [3] proposed a spectral conjugate gradient method by combining conjugate gradient method and spectral gradient method [20] in the following way:

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

where ϑ_k is a parameter and

$$\beta_k = \frac{(\vartheta_k y_{k-1} - s_{k-1})^{\mathrm{T}} g_k}{d_{k-1}^{\mathrm{T}} y_{k-1}}.$$

The reported numerical results show that the above method performs very well if ϑ_k is defined by

$$\vartheta_k = \frac{s_{k-1}^{\mathrm{T}} s_{k-1}}{s_{k-1}^{\mathrm{T}} y_{k-1}},$$

where $s_{k-1} = x_k - x_{k-1}$ and $y_{k-1} = g_k - g_{k-1}$. Unfortunately, the spectral conjugate gradient method [3] cannot guarantee to generate descent directions. In other words, the generated direction d_k does not satisfy the following condition:

$$g_k^{\mathrm{T}} d_k < 0. \tag{1.7}$$

Therefore, based on the quasi-Newton BFGS update formula, Andrei [2] developed scaled conjugate gradient algorithm and proved its global convergence under Wolfe-Powell line search.

In this paper, we take a little modification to the HS method, then the directions generated by the MHS method with the SWP line search always satisfy sufficient descent condition. Under some mild conditions, we prove that the modified method is globally convergent.

The paper is organized as follows. In the next section, we present the modified HS method. In Section 3, we establish global convergence results of the MHS method with the SWP line search. Preliminary numerical results are reported in Section 4. Finally, Section 5 contains some conclusions.

2. Algorithm and preliminaries

It is well-known that PRP and HS methods perform similarly and are regarded as the most efficient conjugate gradient methods. Unfortunately, these two methods are not globally convergent even for the exact line search; see the counterexample of Powell [17]. These facts motivate us to design a new conjugate gradient method that performs similarly to PRP and HS methods and has global convergence. Motivated by Wei [25] and the convergence analysis for the PRP method by Gilbert and Nocedal [9], we propose a MHS method, in which the scalar β_k is defined by

$$\beta_k^{\mathrm{HS}*} = \frac{g_k^{\mathrm{T}}(g_k - \frac{g_k^{\mathrm{T}} g_{k-1}}{\|g_{k-1}\|^2} g_{k-1})}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})}. \tag{2.1}$$

An important feature of this new choice of β_k is that, when the WWP line search is adopted, the value of β_k is greater than zero. From (1.4) and (1.7), we have

$$d_{k-1}^{\mathrm{T}}(g_k - g_{k-1}) \geq (\sigma - 1) g_{k-1}^{\mathrm{T}} d_{k-1} \geq 0, \tag{2.2}$$

which, along with (2.1), gives

$$\begin{aligned} \beta_k^{\mathrm{HS}*} &= \frac{g_k^{\mathrm{T}} g_k - \frac{g_k^{\mathrm{T}} g_{k-1}}{\|g_{k-1}\|^2} g_k^{\mathrm{T}} g_{k-1}}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})} = \frac{\frac{g_k^{\mathrm{T}} g_k \|g_{k-1}\|^2 - g_k^{\mathrm{T}} g_{k-1} g_k^{\mathrm{T}} g_{k-1}}{\|g_{k-1}\|^2}}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})} \\ &= \frac{\frac{\|g_k\|^2 \|g_{k-1}\|^2 - \|g_k\|^2 \|g_{k-1}\|^2 \cos^2 \alpha_k}{\|g_{k-1}\|^2}}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})} = \frac{\frac{\|g_k\|^2 \|g_{k-1}\|^2 (1 - \cos^2 \alpha_k)}{\|g_{k-1}\|^2}}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})} \geq 0, \end{aligned}$$

where α_k is the angle between g_k and g_{k-1}.

Our modified HS conjugate gradient method algorithm is given below.

Algorithm 2.1 (Modified HS method: MHS)

Step 0. Given $x_1 \in \Re^n$, set $d_1 = -g_1, k = 1$. If $g_1 = 0$, then stop.

Step 1. Find a $t_k > 0$ satisfying strong Wolfe-Powell line search (SWP).

Step 2. Let $x_{k+1} = x_k + t_k d_k$ and $g_{k+1} = g(x_{k+1})$. If $g_{k+1} = 0$, then stop.

Step 3. Compute β_k by the formula (2.1) and generate d_{k+1} by (1.2).

Step 4. Set k:=k+1, go to Step 1.

For our later analysis, we make the following basic assumptions on the objective function.

Assumption A The level set $\Omega = \{x \in R^n \mid f(x) \leq f(x_1)\}$ is bounded, where x_1 is given by Algorithm 2.1.

Assumption B In some neighborhood N of Ω, f is continuously differentiable, and its gradient is Lipschitz continuously differentiable, that is, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \tag{2.3}$$

for all $x, y \in \Omega$.

Lemma 2.1 *Suppose that Assumptions A and B hold. Consider any method of the form (1.1)–(1.2) with d_k satisfying (1.7) for all k, and t_k satisfying (1.3) and (1.4). Then*

$$\sum_{k\geq 1} \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2} < +\infty. \tag{2.4}$$

This result was essentially proved by Zoutendijk [27] and Wolfe [21, 22]. We shall refer to (2.4) as the Zoutendijk condition.

Lemma 2.2 *Suppose that Assumptions A and B hold. Consider any method of the form (1.1)–(1.2) with d_k satisfying (1.7) for all k, and t_k satisfying (1.3) and (1.5), then either*

$$\liminf_{k\to\infty} \|g_k\| = 0, \tag{2.5}$$

or

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

Corollary 2.1 *Suppose that Assumptions A and B hold. Consider any method of the form (1.1)–(1.2) with d_k satisfying (1.7) for all k, and t_k satisfying (1.3) and (1.5). If*

$$\sum_{k}^{\infty} \frac{\|g_k\|^t}{\|d_k\|^2} = +\infty$$

for any $t \in [0, 4]$, then the method converges in the sense that (2.5) is true.

3. Global convergence results

In this section we discuss convergence properties of the proposed method in conjunction with strong Wolfe-Powell line search. The following result gives conditions on the line search that guarantee that all search directions satisfy the sufficient descent condition (1.6).

Theorem 3.1 *Suppose that Assumptions A and B hold. Consider the method of the form (1.1)–(1.2) with β_k satisfying (2.1), and the step-size t_k satisfying (1.3) and (1.5) with $0 < \sigma < 1/2$.*

Then the method generates descent directions d_k that satisfy the following inequalities:

$$\frac{-2\sigma-1}{1+\sigma} \le \frac{g_k^{\mathrm{T}} d_k}{\|g_k\|^2} \le \frac{2\sigma-1}{1-\sigma}, \quad k=1,2,\dots. \tag{3.1}$$

Proof The proof is by induction. The result clearly holds for $k=1$, since the middle term equals -1 and $0<\sigma<1/2$. Assume that (3.1) holds for some $k \ge 1$. This implies that (1.7) holds, since $\frac{2\sigma-1}{1-\sigma}<0$, by the condition $0<\sigma<1/2$. Combining with (1.5), it follows that

$$\sigma g_k^{\mathrm{T}} d_k \le g(x_{k+1})^{\mathrm{T}} d_k \le -\sigma g_k^{\mathrm{T}} d_k, \tag{3.2}$$

which implies that

$$(g(x_{k+1})-g_k)^{\mathrm{T}} d_k \ge (\sigma-1) g_k^{\mathrm{T}} d_k \ge 0. \tag{3.3}$$

From (1.2) and (2.1), we have

$$\begin{aligned} g_{k+1}^{\mathrm{T}} d_{k+1} &= -\|g_{k+1}\|^2 + \beta_{k+1}^{\mathrm{HS}*} g_{k+1}^{\mathrm{T}} d_k \\ &= -\|g_{k+1}\|^2 + \frac{g_{k+1}^{\mathrm{T}}(g_{k+1} - \frac{g_{k+1}^{\mathrm{T}} g_k}{\|g_k\|^2} g_k)}{d_k^{\mathrm{T}}(g_{k+1}-g_k)} g_{k+1}^{\mathrm{T}} d_k \\ &= -\|g_{k+1}\|^2 + \frac{g_{k+1}^{\mathrm{T}} d_k \|g_{k+1}\|^2}{d_k^{\mathrm{T}}(g_{k+1}-g_k)} (1 - \frac{(g_{k+1}^{\mathrm{T}} g_k)^2}{\|g_{k+1}\|^2 \|g_k\|^2}) \\ &= -\|g_{k+1}\|^2 + \frac{g_{k+1}^{\mathrm{T}} d_k \|g_{k+1}\|^2}{d_k^{\mathrm{T}}(g_{k+1}-g_k)} (1-\cos^2\alpha_{k+1}), \end{aligned} \tag{3.4}$$

where α_{k+1} is the angle between g_{k+1} and g_k. Dividing both sides of the equality (3.4) by $\|g_{k+1}\|^2$ and combining (3.2) with (3.3), we obtain

$$\begin{aligned} \frac{g_{k+1}^{\mathrm{T}} d_{k+1}}{\|g_{k+1}\|^2} &= -1 + \frac{g_{k+1}^{\mathrm{T}} d_k}{d_k^{\mathrm{T}}(g_{k+1}-g_k)} (1-\cos^2\alpha_{k+1}) \\ &\le -1 - \sigma \frac{g_k^{\mathrm{T}} d_k}{(\sigma-1) g_k^{\mathrm{T}} d_k} (1-\cos^2\alpha_{k+1}) \\ &= -1 + \frac{\sigma(1-\cos^2\alpha_{k+1})}{1-\sigma} \le -1 + \frac{\sigma}{1-\sigma} = \frac{2\sigma-1}{1-\sigma}, \end{aligned}$$

and

$$\begin{aligned} \frac{g_{k+1}^{\mathrm{T}} d_{k+1}}{\|g_{k+1}\|^2} &= -1 + \frac{g_{k+1}^{\mathrm{T}} d_k}{d_k^{\mathrm{T}}(g_{k+1}-g_k)} (1-\cos^2\alpha_{k+1}) \\ &\ge -1 - \sigma \frac{g_k^{\mathrm{T}} d_k}{(\sigma+1) g_k^{\mathrm{T}} d_k} (1-\cos^2\alpha_{k+1}) \\ &= -1 - \frac{\sigma(1-\cos^2\alpha_{k+1})}{1+\sigma} \ge -1 - \frac{\sigma}{1+\sigma} = \frac{-2\sigma-1}{1+\sigma}. \end{aligned}$$

We conclude that (3.1) holds for $k+1$. The proof is completed. □

Lemma 3.1 *Suppose that Assumptions A and B hold. Consider the method of the form (1.1)–(1.2) with β_k satisfying (2.1) and t_k satisfying (1.3) and (1.5). Then (2.4) holds.*

Proof From the line search condition (1.5), we have

$$(g_{k+1}-g_k)^{\mathrm{T}} d_k \ge (\sigma-1) g_k^{\mathrm{T}} d_k. \tag{3.5}$$

By the Lipschitz condition (2.3), we obtain $(g_{k+1}-g_k)^{\mathrm{T}}d_k \le Lt_k\|d_k\|^2$, which, along with (3.5), gives

$$t_k \ge \frac{\sigma-1}{L}\frac{g_k^{\mathrm{T}}d_k}{\|d_k\|^2}. \tag{3.6}$$

By (1.3) and (3.6), we get

$$f(x_k)-f(x_{k+1}) \ge -\delta t_k g_k^{\mathrm{T}}d_k \ge \delta\frac{1-\sigma}{L}\frac{(g_k^{\mathrm{T}}d_k)^2}{\|d_k\|^2}. \tag{3.7}$$

Since the level set Ω is bounded and $\{f(x_k)\}$ is a decreasing sequence, by summing both sides of the inequality (3.7), we have

$$\sum_{k\ge1}\delta\frac{1-\sigma}{L}\frac{(g_k^{\mathrm{T}}d_k)^2}{\|d_k\|^2} \le \sum_{k\ge1}(f(x_k)-f(x_{k+1})) < +\infty.$$

The proof is completed. □

For the remaining section, we assume that convergence does not occur in a finite number of steps, i.e., $g_k \ne 0$ for all k.

Lemma 3.2 *Suppose that Assumptions A and B hold. Consider any method of the form (1.1)–(1.2) with the following three properties: (i) $\beta_k \ge 0$; (ii) t_k satisfies (1.3) and (1.4); (iii) d_k satisfies the sufficient decrease condition (1.6). If there exists a constant $\gamma > 0$ such that*

$$\|g_k\| \ge \gamma, \tag{3.8}$$

then $d_k \ne 0$ and

$$\sum_{k\ge2}\|u_k-u_{k-1}\|^2 < \infty, \tag{3.9}$$

where $u_k = \frac{d_k}{\|d_k\|}$.

The proof of the Lemma can be found in [4]. Of course, condition (3.9) does not imply the convergence of the sequence u_k, but shows that the search directions u_k change slowly.

When a small step-size is generated from the PRP method, the next search direction approaches to the steepest direction automatically. Furthermore, the small step-sizes are not produced successively. Such property essentially owes to the property: β_k^{PRP} tends to zero as the step-size is sufficiently small. This property was firstly introduced and called Property (∗) by Gilbert and Nocedal [9].

Property (∗) Consider any method of the form (1.1)–(1.2) and suppose that

$$0 < \gamma \le \|g_k\| \le \overline{\gamma} \tag{3.10}$$

for all $k \ge 1$. Then we say that the method has Property (∗) if there exist constants $b > 1$ and $\lambda > 0$ such that, for all k: $|\beta_k| \le b$, and $\|s_{k-1}\| \le \lambda \Rightarrow |\beta_k| \le \frac{1}{2b}$.

The following Lemma shows that the modified method has Property (∗).

Lemma 3.3 *Suppose that Assumptions A and B hold. Consider the method of the form (1.1)–(1.2) with β_k satisfying (2.1) and t_k satisfying (1.3) and (1.5). Then Property (∗) holds.*

Proof Consider any constant γ and $\overline{\gamma}$ of satisfying (3.10). We choose $b = \frac{4\overline{\gamma}^2}{c(1-\sigma)\gamma^2} > 1$, and $\lambda = \frac{(1-2\sigma)\gamma^2}{4bL\overline{\gamma}} > 0$, where $c = \frac{1-2\sigma}{1-\sigma}$ and $0 < \sigma < 1/2$. Then, from (2.1), (2.2) and (3.10), we have

$$\begin{aligned}
|\beta_k^{\mathrm{HS}*}| &= \frac{|g_k^{\mathrm{T}}(g_k - \frac{g_k^{\mathrm{T}} g_{k-1}}{\|g_{k-1}\|^2} g_{k-1})|}{|d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})|} \\
&\le \frac{|g_k^{\mathrm{T}}(g_k - g_{k-1} + g_{k-1} - \frac{g_k^{\mathrm{T}} g_{k-1}}{\|g_{k-1}\|^2} g_{k-1})|}{|(\sigma-1) g_{k-1}^{\mathrm{T}} d_{k-1}|} \\
&\le \frac{\|g_k\| \|(g_k - g_{k-1} + g_{k-1} - \frac{g_k^{\mathrm{T}} g_{k-1}}{\|g_{k-1}\|^2} g_{k-1})\|}{|(\sigma-1) g_{k-1}^{\mathrm{T}} d_{k-1}|} \\
&\le \frac{\|g_k\| (\|g_k - g_{k-1}\| + \|g_{k-1} - \frac{g_k^{T} g_{k-1}}{\|g_{k-1}\|^2} g_{k-1}\|)}{|(\sigma-1) g_{k-1}^{\mathrm{T}} d_{k-1}|} \\
&\le \frac{\|g_k\| (\|g_k - g_{k-1}\| + \frac{|\|g_{k-1}\|^2 - g_k^{\mathrm{T}} g_{k-1}|}{\|g_{k-1}\|^2} \|g_{k-1}\|)}{|(\sigma-1) g_{k-1}^{\mathrm{T}} d_{k-1}|} \\
&\le \frac{\|g_k\| (\|g_k - g_{k-1}\| + \frac{\|g_{k-1} - g_k\|}{\|g_{k-1}\|^2} \|g_{k-1}\|^2)}{|(\sigma-1) g_{k-1}^{\mathrm{T}} d_{k-1}|} \\
&= \frac{2\|g_k\| \|g_k - g_{k-1}\|}{|(\sigma-1) g_{k-1}^{\mathrm{T}} d_{k-1}|} \le \frac{2\|g_k\| (\|g_k\| + \|g_{k-1}\|)}{c(1-\sigma)\|g_{k-1}\|^2} \\
&\le \frac{4\overline{\gamma}^2}{c(1-\sigma)\gamma^2} = b.
\end{aligned}$$

When $\|s_{k-1}\| \le \lambda$, we have from (2.3),

$$\begin{aligned}
|\beta_k^{\mathrm{HS}*}| &\le \frac{\|g_k\| (\|g_k - g_{k-1}\| + \frac{\|g_{k-1} - g_k\|}{\|g_{k-1}\|^2} \|g_{k-1}\|^2)}{|(\sigma-1) g_{k-1}^{\mathrm{T}} d_{k-1}|} \\
&= \frac{2\|g_k\| \|g_k - g_{k-1}\|}{|(\sigma-1) g_{k-1}^{\mathrm{T}} d_{k-1}|} \le \frac{2\|g_k\| \|g_k - g_{k-1}\|}{c(1-\sigma)\|g_{k-1}\|^2} \\
&\le 2\frac{L\lambda \|g_k\|}{(1-2\sigma)\|g_{k-1}\|^2} \le 2\frac{L\lambda\overline{\gamma}}{(1-2\sigma)\gamma^2} = \frac{1}{2b}.
\end{aligned}$$

The proof is completed. □

The next Lemma shows that if the gradients are bounded away from zero, and if the method has Property (∗), then a fraction of the steps cannot be too small. Otherwise we can prove that $\|d_k\|$ increases linearly at most. We will consider groups of Δ consecutive iterates, and for this purpose we define $\kappa_{k,\Delta}^{\lambda} = \{i \in Z^+ : k \le i \le k+\Delta-1, \|s_{i-1}\| \ge \lambda\}$. Let $|\kappa_{k,\Delta}^{\lambda}|$ denote the number of elements of $\kappa_{k,\Delta}^{\lambda}$ and let $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ denote, respectively, the floor and ceiling operations.

Lemma 3.4 *Suppose that Assumptions A and B hold. Consider any method of the form (1.1)–(1.2) with the following three properties: (i) d_k satisfies the sufficient condition (1.6); (ii) t_k satisfies (1.3) and (1.4); (iii) Property (∗) holds. Then if (3.8) holds, there exists $\lambda > 0$ such* *that, for any $\Delta \in Z^+$ and any index k_0, there is a greater index $k \geq k_0$ such that*

$$| \kappa_{k,\Delta}^{\lambda} | > \frac{\Delta}{2}.$$

The proof of Lemma 3.4 can be found in [4]. Finally we give the convergence result for the MHS method with $\beta_k^{\mathrm{HS}*}$. Its proof is similar to Theorem 3.3.3 in [4].

Theorem 3.2 *Suppose that assumptions A and B hold. Consider the method of the form (1.1)–(1.2) with the following four properties: (i) β_k satisfies (2.1); (ii) t_k satisfies (1.3) and (1.4); (iii) d_k satisfies (1.6) with $c = \frac{1-2\sigma}{1-\sigma} > 0$; (iv) Property $(*)$ holds. Then we have*

$$\liminf_{k\to\infty} \|g_k\| = 0.$$

Proof We proceed by contradiction. Since $\inf_{k\to\infty} \|g_k\| > 0$, (3.8) must hold. The MHS method has Property $(*)$ and $\beta_k^{\mathrm{HS}*}$ is nonnegative, therefore, the conditions of Lemmas 3.2 and 3.4 hold. Define $u_i = d_i/\|d_i\|$. We have, for any two indices l, k, with $l \geq k$:

$$x_l - x_k = \sum_{i=k}^{l} \|s_{i-1}\| u_{i-1} = \sum_{i=k}^{l} \|s_{i-1}\| u_{k-1} + \sum_{i=k}^{l} \|s_{i-1}\|(u_{i-1} - u_{k-1}),$$

where $s_{i-1} = x_i - x_{i-1}$. Taking its norms, we obtain

$$\sum_{i=k}^{l} \|s_{i-1}\| \leq \|x_l - x_k\| + \sum_{i=k}^{l} \|s_{i-1}\| \|u_{i-1} - u_{k-1}\|.$$

By Assumptions A and B, we have that the sequence $\{x_k\}$ is bounded, and thus there exists a positive constant ξ such that $\|x_k\| \leq \xi$, for all $k \geq 1$. Thus

$$\sum_{i=k}^{l} \|s_{i-1}\| \leq 2\xi + \sum_{i=k}^{l} \|s_{i-1}\| \|u_{i-1} - u_{k-1}\|. \tag{3.11}$$

Let $\lambda > 0$ be given by Lemma 3.4. Following the notion of this lemma, we define $\Delta = \lceil 8\xi/\lambda \rceil$ to be the smallest integer not less than $8\xi/\lambda$. By Lemma 3.2, there exists an index $k_0 \geq 1$ such that

$$\sum_{i \geq k_0} \|u_k - u_{k-1}\|^2 < \frac{1}{4\Delta}. \tag{3.12}$$

With this Δ and k_0, Lemma 3.4 gives an index $k \geq k_0$ such that

$$|\kappa_{k,\Delta}^{\lambda}| > \frac{\Delta}{2}. \tag{3.13}$$

Next, by the Cauchy-Schwarz inequality and (3.12), we have, for any index $i \in [k, k+\Delta-1]$,

$$\begin{aligned} \|u_{i-1} - u_{k-1}\| &\leq \sum_{j=k}^{i-1} \|u_j - u_{j-1}\| \leq (i-k)^{1/2} (\sum_{j=k}^{i-1} \|u_j - u_{j-1}\|^2)^{1/2} \\ &\leq \Delta^{1/2} (\frac{1}{4\Delta})^{1/2} = \frac{1}{2}. \end{aligned}$$

By this relation, (3.13) and (3.11) with $l = k + \Delta - 1$, we have

$$2\xi \geq \frac{1}{2} \sum_{i=k}^{k+\Delta-1} \|s_{i-1}\| > \frac{\lambda}{2} |\kappa_{k,\Delta}^{\lambda}| > \frac{\lambda\Delta}{4}.$$

Thus $\Delta < \frac{8\xi}{\lambda}$, which contradicts the definition of Δ. Therefore, the result follows. □

4. Numerical experiments

In this section we compare MHS (modified Hestenes-Stiefel) and HS (Hestenes-Stiefel) methods with the PRP method on the collection of test problems that are from [15]. The parameters in the strong Wolfe-Powell line search were chosen to be $\delta = 10^{-2}$ and $\sigma = 0.1$. For each test problem, the termination criterion is $\|g(x_k)\| \leq 10^{-5}$.

In order to rank the iterative numerical methods, we can compute the total numbers of function and gradient evaluations by the formula

$$N_{\text{total}} = \text{NF} + m * \text{NG}, \tag{4.1}$$

where NF, NG denote the number of function evaluations and gradient evaluations, respectively, and m is some integer. According to the results on automatic differentiation [8], the value of m can be set to $m = 5$. That is to say, one gradient evaluation is equivalent to m number of function evaluations if by automatic differentiation.

As we know that the PRP method is considered to be the most efficient one among all the conjugate gradient methods. Therefore, in this part, we compare HS and MHS methods with the PRP method as follows: for each testing example i, compute the total numbers of function evaluations and gradient evaluations required by the evaluated method j (EM(j)) and the PRP method by formula (4.1), and denote them by $N_{\text{total},i}(\text{EM}(j))$ and $N_{\text{total},i}(\text{PRP})$; then calculate the ratio

$$r_i(\text{EM}(j)) = \frac{N_{\text{total},i}(\text{EM}(j))}{N_{\text{total},i}(\text{PRP})}.$$

If EM(j_0) does not work for example i_0, then we replace the $N_{\text{total},i_0}(\text{EM}(j_0))$ by a positive constant τ which is defined as $\tau = \max\{r_i(\text{EM}(j)) : \ (i,j) \notin S_1\}$, where

$$S_1 = \{(i,j) : \text{method } j \text{ does not work for example } i\}.$$

The geometric meaning of these ratios for method EM(j) over all the test problems is defined by

$$r(\text{EM}(j)) = (\prod_{i \in S} r_i(\text{EM}(j)))^{1/|S|},$$

where S denotes the set of the test problems and $|S|$ is the cardinality of S. One advantage of the above rule is that, the comparison is relative and hence is not dominated by a few problems for which the method requires a great deal of function evaluations and gradient functions.

The numerical results are summarized in Table 4.1. In Table 4.1, Problem denotes the name of the test problem in MATLAB, Dim denotes the dimension of the problem, NI denotes the number of iterations, NF denotes the number of function evaluations, NG denotes the number of gradient evaluations, $-$ denotes the failure to compute the step-size under the same given conditions.

		PRP	HS	MHS
Problem	Dim	NI/NF/NG	NI/NF/NG	NI/NF/NG
ROSE	2	29/502/65	40/574/117	33/467/71
FROTH	2	12/30/20	7/68/14	15/36/26
BADSCP	2	-	-	-
BADSCB	2	13/80/22	13/80/24	-
BEALE	2	9/126/21	11/81/21	19/94/34
JENSAM	2	-	-	11/30/20
HELIX	3	49/255/83	26/258/50	41/197/73
BARD	3	23/98/37	65/277/105	82/161/121
GAUSS	3	4/57/6	3/8/6	4/9/5
MEYER	3	-	-	-
GULF	3	1/2/2	1/2/2	1/2/2
BOX	3	-	-	-
SING	4	199/611/338	110/425/176	123/445/197
WOOD	4	169/1103/302	265/1500/466	96/419/170
KOWOSB	4	55/300/94	193/676/310	28/297/46
BD	4	-	-	-
OSB1	5	-	-	-
BIGGS	6	264/875/423	155/502/253	-
OSB2	11	254/1061/418	185/799/308	3908/15695/6505
OSB3	20	-	2380/7312/3799	-
ROSEX	8	23/402/59	38/510/108	27/193/66
	50	31/533/77	25/359/60	39/537/103
	100	-	-	38/436/82
SINGX	4	199/611/338	110/425/176	123/445/197
PEN1	2	-	-	5/18/12
PEN2	4	12/134/28	11/177/31	17/142/34
	50	613/2795/1063	1099/3541/1806	141/1105/260
VARDIM	2	3/9/7	2/52/5	3/9/7
	50	-	-	10/52/36
TRIG	3	12/81/24	12/84/26	13/177/23
	50	41/279/72	38/224/68	37/224/70
	100	46/342/87	50/393/92	51/436/91
BV	3	12/25/16	10/20/13	12/25/16
	10	75/241/117	50/148/81	47/188/70
IE	3	5/12/7	5/12/7	5/12/7
	50	6/13/7	6/13/7	6/13/7
	100	6/13/8	6/13/8	6/13/8
	200	6/13/8	6/13/8	6/61/8
	500	6/13/8	6/13/8	6/13/8

Table 4.1 Test results for PRP, HS and MHS methods

		PRP	HS	MHS
Problem	Dim	NI/NF/NG	NI/NF/NG	NI/NF/NG
TRID	3	10/75/16	12/29/17	15/85/22
	50	26/55/31	27/57/31	26/103/30
	100	30/67/36	30/67/36	30/67/36
	200	30/66/36	30/66/37	32/70/38
BAND	3	9/68/13	10/70/14	7/64/12
	50	-	-	21/866/28
	100	-	-	21/865/28
	200	-	-	21/827/29
LIN	2	1/3/3	1/3/3	1/3/3
	50	1/3/3	1/3/3	1/3/3
	500	1/3/3	1/3/3	1/3/3
	1000	1/3/3	1/3/3	1/3/3
LIN1	2	1/51/2	1/51/2	1/51/2
	10	1/3/3	1/3/3	1/3/3
LIN0	4	1/3/3	1/3/3	1/3/3

Table 4.1 (continuous) Test results for PRP, HS and MHS methods

PRP	HS	MHS
1	1.028	0.976

Table 4.2 Relative efficiency of PRP, HS and MHS methods

According to the above rule, it is clear that r(PRP)= 1. The values of r(HS) and r(MHS) are listed in Tables 4.2. From Table 4.2 we can see that the average performances of the HS method is similar to PRP method, and the average performances of the MHS method is a little better than PRP and HS methods. An explanation of this behavior may be that we take a little modification to the HS method such that the generated direction always satisfies sufficient descent condition (1.6), whereas the search directions generated by PRP and HS methods do not guarantee to satisfy the descent property (1.7) for some problems.

5. Conclusion

In this paper, we have proposed a new conjugate gradient algorithm for solving unconstrained optimization problems. The new method is a modification of the HS conjugate gradient method such that the direction generated by the resulting algorithm satisfies the sufficient descent condition. The global convergence result of the modified method with strong Wolfe-Powell line search is established. Preliminary numerical results show that the performance of the MHS method is a little more efficient than PRP and HS methods for given test problems.

References

[1] AL-BAALI M. *Descent property and global convergence of the Fletcher-Reeves method with inexact line search* [J]. IMA J. Numer. Anal., 1985, **5**(1): 121–124.
[2] ANDREI N. *Scaled conjugate gradient algorithms for unconstrained optimization* [J]. Comput. Optim. Appl., 2007, **38**(3): 401–416.
[3] BIRGIN E G, MARTÍNEZ J M. *A spectral conjugate gradient method for unconstrained optimization* [J]. Appl. Math. Optim., 2001, **43**(2): 117–128.
[4] DAI Yuhong, YUAN Yaxiang. *Nonlinear Conjugate Gradient Methods* [M]. Shanghai: Shanghai Scientific and Technical Publishers, 2000.
[5] DAI Yuhong, YUAN Yaxiang. *Further studies on the Polak-Ribière-Polyak method* [R]. Research Report ICM-95-040, Institute of Computational Mathematics and Scientific/ Engineering Computing, Chinese Academy of Sciences, 1995.
[6] DAI Yuhong, NI Qin. *Testing different conjugate gradient methods for large-scale unconstrained optimization* [J]. J. Comput. Math., 2003, **21**(3): 311–320.
[7] FLETCHER R, REEVES C M. *Function minimization by conjugate gradients* [J]. Comput. J., 1964, **7**: 149–154.
[8] GRIEWANK A. *On Automatic Differentiation* [M]. SCIPRESS, Tokyo, 1989.
[9] GILBERT J C, NOCEDAL J. *Global convergence properties of conjugate gradient methods for optimization* [J]. SIAM J. Optim., 1992, **2**(1): 21–42.
[10] GRIPPO L, LUCIDI S. *A globally convergent version of the Polak-Ribière conjugate gradient method* [J]. Math. Programming, Ser.A, 1997, **78**(3): 375–391.
[11] HESTENES M R, STIEFEL E. *Methods of conjugate gradients for solving linear systems* [J]. J. Research Nat. Bur. Standards, 1952, **49**: 409–436.
[12] HAGER W W, ZHANG Hongchao. *A new conjugate gradient method with guaranteed descent and an efficient line search* [J]. SIAM J. Optim., 2005, **16**(1): 170–192.
[13] KHODA K M, LIU Y, STOREY C. *Generalized Polak-Ribière algorithm* [J]. J. Optim. Theory Appl., 1992, **75**(2): 345–354.
[14] LI Guoyin, TANG Chunming, WEI Zengxin. *New conjugacy condition and related new conjugate gradient methods for unconstrained optimization* [J]. J. Comput. Appl. Math., 2007, **202**(2): 523–539.
[15] MORÉ J J, GARBOW B S, HILLSTROM K E. *Testing unconstrained optimization software* [J]. ACM Trans. Math. Software, 1981, **7**(1): 17–41.
[16] NOCEDAL J. *Theory of Algorithms for Unconstrained Optimization* [M]. Cambridge Univ. Press, Cambridge, 1992.
[17] POWELL M J D. *Nonconvex Minimization Calculations and the Conjugate Gradient Method* [M]. Springer, Berlin, 1984.
[18] POLAK E, RIBIÈRE G. *Note sur la convergence de méthodes de directions conjuguées* [J]. Revue Française d'Informatique et de Recherche Opérationnelle, 1969, (16): 35–43. (in France)
[19] POLYAK B T. *The conjugate gradient method in extreme problems* [J]. USSR Comp. Math. Math. Phys., 1969, **9**: 94–112.
[20] RAYDAN M. *The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem* [J]. SIAM J. Optim., 1997, **7**(1): 26–33.
[21] WOLFE P. *Convergence conditions for ascent methods* [J]. SIAM Rev., 1969, **11**: 226–235.
[22] WOLFE P. *Convergence conditions for ascent methods. II. Some corrections* [J]. SIAM Rev., 1971, **13**: 185–188.
[23] WEI Zengxin, LI Guoyin, QI Liqun. *New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems* [J]. Appl. Math. Comput., 2006, **179**(2): 407–430.
[24] WEI Zengxin, LI Guoyin, QI Liqun. *Global convergence of the Polak-Ribière-Polyak conjugate gradient method with an Armijo-type inexact line search for nonconvex unconstrained optimization problems* [J]. Math. Comp., 2008, **77**(264): 2173–2193.
[25] WEI Zengxin, YAO Shengwei, LIU Liying. *The convergence properties of some new conjugate gradient methods* [J]. Appl. Math. Comput., 2006, **183**(2): 1341–1350.
[26] YUAN Yaxiang. *Analysis on the conjugate gradient method* [J]. Optim. Methods Softw., 1993, **2**: 19–29.
[27] ZOUTENDIJK G. *Nonlinear Programming, Computational Methods* [M]. North-Holland, Amsterdam, 1970.