An International Journal of Optimization
and Control: Theories & Applications
Vol.3, No.2, pp.145-152 (2013) © IJOCTA
ISSN:2146-0957 eISSN:2146-5703
DOI:10.11121/ijocta.01.2013.00141
http://www.ijocta.com

Modification of the Armijo line search to satisfy the convergence properties of HS method

Mohammed Belloufi[a], Rachid Benzine[b], and Yamina Laskri[c]

[a]Department of mathematics, Mohamed chérif Messaadi University Souk Ahras, Algeria
Email: mbelloufi_doc@yahoo.fr

[b]Department of mathematics, Badji Mokhtar University Annaba , Algeria
Email: Rachid.Benzine@univ-annaba.org

[c]Department of mathematics, Badji Mokhtar University Annaba , Algeria
Email: yamina.laskri@univ-annaba.org

(Received October 01, 2012; in final form May 26, 2013)

Abstract. The Hestenes-Stiefel (HS) conjugate gradient algorithm is a useful tool of unconstrained numerical optimization, which has good numerical performance but no global convergence result under traditional line searches. This paper proposes a line search technique that guarantee the global convergence of the Hestenes-Stiefel (HS) conjugate gradient method. Numerical tests are presented to validate the different approaches.

Keywords: Unconstrained optimization; conjugate gradient method; line search algorithm.
AMS Classification: 90C06, 90C30, 65K05

1. Introduction

Consider the following unconstrained optimization problem:

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

where f is continuously differentiable and its gradient $g(x) = \nabla f(x)$ is available. Iterative methods are widely used for solving (1) and the iterative formula is given by

$$x_{k+1} = x_k + \alpha_k d_k \tag{2}$$

where $x_k \in \mathbb{R}^n$is the k th approximation to the solution, α_k is a steplength obtained by carrying out a line search, and d_k is a search direction.

There are many kinds of iterative methods that include the Newton method, the steepest descent method and nonlinear conjugate gradient methods, for example. The conjugate gradient methods are the most famous methods for solving unstrained optimization (1), especially in case of large scale optimization problems in scientific and engineering computation due to the simplicity of their iteration and low memory requirements. The search direction d_k is defined by

$$d_k = \begin{cases} -g_k & \text{si } k = 0 \\ -g_k + \beta_k d_{k-1} & \text{si } k \geq 1 \end{cases} \tag{3}$$

where β_k is a scalar and $g_k = g(x_k)$. The original nonlinear conjugate gradient method proposed by Hestenes and Stiefel (HS conjugate gradient method) [15], in which β_k is defined by

$$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})} \tag{4}$$

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

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} \qquad \text{Fletcher-Reeves [9]} \tag{5}$$

Mohammed Belloufi. Email: mbelloufi_doc@yahoo.fr.

$$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T g_{k-1}} \text{ Polak-Ribière [19,20]} \quad (6)$$

$$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}} \quad \text{Conjugate-Descent [8]} \quad (7)$$

$$\beta_k^{LS} = -\frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T g_{k-1}} \quad \text{Liu-Storey [17]} \quad (8)$$

$$\beta_k^{DY} = \frac{g_k^T g_k}{d_{k-1}^T(g_k - g_{k-1})} \quad \text{Dai-Yuan [6]} \quad (9)$$

Considerable attentions have been made on the global convergence for the above methods. Zoutendijk [31] proved that the FR method with exact line search is globally convergent. Al-Baali [2] extended this result to the strong Wolfe-Powell line search. Powell [21] proved that the sequence of gradient norms $\|g_k\|$ could be bounded away from zero only when

$$\sum_{k\geq 0} \frac{1}{\|d_k\|} < +\infty \quad (10)$$

So one can prove that the FR method is globally convergent for general functions by using (10). However, the global convergence has not been established for the PRP method with the strong Wolfe-Powell line search conditions. In fact, Powell proved that even if the steplength was chosen to be the least positive minimizer of the one variable function ($\Phi_k(\alpha) = f(x_k + \alpha d_k)$, $\alpha \in \mathbb{R}$), the PRP method could cycle infinitely without approaching a solution.

Some convergent versions were proposed by using some new complicated line searches or through restricting the parameter β_k to a non-negative number [12,13,25,26,27]. The CD method was proved to have global convergence property under strong Wolfe line search with a strong restriction on the parameters [5] and DY method has global convergence under weak Wolfe line search [6]. Some impressive literature on conjugate gradient methods can be found in [4,5,7,10,11,16,17,22,23,30].

However, to the best of our knowledge, the global convergence of the original LS and HS methods has not been proved under all the mentioned line searches. In this paper, we propose a new line search procedure and investigate the global convergence of the original HS method.

Under the sufficient descent condition

$$g_k^T d_k < -c\|g_k\|^2 \quad (11)$$

for some constant $c \in]0,1[$

Once the descent direction d_k is determined at the k-th iteration, we should seek a step size along the descent direction and complete one iteration.

There are many approaches for finding an available step size. Among them the exact line search is an ideal one, but is cost-consuming or even impossible to use to find the step size. Some inexact line searches are sometimes useful and effective in practical computation, such as Armijo line search [1], Goldstein and Wolfe line search [8,14,28,29]. The Armijo line search is commonly used and easy to implement in practical computation.

Armijo line search

Let $s > 0$ be a constant, $\rho \in (0,1)$ and $\mu \in (0,1)$. Choose α_k to be the largest α in $\{s, s\rho, s\rho^2, ...,\}$ such that

$$f_k - f(x_k + \alpha d_k) \geq -\alpha g_k^T d_k.$$

The drawback of the Armijo line search is how to choose the initial step size s. If s is too large then the procedure needs to call much more function evaluations. If s is too small then the efficiency of related algorithm will be decreased. Thereby, we should choose an adequate initial step size s at each iteration so as to find the step size α_k easily.

In this paper we propose a new Armijo-modified line search in which an appropriate initial step size s is defined and varies at each iteration.

The new Armijo-modified line search enables us to find the step size α_k easily at each iteration and guarantees the global convergence of the original HS conjugate gradient method under some mild conditions.

The global convergence and linear convergence rate are analyzed and numerical results show that HS method with the new Armijo-modified line search is more effective, than other similar methods in solving large scale minimization problems.

2. New Armijo-Modified Line Search

We first assume that

Asumption A. The objective function $f(x)$ is continuously differentiable and has a lower bound on $\mathbb{R}^n$

Asumption B. The gradient $g(x) = \nabla f(x)$ of $f(x)$ is Lipschitz continuous on an open convex set Γ that contains the level set $L(x_0) = \{x \in \mathbb{R}^n \,|f(x) \leq f(x_0)\}$ with x_0 given, i.e., there exists an

$L > 0$ such that

$$\|g(x) - g(y)\| \leq L \|x - y\|, \forall x, y \in \Gamma.$$

New Armijo-modified line search

Given $\mu \in \left]0, \frac{1}{2}\right[$, $\rho \in]0,1[$ and $c \in \left[\frac{1}{2}, 1\right]$, set $s_k = \frac{1-c}{L_k} \frac{d_K^T(g_{k+1}-g_k)}{\|d_k\|^2}$ and α_k is the largest α in $\{s, s\rho s\rho^2, ...,\}$ such that

$$f_k - f(x_k + \alpha d_k) \geq -\alpha \mu g_k^T d_k.$$

$$g(x_k + \alpha d_k)^T d(x_k + \alpha d_k) \leq -c \|g(x_k + \alpha d_k)\|^2$$

and $\|g_{k+1}\| \leq c \|d_k\|$ where

$$d(x_k + \alpha d_k) = -g(x_k + \alpha d_k) + \frac{g(x_k + \alpha d_k)^T [g(x_k + \alpha d_k) - g(x_k)]}{d_k^T [g(x_k + \alpha d_k) - g(x_k)]} d_k$$

for $d_k^T(g_{k+1} - g_k) > 0$ and L_k is an approximation to the Lipschitz constant L of $g(x)$.

3. Algorithm and Convergent Properties

In this section, we will reintroduce the convergence properties of the HS method

Now we give the following algorithm firstly.

Algorithm 1. *Step 0: Given* $x_0 \in \mathbb{R}^n$, *set* $d_0 = g_0$, $k := 0$.

Step 1: If $\|g_k\| = 0$ *then stop else go to Step 2.*

Step 2: Set $x_{k+1} = x_k + a_k d_k$ *where* d_k *is defined by (3),* $\beta_k = \beta_k^{HS}$ *and* α_k *is defined by the new Armijo-modified line search.*

Step 3: Setk $:= k + 1$ *and go to Step 1.*

Some simple properties of the above algorithm are given as follows.

Lemma 1. *Assume that (A) and (B) hold and the HS method with the new Armijo-modified line search generates an infinite sequence* $\{x_k\}$.

If $\alpha_k \leq \frac{1-c}{L} \frac{d_k^T(g_{k+1}-g_k)}{\|d_k\|^2}$ *then:*

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

Proof. By the condition (B), the Cauchy–Schwartz inequality and the HS method, we have

$$\begin{aligned}(1-c)d_k^T(g_{k+1} - g_k) &\geq \alpha_k L \|d_k\|^2 \\ &= \frac{\alpha_k L \|g_{k+1}\| \|d_k\|}{\|g_{k+1}\|^2} \|g_{k+1}\| \|d_k\| \\ &\geq \frac{\|g_{k+1}\| \|g_{k+1} - g_k\|}{\|g_{k+1}\|^2} \|g_{k+1}\| \|d_k\| \\ &\geq \frac{|g_{k+1}^T(g_{k+1}-g_k)|}{\|g_{k+1}\|^2} |g_{k+1}^T d_k| \\ &\geq \frac{g_{k+1}^T(g_{k+1}-g_k)}{d_k^T(g_{k+1}-g_k)} \frac{d_k^T(g_{k+1}-g_k)}{\|g_{k+1}\|^2} (g_{k+1}^T d_k) \\ &= \beta_{k+1}^{HS} \frac{d_k^T(g_{k+1}-g_k)}{\|g_{k+1}\|^2} (g_{k+1}^T d_k)\end{aligned}$$

Therefore

$$(1-c)\, d_k^T(g_{k+1} - g_k) \geq \beta_{k+1}^{HS} \frac{d_k^T(g_{k+1}-g_k)}{\|g_{k+1}\|^2} (g_{k+1}^T d_k)$$

and thus

$$\begin{aligned}-c\|g_{k+1}\|^2 &\geq -\|g_{k+1}\|^2 + \beta_{k+1}^{HS} (g_{k+1}^T d_k) \\ &= g_{k+1}^T d_{k+1}\end{aligned}$$

The proof is finished. □

Lemma 2. *Assume that (A) and (B) hold. Then the new Armijo-modified line search is well defined.*

Proof. On the one hand, since

$\lim_{\alpha \to 0} \frac{f_k - f(x_k + \alpha d_k)}{\alpha} = -g_k^T d_k > -\mu g_k^T d_k$

there is an $\alpha_k^{'} > 0$ such that

$\frac{f_k - f(x_k + \alpha d_k)}{\alpha} \geq -\mu g_k^T d_k, \forall \alpha \in \left[0, \alpha_k^{'}\right]$.

Thus, letting $\alpha_k^{"} = min(s_k, \alpha_k^{'})$ yields

$$\frac{f_k - f(x_k + \alpha d_k)}{\alpha} \geq -\mu g_k^T d_k, \forall \alpha \in \left[0, \alpha_k^{"}\right]. \tag{13}$$

On the other hand, by Lemma 1, we can obtain

$g(x_k + \alpha d_k)^T d(x_k + \alpha d_k) \leq -c \|g(x_k + \alpha d_k)\|^2$.

if $\alpha \leq \frac{1-c}{L} \frac{d_k^T(g_{k+1}-g_k)}{\|d_k\|^2}$.Letting

$\overline{\alpha_k} = \min\left(\alpha_k^{"}, \frac{1-c}{L} \frac{d_k^T(g_{k+1}-g_k)}{\|d_k\|^2}\right)$.

We can prove that the new Armijo-modified line search is well defined when $\alpha \in [0, \overline{\alpha_k}]$. The proof is completed. □

4. Global Convergence

Lemma 3. *Assume that (A) and (B) hold and the HS method with the new Armijo-modified line search generates an infinite sequence* $\{x_k\}$ *and there exist* $m_0 > 0$ *and* $M_0 > 0$ *such that* $m_0 \leq L_k \leq M_0$. *Then,*

$$\|d_k\| \leq \left(1 + \frac{L(1-c)}{m_0}\right) \|g_k\|, \forall k. \tag{14}$$

Proof. For $k = 0$ we have

$\|d_k\| = \|g_k\| \leq \|g_k\| \left(1 + \frac{L(1-c)}{m_0}\right)$

For $k > 0$, by the procedure of the new Armijo-modified line search, we have

$\alpha_k \leq s_k = \frac{1-c}{L_k} \frac{d_k^T(g_{k+1}-g_k)}{\|d_k\|^2} \leq \frac{1-c}{m_0} \frac{d_k^T(g_{k+1}-g_k)}{\|d_k\|^2}$

By the Cauchy–Schwartz inequality, the above inequality and noting the HS formula, we have

$$\|d_{k+1}\| = \left\|-g_{k+1} + \beta_{k+1}^{HS} d_k\right\|$$
$$\leq \|g_{k+1}\| + \frac{\left|g_{k+1}^T(g_{k+1}-g_k)\right|}{\left|d_k^T(g_{k+1}-g_k)\right|}\|d_k\|$$
$$\leq \|g_{k+1}\|\left[1 + L\alpha_k \frac{\|d_k\|^2}{d_k^T(g_{k+1}-g_k)}\right]$$
$$\leq \left(1 + L\frac{(1-c)}{m_0}\right)\|g_{k+1}\|$$

The proof is completed. □

Theorem 1. *Assume that (A) and (B) hold, the HS method with the new Armijo-modified line search generates an infinite sequence $\{x_k\}$ and there exist $m_0 > 0$ and $M_0 > 0$ such that $m_0 \leq L_k \leq M_0$. Then*

$$\lim_{k\to\infty} \|g_k\| = 0. \quad (15)$$

Proof. Let $\eta_0 = inf_{\forall k}\{\alpha_k\}$.

If $\eta_0 > 0$ then we have

$$f_k - f_{k+1} \geq -\mu\alpha_k g_k^T d_k \geq \mu\eta_0 c\|g_k\|^2.$$

By **(A)** we have

$$\sum_{k=0}^{+\infty} \|g_k\|^2 < +\infty$$

and thus,

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

In the following, we will prove that $\eta_0 > 0$. For the contrary, assume that $\eta_0 = 0$.
Then, there exists an infinite subset $K \subseteq \{0, 1, 2, ..., \}$ such that

$$\lim_{k\in K, k\to\infty} \alpha_k = 0 \quad (16)$$

By Lemma 3 we obtain

$$s_k = \frac{1-c}{L_k}\frac{d_k^T(g_{k+1}-g_k)}{\|d_k\|^2} \geq \frac{1-c}{m_0}\left(1 + \frac{L(1-c)}{m_0}\right)^{-2} > 0$$

Therefore, there is a k' such that

$\alpha_k/\rho \leq s_k$, $k \geq k'$ and $k \in K$.

Let $\alpha = \alpha_k/\rho$, at least one of the following two inequalities

$$f_k - f(x_k + \alpha d_k) \geq -\alpha\mu g_k^T d_k \quad (17)$$

and

$$g(x_k + \alpha d_k)^T d(x_k + \alpha d_k) \leq -c\|g(x_k + \alpha d_k)\|^2 \quad (18)$$

does not hold. If (17) does not hold, then we have

$$f_k - f(x_k + \alpha d_k) < -\alpha\mu g_k^T d_k.$$

Using the mean value theorem on the left-hand side of the above inequality, there exists $\theta_k \in [0, 1]$ such that

$$-\alpha g(x_k + \alpha d_k)^T d_k < -\alpha\mu g_k^T d_k$$

Thus

$$g(x_k + \alpha d_k)^T d_k > \mu g_k^T d_k \quad (19)$$

By (B), the Cauchy–Schwartz inequality, (19) and Lemma 1, we have

$$L\alpha\|d_k\|^2 \geq \|g(x_k + \alpha\theta_k d_k) - g(x_k)\|\,\|d_k\|$$
$$\geq \left|(g(x_k + \alpha\theta_k d_k) - g(x_k))^T \; d_k\right|$$
$$\geq -(1-\mu)g_k^T d_k$$
$$\geq c(1-\mu)\|g_k\|^2.$$

We can obtain from Lemma 3 that

$$\alpha_k \geq \frac{c\rho(1-\mu)}{L}\frac{\|g_k\|^2}{\|d_k\|^2}$$
$$\geq \frac{c\rho(1-\mu)}{L}\frac{1}{\left(\frac{1}{1+\frac{L(1-\mu)}{m_0}}\right)^2} > 0,$$

$k \geq k', k \in K$.

which contradicts (16).

If (18) does not hold, then we have

$$g(x_k + \alpha d_k)^T d(x_k+\alpha d_k) > -c\|g(x_k + \alpha d_k)\|^2.$$

and thus,

$$\frac{g(x_k+\alpha d_k)^T[g(x_k+\alpha d_k)-g_k]}{d_k^T[g(x_k+\alpha d_k)-g_k]} g(x_k + \alpha d_k)^T d_k$$
$$> (1-c)\|g(x_k + \alpha d_k)\|^2.$$

By using the Cauchy–Schwartz inequality on the left-hand side of the above inequality we have

$$\alpha L\frac{\|d_k\|^2}{d_k^T(g(x_k+\alpha d_k)-g_k)} > (1-c)$$

Combining Lemma 3 we have

$$\alpha_k > \frac{\rho(1-c)}{L}\frac{d_k^T(g(x_k+\alpha d_k)-g_k)}{\left(1+\frac{L(1-\mu)}{m_0}\right)^2\|g_k\|^2} > 0,$$

$k \geq k', k \in K$.

which also contradicts (16). This shows that $\eta_0 > 0$. The whole proof is completed. □

5. Linear Convergence Rate

In this section we shall prove that the HS method with the new Armijo-modified line search has linear convergence rate under some mild conditions.

We further assume that

Asumption C. The sequence $\{x_k\}$ generated by the HS method with the new Armijo-type line search converges to x^*, $\nabla^2 f(x^*)$ is a symmetric positive definite matrix and $f(x)$ is twice continuously differentiable on

$$N(x^*, \varepsilon_0) = \{x/\|x - x^*\| < \varepsilon_0\}.$$

Lemma 4. *Assume that Asumption (C) holds. Then there exist $m^{'}, M^{'}$ and ε_0 with $0 < m^{'} \leq M^{'}$ and $\varepsilon \leq \varepsilon_0$ such that*

$$m^{'} \|y\|^2 \leq y^T \nabla^2 f(x) y \leq M^{'} \|y\|^2, \quad \forall x, y \in N(x^*, \varepsilon) \tag{20}$$

$$\tfrac{1}{2} m^{'} \|x - x^*\|^2 \leq f(x) - f(x^*) \leq h \|x - x^*\|^2, \quad h = \frac{1}{2} M^{'}, \forall x \in N(x^*, \varepsilon) \tag{21}$$

$$M^{'} \|x - y\|^2 \geq (\triangle\, g)^T (x - y) \geq m^{'} \|x - y\|^2, \quad \triangle\, g = g(x) - g(x^*), \forall x, y \in N(x^*, \varepsilon) \tag{22}$$

and thus

$$M^{'} \|x - x^*\|^2 \geq g(x)^T (\triangle\, x) \geq m^{'} \|x - x^*\|^2, \quad \triangle\, x = (x - x^*), \forall x \in N(x^*, \varepsilon) \tag{23}$$

By (23) and (22) we can also obtain, from the Cauchy–Schwartz inequality, that

$$M^{'} \|x - x^*\| \geq \|g(x)\| \geq m^{'} \|x - x^*\|, \quad \forall x \in N(x^*, \varepsilon) \tag{24}$$

and

$$\|g(x) - g(y)\| \leq m^{'} \|x - y\|, \quad \forall x, y \in N(x^*, \varepsilon) \tag{25}$$

Proof. Its proof can be seen from the literature (e.g. [29]). □

Theorem 2. *Assume that Asumption (C) holds, the HS method with the new Armijo-type line search generates an infinite sequence $\{x_k\}$ and there exist $m^{'} > 0$ and $M^{'} > 0$ such that $m_0 \leq L_k \leq M_0$. Then $\{x_k\}$ converges to x^* at least R-linearly.*

Proof. Its proof can be seen from the literature (e.g. [26]). □

6. Numerical Reports

In this section, we shall conduct some numerical experiments to show the efficiency of the new Armijo-modified line search used in the HS method.

The Lipschitz constant L of $g(x)$ is usually not a known priori in practical computation and needs to be estimated. In the sequel, we shall discuss the problem and present some approaches for estimating L. In a recent paper [24], some approaches for estimating L were proposed. If $k \geq 1$ then we set $\delta_{k-1} = x_k - x_{k-1}$ and $y_{k-1} = g_k - g_{k-1}$ and obtain the following three estimating formula

$$L \simeq \frac{\|y_{k-1}\|}{\|\delta_{k-1}\|}, \tag{31}$$

$$L \simeq \frac{\|y_{k-1}\|^2}{\delta_{k-1}^T y_{k-1}}, \tag{32}$$

$$L \simeq \frac{\delta_{k-1}^T y_{k-1}}{\|\delta_{k-1}\|^2}. \tag{33}$$

In fact, if L is a Lipschitz constant then any $L^{'}$ greater than L is also a Lipschitz constant, which allows us to find a large Lipschitz constant. However, a very large Lipschitz constant possibly leads to a very small step size and makes the HS method with the new Armijo-modified line search converge very slowly. Thereby, we should seek as small as possible Lipschitz constants in practical computation.

In the k-th iteration we take the Lipschitz constants as respectively

$$L_k = \max\left(L_0, \frac{\|y_{k-1}\|}{\|\delta_{k-1}\|}\right), \tag{34}$$

$$L_k = \max\left(L_0, \min\left(\frac{\|y_{k-1}\|^2}{\delta_{k-1}^T y_{k-1}}, M_0^{'}\right)\right), \tag{35}$$

$$L_k = \max\left(L_0, \frac{\delta_{k-1}^T y_{k-1}}{\|\delta_{k-1}\|^2}\right), \tag{36}$$

with $L_0 > 0$ and $M_0^{'}$ being a large positive number.

Lemma 5. *Assume that (H1) and (H2) hold, the HS method with the new Armijo-modified line search generates an infinite sequence $\{x_k\}$ and L_k is evaluated by (34), (35) or (36). Then, there exist $m_0 > 0$ and $M_0 > 0$ such that*

$$m_0 = L_k = M_0 \tag{37}$$

Proof. Obviously, $L_k = L_0$, and we can take $m_0 = L_0$. For (34) we have

$$L_k = \max\left(L_0, \tfrac{\|y_{k-1}\|}{\|\delta_{k-1}\|}\right) \leq \max(L_0, L)$$

For (35) we have

$$L_k = \max\left(L_0, \min\left(\tfrac{\|y_{k-1}\|^2}{\delta_{k-1}^T y_{k-1}}, M_0^{'}\right)\right) \leq \max\left(L_0, M_0^{'}\right)$$

For (36), by using the Cauchy–Schwartz inequality, we have

$$L_k = \max\left(L_0, \frac{\delta_{k-1}^T y_{k-1}}{\|\delta_{k-1}\|^2}\right) \leq \max(L_0, L).$$

By letting $M_0^{'} = \max(L_0, L, M_0^{'})$, we complete the proof. □

HS1, HS2, and HS3 denote the HS methods with the new Armijo-modified line search corresponding to the estimations (34)–(36), respectively. HS denotes the original HS method with strong Wolfe line search. PRP+ denotes the PRP method with

$$\beta_k = \max\left(0, \beta_k^{PRP}\right)$$

and strong Wolfe line search .

Birgin and Martinez developed a family of scaled conjugate gradient algorithms, called the spectral conjugate gradient method (abbreviated as SCG) [3]. Numerical experiments showed

that some special SCG methods were effective. In one SCG method, the initial choice of α at the $k-$th iteration in SCG method was

$$\overline{\alpha}\left(k, d_k, d_{k-1}, \alpha_{k-1}\right) = \begin{cases} 1 & \text{if } k = 0 \\ \alpha_{k-1}\frac{\|d_{k-1}\|}{\|d_k\|} & \text{otherwise} \end{cases} \tag{38}$$

Table 1. Number of iterations and number of functional evaluations

P	n	HS1	HS2	HS3	HS
21	10^4	31/109	36/124	54/123	56/176
22	10^4	32/122	50/217	56/276	61/427
23	10^4	26/103	30/126	32/119	39/196
24	10^4	33/136	46/234	53/270	80/295
25	10^4	44/127	43/221	56/227	44/125
26	10^4	35/142	31/105	37/127	51/196
27	10^4	32/210	44/229	36/196	50/237
28	10^4	63/321	66/296	74/280	82/335
29	10^4	31/132	27/209	30/130	44/209
30	10^4	42/291	37/126	36/160	41/279
31	10^4	61/196	76/179	67/231	74/279
32	10^4	71/221	74/339	81/341	85/419
33	10^4	35/211	41/133	47/161	35/134
34	10^4	40/217	33/327	36/235	56/383
35	10^4	52/276	66/281	51/357	70/274
CPU	-	89 s	120 s	180 s	257 s

We chose 15 test problems (Problems 21–35) with the dimension n = 10 000 and initial points from the literature [18] to implement the HS method with the new Armijo-modified line search. We set the parameters as $\mu = 0.25$, $\rho = 0.75$, $c = 0.75$ and $L_0 = 1$ in the numerical experiment. five conjugate gradient algorithms (HS1, HS2, HS3, HS and HS+) are compared in numerical performance.

The stop criterion is $\|g_k\| \le 10^{-8}$, and the numerical results are given in Table 1.

In Table 1, CPU denotes the total CPU time (seconds) for solving all the 15 test problems. A pair of numbers means the number of iterations and the number of functional evaluations. It can be seen from Table 1 that the HS method with the new Armijo-modified line search is effective for solving some large scale problems. In particular, method HS1 seems to be the best one among the five algorithms because it uses the least number of iterations and functional evaluations when the algorithms reach the same precision. This shows that the estimating formula (34) may be more reasonable than other formula. In fact, if $\delta_{k-1}^T y_{k-1} > 0$ then we have

$$\frac{\delta_{k-1}^T y_{k-1}}{\|\delta_{k-1}\|^2} \le \frac{\|y_{k-1}\|}{\|\delta_{k-1}\|} \le \frac{\|y_{k-1}\|^2}{\delta_{k-1}^T y_{k-1}}.$$

This motivates us to guess that the suitable Lipschitz constant should be chosen in the interval

$$\left[\frac{\delta_{k-1}^T y_{k-1}}{\|\delta_{k-1}\|^2}, \frac{\|y_{k-1}\|^2}{\delta_{k-1}^T y_{k-1}}\right].$$

It can be seen from Table 1 that HS methods with the new line search are superior to HS and PRP+ conjugate gradient methods. Moreover, the HS method may fail in some cases if we choose inadequate parameters. Although the PRP+ conjugate gradient method has global convergence, its numerical performance is not better than that of the HS method in many situations.

Numerical experiments show that the new line search proposed in this paper is effective for the HS method in practical computation. The reason is that we used Lipschitz constant estimation in the new line search and could define an adequate initial step size s_k so as to seek a suitable step size α_k for the HS method, which reduced the function evaluations at each iteration and improved the efficiency of the HS method.

It is possible that the initial choice of step size (38) is reasonable for the SCG method in practical computation.All the facts show that choosing an adequate initial step size at each iteration is very important for line search methods, especially for conjugate gradient methods.

7. Conclusion

In this paper, a new form of Armijo-modified line search has been proposed for guaranteeing the global convergence of the HS conjugate gradient method for minimizing functions that have Lipschitz continuous partial derivatives. It needs one to estimate the local Lipschitz constant of the derivative of objective functions in practical computation. The global convergence and linear convergence rate of the HS method with the new Armijo-modified line search were analyzed under some mild conditions. Numerical results showed that the corresponding HS method with the new Armijo-modified line search was effective and superior to the HS conjugate gradient method with strong Wolfe line search. For further research we should not only find more techniques of estimating parameters but also carry out numerical experiments.

Acknowledgments

We would like to thank the two anonymous referees for carefully reading a draft of this paper and making a number of valuable comments.

References

[1] Armijo, L., Minimization of functions having Lipschits continuous partial derivatives, *Pacific J. Math.*, 16, (1966) 1–3.

[2] Al-Baali, M., Descent property and global convergence of the Fletcher–Reeves method with inexact line search, *IMA J. Numer. Anal.*, 5, 121–124 (1985).

[3] Birgin, E.G., Martinez, J.M., A spectral conjugate gradient method for unconstrained optimization, *Appl. Math. Optim.*, 43, 117–128 (2001).

[4] Chen, X.D., Sun, J., Global convergence of a two-parameter family of conjugate gradient methods without line search, *Journal of Computational and Applied Mathematics*, 146, 37–45 (2002).

[5] Dai, Y.H., Yuan, Y., Convergence properties of the conjugate descent method, *Advances in Mathematics*, 25, 552–562 (1996).

[6] Dai, Y.H., Yuan, Y., A nonlinear conjugate gradient method with a strong global convergence property, *SIAM Journal of Optimization*, 10, 177–182 (1999).

[7] Dai, Y.H., Han, J.Y., Liu, G.H., Sun, D.F., Yin, H.X., Yuan, Y.X., Convergence properties of nonlinear conjugate gradient methods, *SIAM Journal of Optimization*, 10, 345–358 (1999).

[8] Fletcher, R., *Practical Method of Optimization*, second ed., vol. I: Unconstrained Optimization, Wiley, New York, (1997).

[9] Fletcher, R., Reeves, C., Function minimization by conjugate gradients, *Comput. J.*, 7, 149–154 (1964).

[10] Fasano, G., Planar conjugate gradient algorithm for large scale unconstrained optimization. Part 1: Theory, *Journal of Optimization Theory and Applications*, 125, 523–541 (2005).

[11] Fasano, G., Planar conjugate gradient algorithm for large-scale unconstrained optimization. Part 1: Applications, *Journal of Optimization Theory and Applications*, 125, 543–558 (2005).

[12] Grippo, L., Lucidi, S., A globally convergent version of the Polak–Ribière conjugate gradient method, *Mathematical Programming*, 78, 375–391 (1997).

[13] Grippo, L., Lucidi, S., Convergence conditions, line search algorithms and trust region implementations for the Polak Ribière conjugate gradient method, *Optimization Methods and Software*, 20(1), 71–98 (2005).

[14] Goldstein, A.A., On steepest descent, *SIAM J. Control*, 3, 147–151 (1965).

[15] Hestenes, M.R., and Stiefel, E.L., Methods of conjugate gradients for solving linear systems, *J. Res. Nat. Bur. Standars Sect.*, 5(49), 409-436 (1952).

[16] Hu, Y.F., Storey, C., Global convergence result for conjugate gradient methods, *Journal of Optimization Theory and Applications*, 71, 399–405 (1991).

[17] Liu, Y., Storey, C., Efficient generalized conjugate gradient algorithms. Part 1: Theory, *Journal of Optimization Theory and Applications*, 69, 129–137 (1991).

[18] Moré, J.J., Garbow, B.S., Hillstrom, K.E., Testing unconstrained optimization software, *ACM Trans. Math. Softw.*, 7, 17–41 (1981).

[19] Polak, E., Ribière, G., Note Sur la convergence de directions conjugèes, Rev. Francaise Informat Recherche Operationelle, 3e Annèe16, 35–43 (1969).

[20] Polyak, B.T., The conjugate gradient method in extreme problems, *USSR Comp. Math. Math. Phys.*, 9, 94–112 (1969).

[21] Powell, M.J.D., Nonconvex minimization calculations and the conjugate gradient method, Lecture Notes in Mathematics, 1066, Springer-Verlag, Berlin, 122–141 (1984).

[22] Sun, J., Zhang, J.P., Convergence of conjugate gradient methods without line search, *Annals of Operations Research*, 103, 161–173 (2001).

[23] Sun, J., Yang, X.Q., Chen, X.D., Quadratic cost flow and the conjugate gradient method, *European Journal of Operational Research*, 164, 104–114 (2005).

[24] Shi, Z.J., Shen, J., Convergence of descent method without line search, *Appl. Math. Comput.*, 167, 94–107 (2005).

[25] Shi, Z.J., Restricted PR conjugate gradient method and its global convergence, *Advances*

in Mathematics, 31(1), 47–55 (2002)(in Chinese).

[26] Shi, Z.J., Shen, J., Convergence of Polak Rebière Polyak conjugate gradient method, *Nonlinear Analysis*, 66, 1428–1441 (2007).

[27] Shi, Z.J., Shen, J., Convergence of Liu-Storey conjugate gradient method, *European Journal of Operational Research*,182(2), 552–560 (2007).

[28] Wolfe, P. Convergence conditions for ascent methods, *SIAM Rev.*, 11, 226–235 (1969).

[29] Yuan, Y. *Numerical Methods for Nonlinear Programming*, Shanghai Scientific & Technical Publishers, (1993).

[30] Yuan, Y. Analysis on the conjugate gradient method, *Optimization Methods and Software*, 2, 19–29 (1993).

[31] Zoutendijk, G. Nonlinear programming computational methods, in: J. Abadie (Ed.), Integer and Nonlinear Programming, North-Holland, Amsterdam, 37–86 (1970).

Mohammed Belloufi *is a Professor in Department of Mathematics, Mohamed chérif Messaadi University Souk Ahras, Algeria. His research interests include unconstrained optimization.*

Rachid Benzine *is a Professor in Department of Mathematics, Badji Mokhtar University Annaba , Algeria. His research interests include unconstrained optimization.*

Yamina Laskri *is a Professor in Department of Mathematics, Badji Mokhtar University Annaba , Algeria. His research interests include unconstrained optimization.*