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

Mohammed Belloufi

Laboratory Informatics and Mathematics (LiM)
Faculty of Science and Technology
Department of Mathematics and Informatics
Mohamed Cherif Messaadia University, Souk Ahras, Algeria.
Email: mbelloufi_doc@yahoo.fr

Rachid Benzine

LANOS
Badji Mokhtar University
Annaba, Algeria
Email: rachid.benzine@univ-annaba.org

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.

Mathematics Subject Classification: 90C06; 90C30; 65K05

Keywords: Unconstrained optimization, Conjugate gradient method, Armijo line search, Global convergence.

1   Introduction

Consider the unconstrained optimization problem

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

where $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is continuously differentiable. The line search method usually takes the following iterative formula

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

for (1.1), where x_k is the current iterate point, $\alpha_k > 0$ is a steplength and d_k is a search direction. Different choices of d_k and α_k will determine different line search methods([23,25,26]). We denote $f(x_k)$ by f_k ,$\nabla f(x_k)$ by g_k, and $\nabla f(x_{k+1})$ by g_{k+1}, respectively. $\|.\|$ denotes the Euclidian norm of vectors and define $y_k = g_{k+1} - g_k$.

We all know that a method is called steepest descent method if we take $d_k = -g_k$ as a search direction at every iteration, which has wide applications in solving large-scale minimization problems ([23, 24,28]). One drawback of the method is often yielding zigzag phenomena in solving practical problems, which makes the algorithm converge to an optimal solution very slowly, or even fail to converge ([16, 18]).

If we take $d_k = -H_k g_k$ as a search direction at each iteration in the algorithm, where H_k is an $n \times n$ matrix approximating $[\nabla^2 f(x_k)]^{-1}$, then the corresponding method is called the Newton-like method ([16, 18, 28]) such as the Newton method, the quasi-Newton method, variable metric method, etc. Many papers have proposed this method for optimization problems ([4, 5, 8, 19]).

However, the Newton-like method needs to store and compute matrix H_k at each iteration and thus adds to the cost of storage and computation. Accordingly, this method is not suitable to solve large-scale optimization problems in many cases.

Due to its simplicity and its very low memory requirement, the conjugate gradient method is a powerful line search method for solving the large-scale optimization problems. In fact, the CG method is not among the fastest or most robust optimization algorithms for nonlinear problems available today, but it remains very popular for engineers and mathematicians who are interested in solving large problems (cf. [1,15,17,27]). The conjugate gradient method is designed to solve unconstrained optimization problem (1). More explicitly, the conjugate gradient method is an algorithm for finding the nearest local minimum of a function of variables which presupposes that the gradient of the function can be computed. We consider only the case where the method is implemented without regular restarts. The iterative formula of the conjugate gradient method is given by (2), where α_k is a steplength which is computed by carrying out a line search, and d_k is the search direction defined by

$$d_{k+1} = \begin{cases} -g_k & \text{si } k = 1 \\ -g_{k+1} + \beta_k d_k & \text{si } k \geq 2 \end{cases} \tag{3}$$

where β_k is a scalar and $g(x)$ denotes $\nabla f(x)$. If f is a strictly convex

quadratic function, namely,

$$f(x) = \frac{1}{2}x^T H x + b^T x, \tag{4}$$

where H is a positive definite matrix and if α_k is the exact one-dimensional minimizer along the direction d_k, i.e.,

$$\alpha_k = \arg\min_{\alpha > 0} \{f(x + \alpha d_k)\} \tag{5}$$

then (1.2)–(1.3) is called the linear conjugate gradient method. Otherwise, (1.2)–(1.3) is called the nonlinear conjugate gradient method.

Conjugate gradient methods differ in their way of defining the scalar parameter β_k. In the literature, there have been proposed several choices for β_k which give rise to distinct conjugate gradient methods. The most well known conjugate gradient methods are the Hestenes–Stiefel (HS) method [11], the Fletcher–Reeves (FR) method [9], the Polak-Ribière-Polyak (PR) method [20,22], the Conjugate Descent method(CD) [8], the Liu-Storey (LS) method [14], the Dai-Yuan (DY) method [6], and Hager and Zhang (HZ) method [12]. The update parameters of these methods are respectively specified as follows:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}, \ \beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \ \beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}, \ \beta_k^{CD} = -\frac{\|g_{k+1}\|^2}{d_k^T g_k} \ ,$$

$$\beta_k^{LS} = -\frac{g_{k+1}^T y_k}{d_k^T g_k}, \ \beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k}, \ \beta_k^{HZ} = \left(y_k - 2d_k \frac{\|y_k\|^2}{d_k^T y_k}\right)^T \frac{g_{k+1}}{d_k^T y_k}$$

The convergence behavior of the above formulas with some line search conditions has been studied by many authors for many years. The FR method with an exact line search was proved to globally convergent on general functions by Zoutendijk [29]. However, the PRP method and the HS method with the exact line search are not globally convergent, see Powell's counterexample [21]. In the already-existing convergence analysis and implementations of the conjugate gradient method, the **Armijo** conditions, namely,

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

$$f_k - f(x_k + \alpha d_k) \geq -\alpha g_k^T d_k. \tag{6}$$

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 addition, the sufficient descent condition:

$$g_k^T d_k \leq -c \|g_k\|^2 \tag{7}$$

has often been used in the literature to analyze the global convergence of conjugate gradient methods with inexact line searches. For instance, Al-Baali [1], Toouati-Ahmed and Storey [3], Hu and Storey [13], Gilbert and Nocedal [10] analyzed the global convergence of algorithms related to the Fletcher–Reeves method with the strong Wolfe line search. Their convergence analysis used the sufficient descent condition (1.8). As for the algorithms related to the PRP method, Gilbert and Nocedal [10] investigated wide choices of β_k that resulted in globally convergent methods. In order for the sufficient descent condition to hold, they modified the strong Wolfe line search to the two-stage line search, the first stage is to find a point using the strong Wolfe line search, and the second stage is when, at that point the sufficient descent condition does not hold, more line search iterations will proceed until a new point satisfying the sufficient descent condition is found. They hinted that the sufficient descent condition may be crucial for conjugate gradient methods.

In this paper we propose a new Armijo line search in which an appropriate initial step size s is defined and varies at each iteration. The new Armijo 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 line search is more effective than other similar methods in solving large scale minimization problems.

2   New Armijo 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

$$\|g(x) - g(y)\| \leq L\|x - y\|, \forall x, y \in \Gamma. \tag{8}$$

New Armijo line search

Given $\mu \in \left]0, \frac{1}{2}\right[$, $\rho \in \left]0, 1\right[$ 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

$$f_k - f(x_k + \alpha d_k) \geq -\alpha\mu g_k^T d_k. \tag{9}$$

$$g(x_k + \alpha d_k)^T d(x_k + \alpha d_k) \leq -c\|g(x_k + \alpha d_k)\|^2 \tag{10}$$

and

$$\|g_{k+1}\| \leq c \|d_k\| \tag{11}$$

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

$$d_k^T g_{k+1} > d_k^T g_k$$

and L_k is an approximation to the Lipschitz constant L of $g(x)$.

3 Algorithm and Convergent properties

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

Now we give the following algorithm firstly.   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 $x_{k+1} = x_k + a_k d_k$ whered_k is defined by (3), $\beta_k = \beta_k^{HS}$ and α_k is

defined by the new Armijo line search.

Step 3. Set$k := k + 1$ and go to Step 1.   Some simple properties of the above algorithm are given as follows.

Assume that (A) and (B) hold and the **HS** method with the new Armijo 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:

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

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

$$-c \left\| g_{k+1} \right\|^2 \geq - \left\| g_{k+1} \right\|^2 + \beta_{k+1}^{HS} \left(g_{k+1}^T d_k \right) = g_{k+1}^T d_{k+1}$$

The proof is finished.   Assume that (A) and (B) hold. Then the new Armijo line search is well defined.

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 2, we can obtain

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

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

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

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

4   Global convergence

Assume that **(A)** and **(B)** hold and the **HS** method with the new Armijo 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,

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

For $k = 0$ we have

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

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

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

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

$$\left\| d_{k+1} \right\| = \left\| -g_{k+1} + \beta_{k+1}^{HS} d_k \right\|$$

$$\leq \left\| g_{k+1} \right\| + \frac{\left| g_{k+1}^T (g_{k+1} - g_k) \right|}{\left| d_k^T (g_{k+1} - g_k) \right|} \left\| d_k \right\|$$

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

$$\leq \left(1 + L \frac{(1-c)}{m_0} \right) \left\| g_{k+1} \right\|$$

The proof is completed.   Assume that **(A)** and **(B)** hold, the **HS** method with the new Armijo 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

$$\lim_{k \to \infty} \|g_k\| = 0. \tag{15}$$

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

$$\lim_{k \in K, k \to \infty} \alpha_k = 0 \tag{16}$$

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

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

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

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

and

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

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

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

Thus

$$g \left(x_k + \alpha d_k\right)^T d_k > \mu g_k^T d_k \qquad (19)$$

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

$$L\alpha \left\|d_k\right\|^2 \geq \left\|g \left(x_k + \alpha \theta_k d_k\right) - g \left(x_k\right)\right\| \left\|d_k\right\|$$

$$\geq \left|\left(g \left(x_k + \alpha \theta_k d_k\right) - g \left(x_k\right)\right)^T d_k\right|$$

$$\geq -(1 - \mu) g_k^T d_k$$

$$\geq c(1 - \mu) \left\|g_k\right\|^2 .$$

We can obtain from Lemma 3 that

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

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

$k \geq k^{'}, k \in K.$

which contradicts (16).

If (18) does not hold, then we have

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

and thus,

$$\frac{g(x_k+\alpha d_k)^T [g(x_k+\alpha d_k)-g_k]}{d_k^T [g(x_k+\alpha d_k)-g_k]} g \left(x_k + \alpha d_k\right)^T d_k$$

$$> (1 - c) \left\|g \left(x_k + \alpha d_k\right)\right\|^2 .$$

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

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

Combining Lemma 3 we have

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

$k \geq k^{'}, k \in K.$

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

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

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

$$m' \|y\|^2 \leq y^T \nabla^2 f(x) y \leq M' \|y\|^2 ,$$

$$\forall x, y \in N(x^*, \varepsilon) \tag{20}$$

$$\tfrac{1}{2} m' \|x - x^*\|^2 \leq f(x) - f(x^*) \leq h \|x - x^*\|^2 ,$$

$$h = \frac{1}{2} M', \forall x \in N(x^*, \varepsilon) \tag{21}$$

$$M' \|x - y\|^2 \geq (g)^T (x - y) \geq m' \|x - y\|^2 ,$$

$$g = g(x) - g(x^*), \forall x, y \in N(x^*, \varepsilon) \tag{22}$$

and thus

$$M' \|x - x^*\|^2 \geq g(x)^T(x) \geq m' \|x - x^*\|^2 ,$$

$$x = (x - x^*), \forall x \in N(x^*, \varepsilon) \tag{23}$$

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

$$M' \|x - x^*\| \geq \|g(x)\| \geq m' \|x - x^*\| ,$$

$$\forall x \in N(x^*, \varepsilon) \tag{24}$$

and

$$\|g(x) - g(y)\| \leq m' \|x - y\| ,$$

$$\forall x, y \in N(x^*, \varepsilon) \tag{25}$$

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

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.

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.

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

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

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

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

For (35) we have

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

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

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

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

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

tableNumber 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 $\|g_k\| \leq 10^{-8}$, 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

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

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.

References

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

[2] N. Andrei, An unconstrained optimization test functions collection, Adv. Modell. Optim. 10 (2008) 147–161.

[3] T. Ahmed, D. Storey, Efficient hybrid conjugate gradient techniques, J. Optimiz. Theory Appl. 64 (1990) 379–394.

[4] Y. Dai, Convergence properties of the BFGS algorithm. SIAM Journal on Optimization, 13 (2003) 693–701.

[5] J. E. Dennis, Jr., J. J. Moré, (1977). Quasi-Newton methods, motivation and theory. SIAM Review, 19, 46–89.

[6] Y. Dai, Y. Yuan, A nonlinear conjugate gradient with a strong global convergence properties, SIAM J. Optimiz. 10 (2000) 177–182.

[7] Y. Dai, Y. Yuan, Convergence properties of the Fletch–Reeves method, IMA J. Numer. Anal. 16 (2) (1996) 155–164.

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

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

[10] J.C. Gibert, J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, SIAM J. Optimiz. 2 (1992) 21–42.

[11] M.R. Hestenes, E. Stiefel, Method of conjugate gradient for solving linear equations, J. Res. Nat. Bur. Stand. 49 (1952) 409–436.

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

[13] Y.F. Hu, C. Storey, Global convergence result for conjugate gradient method, J. Optimiz. Theory Appl. 71 (1991) 399–405.

[14] Y. Liu, C. Storey, Efficient generalized conjugate gradient algorithms. Part 1: Theory, J. Optimiz. Theory Appl. 69 (1992) 129–137.

[15] G. Liu, J. Han, H. Yin, Global convergence of the Fletcher–Reeves algorithm with inexact line search, Appl. Math. JCU 10B (1995) 75–82.

[16] D. C. Luenerger, Linear and nonlinear programming (2nd ed.). Reading, MA: Addition Wesley, 1989.

[17] J. Nocedal, Conjugate gradient methods and nonlinear optimization, in: L. Adams, J.L. Nazareth (Eds.), Linear and Nonlinear Conjugate Gradient Related Methods, SIAM, Philadelphia, PA, 1995, pp. 9–23.

[18] J. Nocedal, S.J. Wright, Numerical optimization, Springer Series in Operations Research, Springer, New York, 1999.

[19] M.J.D. Powell, A new algorithm for unconstrained optimization. In J. B. Rosen, O. L. Mangasarian, & K. Ritter (Eds.), Nonlinear programming. New York: Academic Press, (1970).

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

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

[22] E. Polak, G. Ribière, Note Sur la convergence de directions conjugèes, Rev. Francaise Informat Recherche Operationelle 3e Annèe 16(1969) 35–43.

[23] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. SIAM Journal of Optimization, 7 (1997) 26–33.

[24] M. Raydan, B. F. Svaiter, Relaxed steepest descent and Cauchy–Barzilai–Borwein method. Computational Optimization and Applications, 21 (2002) 155–167.

[25] J. Schropp, A note on minimization problems and multistep methods. Numeric Mathematic, 78 (1997) 87–101.

[26] J. Schropp, One-step and multistep procedures for constrained minimization problems. IMA Journal of Numerical Analysis, 20 (2000) 135–152.

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

[28] Y. Yuan, W. Sun, Theory and Methods of Optimization, Science Press of China, Beijing, 1999.

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