AMO - Advanced Modeling and Optimization, Volume 14, Number 1, 2012

A Note on the Convergence Properties of the Original Three-term Hestenes-Stiefel Method

Li Zhang[1] and Youhua Zhou
Department of Mathematics, Changsha University of Science and Technology,
Changsha 410004, China

Abstract

Recently, Zhang, Zhou and Li [Optim. Methods Softw., 22 (2007), pp. 697-711] proposed a three-term Hestenes-Stiefel (THS) nonlinear conjugate gradient method for optimization and proved its global convergence for strongly convex functions. In this note we further investigate the convergence properties of the THS method on convex optimization. We show that the THS method converges globally for convex functions with the strong Wolfe line search and obtain its R-linear convergence rate under suitable assumptions.

Keywords. The THS method, global convergence, linear convergence rate.

AMS subject classification 2010. 65K05, 90C30.

1 Introduction

In this paper we consider the following unconstrained optimization problem:

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

where $f : R^n \to R$ is a smooth function and its gradient $g(x) \triangleq \nabla f(x)$ is available.

In [6], Zhang et al. proposed a three-term Hestenes-Stiefel (THS) nonlinear conjugate gradient method for nonlinear optimization, that is, the search direction generated by the THS method is defined by

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -g_k + \beta_k^{HS} d_{k-1} - \theta_k y_{k-1}, & \text{if} \quad k \geq 1, \end{cases} \tag{1.2}$$

where

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \quad \theta_k = \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}},$$

$y_{k-1} = g_k - g_{k-1}$ and $g_k = g(x_k)$. The THS method was proved to be globally convergent for strongly convex functions [6]. To ensure global convergence of the THS method for

[1]This work was supported by the NSF foundation (10901026) of China and The Open Fund Project of Key Research Institute of Philosophies and Social Sciences in Hunan Universities.

general functions, based on the MBFGS method and CBFGS method [1, 2], two variants of the THS method were proposed [6]. One is the modified THS method whose search direction is given by

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -g_k + \beta_k^{MHS} d_{k-1} - \theta_k^M \gamma_{k-1}, & \text{if} \quad k \geq 1, \end{cases}$$

where

$$\beta_k^{MHS} = \frac{g_k^T \gamma_{k-1}}{d_{k-1}^T \gamma_{k-1}}, \quad \theta_k^M = \frac{g_k^T d_{k-1}}{d_{k-1}^T \gamma_{k-1}},$$

$$\gamma_{k-1} = y_{k-1} + t_k s_{k-1}, \quad t_k = \max\left\{0, -\frac{y_{k-1}^T s_{k-1}}{\|s_{k-1}\|^2}\right\} + \mu$$

with a given constant $\mu > 0$. Another is the cautious THS method which generates the search direction

$$d_k = \begin{cases} -g_k, & \text{if} \quad s_{k-1}^T y_{k-1} < \epsilon_1 \|g_{k-1}\| s_{k-1}\|^2, \\ -g_k + \beta_k^{HS} d_{k-1} - \theta_k y_{k-1}, & \text{otherwise}, \end{cases}$$

where ϵ_1 is a given positive constant and $s_{k-1} = x_k - x_{k-1}$. These two methods converge globally even for nonconvex optimization [6]. Extensive numerical results in [6] show that these three methods perform very well especially for large-scale problems.

The purpose of this paper is to further study the convergence properties of the original (unmodified) THS method (1.2) on convex optimization. In fact, in the next section we show its global convergence for convex functions and obtain its R-linear convergence rate for strongly convex functions under some conditions.

2 Convergence properties

In this paper we consider the following iterative process

$$x_{k+1} = x_k + \alpha_k d_k, \; k = 0, 1, \ldots, \tag{2.1}$$

where d_k is determined by the THS method (1.2) and the stepsize α_k is computed by the strong Wolfe line search

$$\begin{cases} f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k, \\ |g(x_k + \alpha_k d_k)^T d_k| \leq -\sigma g_k^T d_k, \end{cases} \tag{2.2}$$

where $0 < \delta < \sigma < \frac{1}{2}$. An important property of (1.2) is that the direction d_k satisfies

$$g_k^T d_k = -\|g_k\|^2, \tag{2.3}$$

which is independent of line search used and the convexity of the objective function [6]. In this section, we make the following assumptions for global convergence analysis.

Assumption 2.1

(i) The level set $\Omega_0 = \{x \in R^n | \; f(x) \leq f(x_0)\}$ is bounded;

(ii) $f \in C^2$ and f is a convex function.

Assumption 2.1 implies that the gradient g satisfies Lipschitz condition, that is, there exist an neighborhood Ω of Ω_0 and a positive constant L such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \ \forall x, y \in \Omega. \tag{2.4}$$

Now we present some useful lemmas for global convergence of the THS method.

Lemma 2.1. [3, Lemma 3.4] *Let Assumption 2.1 hold and the sequence $\{x_k\}$ be generated by (2.1), (1.2) and (2.2). Then there exists a positive constant $M_1 > 0$ such that*

$$\frac{\|y_k\|^2}{s_k^T y_k} \leq M_1. \tag{2.5}$$

Lemma 2.2. *Let Assumption 2.1 hold and the sequence $\{x_k\}$ be generated by (2.1), (1.2) and (2.2). Then there is a positive constant m_1 such that*

$$\sum_{k=0}^{\infty} -\alpha_k g_k^T d_k < \infty, \quad \alpha_k \geq m_1 \frac{-g_k^T d_k}{\|d_k\|^2} = m_1 \frac{\|g_k\|^2}{\|d_k\|^2}. \tag{2.6}$$

Proof. The first inequality of (2.6) follows from the first condition in (2.2). From the second condition in (2.2), (2.4) and (2.3), we have

$$L\alpha_k\|d_k\|^2 \geq d_k^T y_k \geq -(1-\sigma) g_k^T d_k = (1-\sigma)\|g_k\|^2,$$

which implies that the second inequality of (2.6) holds with $m_1 = \frac{1-\sigma}{L}$. $\square$

It is clear that (2.6) and (2.3) imply

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \tag{2.7}$$

The following result shows that the THS method converges globally.

Theorem 2.1. *Let Assumption 2.1 hold and the sequence $\{x_k\}$ be generated by (2.1), (1.2) and (2.2). Then*

$$\liminf_{k \to \infty} \|g_k\| = 0. \tag{2.8}$$

Proof. From the second inequality in (2.2) and the fact $\alpha_k > 0$, we have

$$s_k^T y_k \leq -(1+\sigma)\alpha_k g_k^T d_k,$$

which together with the first inequality of (2.6) means that

$$s_k^T y_k \to 0.$$

This and (2.5) yield

$$\|y_k\| \to 0. \tag{2.9}$$

Now we assume the inequality (2.8) is not true. Then there exists a constant $\epsilon > 0$ such that

$$\|g_k\| \geq \epsilon, \ \forall k \geq 0. \tag{2.10}$$

Moreover, the second inequality in (2.2) and (2.3) imply

$$d_{k-1}^T y_{k-1} \geq -(1-\sigma) g_{k-1}^T d_{k-1} = (1-\sigma)\|g_{k-1}\|^2. \tag{2.11}$$

From the definition of the search direction (1.2), the relation (2.3) and (2.11), we have

$$\begin{aligned}
\|d_k\|^2 &= -\|g_k\|^2 - 2g_k^T d_k + \|\beta_k^{HS} d_{k-1} - \theta_k y_{k-1}\|^2 \\
&= \|g_k\|^2 + \|\beta_k^{HS} d_{k-1} - \theta_k y_{k-1}\|^2 \\
&\leq \|g_k\|^2 + 4\|g_k\|^2 \|y_{k-1}\|^2 \frac{\|d_{k-1}\|^2}{(d_{k-1}^T y_{k-1})^2} \\
&\leq \|g_k\|^2 + \frac{4}{(1-\sigma)^2} \|g_k\|^2 \|y_{k-1}\|^2 \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4}.
\end{aligned}$$

This together with (2.9) and (2.10) implies that

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leq \frac{1}{\|g_k\|^2} + \frac{4\|y_{k-1}\|^2}{(1-\sigma)\|g_k\|^2} \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} \leq M_1 + \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} \leq M_1 k + M_0$$

holds for two positive constant M_0 and M_1, which contradicts to (2.7). This finishes the proof. □

Now we begin to prove the R-linear convergence of the THS method. We present the following assumption.

Assumption 2.2 $f \in C^2$ and f is strongly convex, namely, there are two positive constants m and M such that

$$m\|d\|^2 \leq d^T \nabla^2 f(x) d \leq M\|d\|^2, \ \forall d \in R^n, x \in \Omega. \tag{2.12}$$

Under Assumption 2.2, the sequence $\{x_k\}$ converges to the unique minimizer x^* of the problem (1.1). Moreover, Assumption 2.2 yields

$$\frac{1}{2} m\|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{1}{m}\|g(x)\|^2. \tag{2.13}$$

Theorem 2.2. *Let Assumption 2.2 hold and the sequence $\{x_k\}$ be generated by (2.1), (1.2) and (2.2). Then $\{x_k\}$ is R-linearly convergent in the sense that*

$$f(x_{k+1}) - f(x^*) \leq r^k \big(f(x_0) - f(x^*)\big), \ \|x_{k+1} - x^*\| \leq \sqrt{r}^k \big(\frac{2}{m}(f(x_0) - f(x^*))\big)^{\frac{1}{2}},$$

where $0 < r < 1$ is a constant.

Proof. Assumption 2.2 implies that

$$d_{k-1}^T y_{k-1} \geq m\alpha_{k-1}\|d_{k-1}\|^2.$$

This together with (1.2) and (2.4) means

$$\|d_k\| \leq \|g_k\| + \frac{2\|g_k\| L\alpha_{k-1}\|d_{k-1}\|^2}{d_{k-1}^T y_{k-1}} \leq \|g_k\|\big(1 + \frac{2L}{m}\big). \tag{2.14}$$

From (2.6) and (2.14), we obtain

$$\delta\alpha_k \geq \frac{\delta m m_1}{m+2L} \triangleq m_2. \tag{2.15}$$

From the line search (2.2), (2.6) and (2.13), we have

$$f(x_k) - f(x_{k+1}) \geq \delta\alpha_k \|g_k\|^2 \geq m_2\|g_k\|^2 \geq m_2 m\big(f(x_k) - f(x^*)\big),$$

which implies that

$$f(x_{k+1}) - f(x^*) \leq (1 - m_2 m)\big(f(x_k) - f(x^*)\big) \leq r^k\big(f(x_0) - f(x^*)\big), \tag{2.16}$$

where $0 < r \triangleq 1 - m_2 m < 1$. Then from the above inequality and (2.13), we have

$$\|x_{k+1} - x^*\| \leq \sqrt{r}^k \big(\frac{2}{m}(f(x_0) - f(x^*))\big)^{\frac{1}{2}}.$$

The proof is then completed. □

References

[1] D. Li and M. Fukushima, A modified BFGS method and its global convergence in nonconvex minimization, Journal of Computational and Applied Mathematics, 129 (2001), pp. 15-35.

[2] D. Li and M. Fukushima, On the global convergence of the BFGS method for nonconvex unconstrained optimization problems, SIAM Journal on Optimization, 11 (2001), pp. 1054-1064.

[3] G. H. Liu, L. L. Jing, L. X. Han and D. Han, A class of nonmonotone conjugate gradient methods for unconstrained optimization, Journal of Optimization Theory and Applications, 101 (1999), pp. 127-140.

[4] L. Zhang, W. Zhou and D. Li, A descent modified Polak-Ribiére-Polyak conjugate gradient method and its global convergence, IMA Journal of Numerical Analysis, 26 (2006), pp. 629-640.

[5] L. Zhang, W. Zhou and D. Li, Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search, Numerische Mathematik, 104 (2006), pp. 561-572.

[6] L. Zhang, W. Zhou and D. Li, Some descent three-term conjugate gradient methods and their global convergence, Optimization Methods and Software, 22 (2007), pp. 697-711.