RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2016.42.146

ScienceAsia 42 (2016): **146–149**

Global convergence of the original Liu-Storey conjugate gradient method

Yao Ding

College of General Education, Chongqing College of Electronic Engineering, Chongqing 401331 China

e-mail: yaoding_math@126.com

Received 28 Dec 2015
Accepted 23 Mar 2016

ABSTRACT: The original Liu-Storey (LS) method is one of the most effective nonlinear conjugate gradient methods for solving unconstrained optimization problems. Its global convergence is only obtained by using some descent backtracking line searches, which can force the original LS method to generate the sufficient descent direction at each iteration. In this paper, we prove the global convergence of the original LS method with a non-declining backtracking line search.

KEYWORDS: unconstrained optimization, nonlinear conjugate gradient method, non-declining backtracking line search

MSC2010: 90C30

INTRODUCTION

In this paper, we consider the global convergence of the original Liu-Storey (LS) conjugate gradient method[1] for unconstrained optimization problem

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

where $f:\mathbb{R}^n\to\mathbb{R}$ is a continuously differentiable function, and its gradient g is available. For solving problem (1), starting from an initial guess $x_0\in\mathbb{R}^n$, the original LS method generates a sequence $\{x_k\}$ satisfying

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

where the step-size $\alpha_k>0$ is obtained by using some line searches, and the search direction d_k is generated using

$$d_k=\begin{cases}-g_k, & k=0,\\ -g_k+\beta_k d_{k-1}, & k\geqslant 1.\end{cases} \tag{3}$$

In (3), β_k is known as the conjugate gradient parameter and defined as

$$\beta_k=\beta_k^{\mathrm{LS}}=-\frac{g_k^{\mathrm{T}}y_{k-1}}{d_{k-1}^{\mathrm{T}}g_{k-1}}, \tag{4}$$

where $g_k=\nabla f(x_k)$ and $y_{k-1}=g_k-g_{k-1}$.

It is well known that the PRP method[2,3] has a built-in restart feature that avoids the jamming problem: when the step-size α_{k-1} is sufficiently small, the factor y_{k-1} in the numerator of the parameter β_k^{PRP} tends to zero. Hence the parameter β_k^{PRP} becomes sufficiently small and the next search direction d_k is essentially the steepest descent direction $-g_k$. This may be one reason that the PRP method is one of the most effective conjugate gradient methods for solving unconstrained optimization problems. The parameter β_k^{LS} shares the common numerator $g_k^{\mathrm{T}}y_{k-1}$ with the parameter β_k^{PRP}. This means that the original LS method is able to adjust β_k^{LS} to avoid the jamming automatically. Thus some variants of the original LS method have been widely studied. For example, Zhang[4] proposed a modified LS (MLS) method in which the search direction d_k is defined as

$$d_k=\begin{cases}-g_k, & k=0,\\ -g_k+\beta_k^{\mathrm{MLS}}d_{k-1}+\theta_k(y_{k-1}-s_{k-1}), & k\geqslant 1,\end{cases}$$

where $s_{k-1}=x_k-x_{k-1}$, and the parameters β_k^{MLS} and θ_k are defined as follows:

$$\beta_k^{\mathrm{MLS}}=-\frac{g_k^{\mathrm{T}}y_{k-1}}{g_{k-1}^{\mathrm{T}}d_{k-1}}-t\frac{\|y_{k-1}\|^2 g_k^{\mathrm{T}}d_{k-1}}{(g_{k-1}^{\mathrm{T}}d_{k-1})^2},$$

$$\theta_k=-\frac{g_k^{\mathrm{T}}d_{k-1}}{g_{k-1}^{\mathrm{T}}d_{k-1}},$$

with $t>1$. An attractive property of the MLS method is that it always generates the sufficient descent direction at each iteration, which is independent of any line search. Zhang also proved its global convergence under the strong Wolfe line search. Subsequently, based on the parameter β_k^{MLS}, using the structure of β_k in the CG_DESCENT method[5], Li and Feng[6] presented another sufficient modified LS method. Its search direction d_k is given by (3), and

$$\beta_k = \max\{\beta_k^{\mathrm{MLS}}, \eta_k\},$$

where

$$\eta_k = \frac{-1}{\|d_{k-1}\| \min\{\eta, \|g_{k-1}\|\}},$$

and η is a positive constant. Liu and Li[7] proposed a new hybrid DY-LS conjugate gradient method. A remarkable feature is that its search direction not only satisfies the Dai-Liao conjugate condition but is also the same as the Newton direction. This method enjoys the convergent stability of the DY method[8] and the numerical effectiveness of the LS method. Liu[9] studied a class of conjugate gradient methods based on the original LS method for solving unconstrained optimization problems. In Ref. 9, the search direction d_k is given by (3). Set

$$t_k = \frac{u}{\sigma(1+\mu_k)}, \; \mu_k = \frac{\left|g_k^{\mathrm{T}} g_{k-1}\right|}{\|g_k\|^2}, \quad u \in (0, 0.5).$$

If $|\beta_k| \leqslant t_k \left|\beta_k^{\mathrm{LS}}\right|$, the corresponding method always generates a sufficient descent direction at each iteration and converges globally under the strong Wolfe line search.

Although the variants of the original LS method have attracted much attention, and have been accepted as one of the most useful tools for solving unconstrained optimization problems, there is little about the original LS method in the literature. To prove the global convergence of the original LS method, some line searches have been used which force the original LS method to generate a sufficient descent direction at each iteration. For example, Shi and Shen[10] proved the global convergence of the original LS method under a new Armijo-type line search, i.e., the step-size $\alpha_k = \max\{s_k, s_k\rho, s_k\rho^2, \ldots,\}$ such that

$$f(x_k) - f(x_k + \alpha d_k) \geqslant -\alpha\mu g_k^{\mathrm{T}} d_k,$$
$$g(x_k + \alpha d_k)^{\mathrm{T}} d(x_k + \alpha d_k) \leqslant -c \, \|g(x_k + \alpha d_k)\|^2,$$

where $s_k = -((1-c)/L_k) g_k^{\mathrm{T}} d_k / \|d_k\|^2$, $\mu \in (0, 0.5)$, ρ and $c \in (0, 1)$, and L_k is the approximate Lipschitz constant at the kth iteration.

Most recently, Zhou[11] proved the global convergence of the unmodified PRP method with the approximate function descent backtracking line search. By using the idea of Zhou[11], the purpose of this paper is to prove the global convergence of the original LS method using the following non-declining backtracking line search. The step-size $\alpha_k = \max\{1, \rho, \rho^2, \ldots\}$ satisfies

$$f(x_k + \alpha_k d_k) \leqslant f(x_k) - \sigma\alpha_k^2 \, \|d_k\|^2 + \epsilon_k, \qquad (5)$$

$$\left|g_k^{\mathrm{T}} d_k\right| \geqslant \delta \, \|g_k\|^2, \qquad (6)$$

where $\{\epsilon_k\}$ is a positive sequence satisfying

$$\sum_{k=0}^{\infty} \epsilon_k \leqslant \epsilon < +\infty, \qquad (7)$$

$\rho \in (0, 1)$, σ, δ and ϵ are positive constants.

The proposed line search is similar to the approximate function descent backtracking line search proposed by Zhou[11]. However, from additional condition (6), it is not difficult to show that the proposed line search is a non-declining backtracking line search.

The condition (6) means that $g_{k-1}^{\mathrm{T}} d_{k-1}$ is always non-zero when the optimal solution is not achieved. Thus the parameter β_k^{LS} is defined well.

The remainder of this paper establishes the global convergence of the original LS method by using the proposed line search and some mild assumptions.

GLOBAL CONVERGENCE ANALYSIS

We first describe the steps of the original LS method.

Algorithm 1

Step 1: Choose the initial point $x_0 \in \mathbb{R}^n$. Set σ, δ, ϵ, $\varepsilon > 0$, and $\rho \in (0, 1)$. Set $k = 0$.
Step 2: If $\|g_k\| \leqslant \varepsilon$, stop.
Step 3: Generate the step-size α_k by (5) and (6).
Step 4: Set $x_{k+1} = x_k + \alpha_k d_k$. If $\|g_{k+1}\| \leqslant \varepsilon$, stop.
Step 5: Compute $\beta_{k+1}^{\mathrm{LS}}$, and generate the next search direction d_{k+1} by (3).
Step 6: Set $k := k + 1$. Go to Step 2.

To prove the global convergence of Algorithm 1, the objective function $f(x)$ needs the following assumption. *Assumption H*:

(i) The level set $\Omega = \{x \in \mathbb{R}^n \mid f(x) \leqslant f(x_0) + \epsilon\}$ is bounded.
(ii) In a neighbourhood Λ of Ω, f is continuously differentiable and its gradient g is Lipschitz continuous, i.e., there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leqslant L \, \|x - y\|, \quad \forall x, y \in \Lambda. \quad (8)$$

It follows from Assumption H that there exists a constant $\gamma > 0$, such that

$$\|g(x)\| \leqslant \gamma, \quad \forall x \in \Lambda. \tag{9}$$

Lemma 1 *Suppose that Assumption H holds. Let the sequence $\{x_k\}$ be generated by Algorithm 1. Then*

$$\sum_{k=0}^{\infty} \alpha_k^2 \|d_k\|^2 < \infty. \tag{10}$$

Proof: This follows from (5) and (7) directly. □

From (10), it is not difficult to obtain

$$\alpha_k \|d_k\| = 0. \tag{11}$$

Lemma 2 *Suppose that Assumption H holds. Let the sequence $\{x_k\}$ be generated by Algorithm 1. If there exists a positive constant r such that*

$$\|g_k\| \geqslant r, \quad \forall k \geqslant 0, \tag{12}$$

then there exists a positive constant M such that

$$\|d_k\| \leqslant M, \quad \forall k \geqslant 0. \tag{13}$$

Proof: From (6) and (12),

$$\left|g_k^{\mathrm{T}} d_k\right| \geqslant \delta \|g_k\|^2 \geqslant \delta r^2, \quad \forall k \geqslant 0.$$

This inequality together with (8) and (11) gives

$$\begin{aligned}\left|\beta_k^{\mathrm{LS}}\right| &\leqslant \frac{\|g_k\| \cdot \|g_k - g_{k-1}\|}{\delta r^2} \leqslant \frac{\gamma L \|x_k - x_{k-1}\|}{\delta r^2} \\ &= \frac{\gamma L}{\delta r^2} \|\alpha_k d_k\| \to 0, \end{aligned}\tag{14}$$

where the first inequality follows from the Cauchy-Schwartz inequality, and the second inequality follows from (9). Thus there exists a constant $N_0 \in N^+$ and a constant $p \in (0,1)$ such that

$$\left|\beta_k^{\mathrm{LS}}\right| \leqslant p, \quad \forall k \geqslant N_0. \tag{15}$$

From (14) and (15), $\forall k \geqslant k_0$, it is not difficult to obtain

$$\begin{aligned}\|d_k\| &\leqslant \|g_k\| + \left|\beta_k^{\mathrm{LS}}\right| \cdot \|d_{k-1}\| \leqslant \gamma + p \|d_{k-1}\| \\ &\leqslant \gamma(1 + p + p^2 + \cdots + p^{k-N_0-1}) \\ &\quad + p^{k-N_0} \left\|d_{N_0}\right\| \\ &\leqslant \frac{\gamma}{1-p} + \left\|d_{N_0}\right\| \leqslant M, \end{aligned}\tag{16}$$

where $M = \max\{\|d_0\|, \|d_1\|, \dots, \left\|d_{N_0}\right\|, \gamma/(1-p) + \left\|d_{N_0}\right\|\}$. □

The following proof of the global convergence is similar to Theorem 1 in Ref. 11, and it reappears here only for completeness.

Theorem 1 *Suppose that Assumption H holds. Let the sequence $\{x_k\}$ be generated by Algorithm 1. Then*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{17}$$

Proof: By contradiction, we assume that (17) does not hold. Then (12) holds. This means that (13) is true. From (6), it is easy to obtain

$$\|d_k\| \geqslant \delta \|g_k\|. \tag{18}$$

In the following, we consider two possible cases. Firstly, if

$$\liminf_{k\to\infty} \|d_k\| = 0,$$

it follows from (18) that

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

This contradicts (12). Secondly, if

$$\liminf_{k\to\infty} \|d_k\| > 0,$$

from (11), we have

$$\lim_{k\to\infty} \alpha_k = 0. \tag{19}$$

From the definition of the step-size α_k, $\tilde{\alpha}_k = \alpha_k/\rho$ does not satisfy (5), i.e.,

$$\begin{aligned} f(x_k + \tilde{\alpha}_k d_k) &> f(x_k) - \sigma \tilde{\alpha}_k^2 \|d_k\|^2 + \epsilon_k \\ &\geqslant f(x_k) - \sigma \tilde{\alpha}_k^2 \|d_k\|^2, \end{aligned}$$

which implies that

$$\frac{f(x_k + \tilde{\alpha}_k d_k) - f(x_k)}{\tilde{\alpha}_k} \geqslant -\sigma \tilde{\alpha}_k \|d_k\|^2.$$

For the left part of the above inequality, using the mean-value theorem, there exists a constant $\theta_k \in (0,1)$ such that

$$\begin{aligned} -\sigma \tilde{\alpha}_k \|d_k\|^2 &\leqslant g(x_k + \theta_k \tilde{\alpha}_k d_k)^{\mathrm{T}} d_k \\ &= -g(x_k + \theta_k \tilde{\alpha}_k d_k)^{\mathrm{T}} g_k \\ &\quad + \beta_k^{\mathrm{LS}} g(x_k + \theta_k \tilde{\alpha}_k d_k)^{\mathrm{T}} d_{k-1}. \end{aligned}\tag{20}$$

From (9), (13), and (14), it is not difficult to obtain

$$\liminf_{k\to\infty} \beta_k^{\mathrm{LS}} g(x_k + \theta_k \tilde{\alpha}_k d_k)^{\mathrm{T}} d_{k-1} = 0. \tag{21}$$

Due to the boundedness of the sequence $\{x_k\} \subseteq \Omega$, there exists an accumulation point x^* and an infinite index set K such that

$$\liminf_{k\to\infty} x_k = x^*, \quad k \in K. \tag{22}$$

The sequence $\{g_k\}$ is also bounded, and so there also exists an infinite index $K_1 \subseteq K$ and an accumulation point $\tilde{g}$ such that

$$\liminf_{k\to\infty} g_k = \tilde{g}, \quad k \in K_1. \tag{23}$$

Thus by taking the limit as $k \to \infty$ in both sides of (20) for $k \in K_1$, from (11), (21)–(23) we have

$$-\tilde{g}(x^*)^{\mathrm{T}}\tilde{g}(x^*) \geqslant 0,$$

which implies that $\|\tilde{g}(x^*)\| = 0$. This contradicts (12). Thus

$$\liminf_{k\to\infty} \|g_k\| > 0$$

is not true. □

Acknowledgements: The author wishes to express their heartfelt thanks to the referees and editors for their detailed and helpful suggestions for revising the manuscript.

REFERENCES

1. Liu Y, Storey C (1991) Efficient generalized conjugate gradient algorithms. *J Optim Theor Appl* **69**, 177–82.
2. Polak E, Ribière G (1969) Note sur la convergence de méthodes de directions conjugées. *Rev Fr Informat Rech Opér* **16**, 35–43.
3. Polyak BT (1969) The conjugate gradient method in extreme problems. *USSR Comput Math Math Phys* **9**, 94–112.
4. Zhang L (2009) A new Liu-Storey type nonlinear conjugate gradient method for unconstrained optimization problems. *J Comput Appl Math* **225**, 146–57.
5. Hager WW, Zhang H (2005) A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM J Optim* **16**, 170–92.
6. Li M, Feng HY (2011) A sufficient descent LS conjugate gradient method for unconstrained optimization problems. *Appl Math Comput* **218**, 1577–86.
7. Liu JK, Li SJ (2014) A new hybrid conjugate gradient method for unconstrained optimization. *Appl Math Comput* **245**, 36–43.
8. Dai YH, Yuan YX (1999) A nonlinear conjugate gradient method with a strong global convergence property. *SIAM J Optim* **10**, 177–82.
9. Liu JK (2013) Convergence properties of a class of nonlinear conjugate gradient methods. *Comput Oper Res* **40**, 2656–61.
10. Shi ZJ, Shen J (2007) Convergence of Liu-Storey conjugate gradient method. *Eur J Oper Res* **182**, 552–60.
11. Zhou WJ (2013) A short note on the global convergence of the unmodified PRP method. *Optim Lett* **7**, 1367–72.