**Article ID:** 1000-5641(2010)01-0044-08

# Global convergence of a new conjugate gradient method for modified Liu-Storey formula

CAO Wei, WANG Kai-rong

(*College of Mathematics and Physics, Chongqing University, Chongqing* 400030, *China*)

**Abstract:** In this paper, a modified conjugate gradient formula $\beta_k^{MLS}$ based on the formula of the Liu-Storey(LS) nonlinear conjugate gradient method was proposed. It was proved that under the Wolfe-Powell line search and even under the strong Wolfe-Powell line search, with parameter $\sigma \in \left(0, \frac{1}{2}\right)$, the new method has sufficient descent and global convergence properties. Preliminary numerical results show that the method is very promising.

**Key words:** unconstrained optimization; conjugate gradient method; SWP line search; global convergence

**CLC number:** O221.2 **Document code:** A

## 一种新的修正 Liu-Storey 共轭梯度法的全局收敛性

曹 伟, 王开荣

(重庆大学 数理学院, 重庆 400030)

**摘要**: 在 Liu-Storey(LS)公式的基础上给出了一个修正的共轭梯度公式 $\beta_k^{MLS}$. 证明了该新公式在 Wolfe-Powell 线搜索下, 甚至在强 Wolfe-Powell 线搜索下, 在满足$\sigma \in \left(0, \frac{1}{2}\right)$ 的同时, 新算法具有充分下降性和全局收敛性. 数值结果展现了算法的可行性.

**关键词**: 无约束优化; 共轭梯度法; SWP线搜索; 全局收敛性

## 0 Introduction

In this paper, we consider the unconstrained optimization problem

$$\min_{x \in \mathbf{R}^n} f(x), \ x \in \mathbf{R}^n, \tag{0.1}$$

where $f$ is smooth and its gradient $g$ is available.

收稿日期: 2009-10

第一作者: 曹伟, 男, 研究生, 研究方向为最优化技术及方法. E-mail: caowei220@sina.com.

通信作者: 王开荣, 男, 博士, 副教授, 硕士生导师, 主要从事最优化理论研究.

Conjugate gradient methods are very efficient for solving large-scale unconstrained optimization problems (0.0). The iterates of conjugate gradient methods are obtained by

$$x_{k+1} = x_k + \alpha_k d_k, \tag{0.1}$$

with

$$d_k = \begin{cases} -g_k, & if \quad k = 1, \\ -g_k + \beta_k d_{k-1}, & if \quad k > 1, \end{cases} \tag{0.2}$$

where stepsize $\alpha_k$ is positive, $g_k = \nabla f(x_k)$ and $\beta_k$ is a scalar. In addition, $\alpha_k$ is a step length which is computed by carrying out some line search. There are several line search rules for choosing step-size $\alpha_k$, (see [1]) for example, exact minimization rule, Armijo rule, Goldstein rule, Wolfe rule, etc. In this paper we analyze the general results on convergence of line search methods with the following two line search rules.

The Wolfe-Powell(WWP) line search:

$$f(x_k + \alpha_k d_k) - f(x_k) \leqslant \delta \alpha_k g_k^{\mathrm{T}} d_k, \tag{0.3}$$

$$g(x_k + \alpha_k d_k)^{\mathrm{T}} d_k \geqslant \sigma g_k^{\mathrm{T}} d_k, \tag{0.4}$$

and the strongWolfe-Powell (SWP) line search: (0.3) and

$$\left| g(x_k + \alpha_k d_k)^{\mathrm{T}} d_k \right| \leqslant -\sigma g_k^{\mathrm{T}} d_k, \tag{0.5}$$

where $\delta \in (0,1)$ and $\sigma \in \left(\delta, \frac{1}{2}\right)$.

Since 1952, there have been many formulas for the scalar, for example,

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad \text{(Fletcher-Reeves)} \tag{0.6}$$

$$\beta_k^{PRP} = \frac{g_k^{\mathrm{T}} y_{k-1}}{\|g_{k-1}\|^2}, \quad \text{(Ploak-Ribiere-Polyak)} \tag{0.7}$$

$$\beta_k^{HS} = \frac{g_k^{\mathrm{T}} y_{k-1}}{d_{k-1}^{\mathrm{T}} y_{k-1}}, \quad \text{Hestenes-Stiefel)} \tag{0.8}$$

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^{\mathrm{T}} y_{k-1}}, \quad \text{(Dai-Yuan)} \tag{0.9}$$

$$\beta_k^{LS} = -\frac{g_k^{\mathrm{T}} y_{k-1}}{g_{k-1}^{\mathrm{T}} d_{k-1}}, \quad \text{(Liu-Storey)} \tag{1.1}$$

$$\beta_k^{CD} = -\frac{\|g_k\|^2}{d_{k-1}^{\mathrm{T}} g_{k-1}}, \quad \text{(Fletcher)} \tag{1.1}$$

where $y_{k-1} = g_k - g_{k-1}$ and $\|.\|$ stands for the Euclidean norm, which were called FR, PRP, HS, DY, LS and CD methods, respectively, correlative conjugate gradient methods can be found in [2-8].

The PRP, and HS methods are two well-known conjugate gradient methods in practical computation and studied extensively. Although the LS nonlinear conjugate gradient method has a similar structure as the well-known PRP and HS methods, research about this method is rare.

In this paper, we are concerned with the LS method that we expect the techniques developed for the analysis of the above two methods can apply to.

Now let us simply introduce some existing results on the LS method in recent years. ZHANG[9] proposed a new LS type method, which converged globally for general functions with the Grippo-Lucidi line search, and modified this new LS method such that it was globally convergent for nonconvex minimization if the strong Wolfe line search was used. The method combined the Liu-Storey conjugate gradient formula and a new inexact line search and proved that the new method was globally convergent in [10], and it also proved that the LS method was globally convergent and improved the efficiency of LS method in practical computation in [11].

The remainder of the paper is organized as follows. In Section 1, we describe the algorithm of the modified Liu–Storey method and some assumptions. In Section 2, we exhaustive analyze the global convergence, and the numerical results and comparison are reported in Section 3.

# 1 The modified Liu-Storey method

If exact line search is used, the new method is identical to the LS methods. The new conjugate gradient formula is as follows:

$$\beta_k^{MLS} = \frac{g_k^{\mathrm{T}} \bar{y}_{k-1}}{\mu \left| g_k^{\mathrm{T}} d_{k-1} \right| - g_{k-1}^{\mathrm{T}} d_{k-1}}, \tag{1.2}$$

where $\bar{y}_{k-1} = g_k - \dfrac{\|g_k\|}{\|g_{k-1}\|} g_{k-1}, \mu \in (1, +\infty)$ is a parameter. We call the method (0.1) and (0.2) with $\beta_k = \beta_k^{MLS}$ as the MLS method. Now we present concrete algorithm as follows:

**Algorithm**

Step 0 Give $x_1 \in \mathrm{R^n}, \varepsilon \geqslant 0$. Set $d_1 = -g_1 = -\nabla f(x_1), k := 1$, if $\|g_1\| \leqslant \varepsilon$, then stop.

Step 1 Find $\alpha_k > 0$ satisfying the WWP(SWP) conditions (0.3) and (0.4).

Step 2 Let $x_{k+1} = x_k + \alpha_k d_k$ and $g_{k+1} = g(x_{k+1})$. If $\|g_{k+1}\| \leqslant \varepsilon$, then stop.

Step 3 Compute $\beta_{k+1}$ by the formulae (1.2), then generate $d_{k+1}$ by (0.2).

Step 4 Set $k\!:= k{+}1$, go to Step 1.

**Assumption** In the global convergence analysis of conjugate gradient methods, the following assumption is often needed.

(i) The level set $\Omega = \{x \in \mathbf{R^n} / f(x) \leqslant f(x_1)\}$ is bounded, and $f(\mathrm{x})$ is bounded blow in $\Omega$.

(ii) In some neighborhood $\mathbf{N}$ of $\Omega$, $f(\mathrm{x})$ is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leqslant L \|x - y\|, \forall x, y \in \mathbf{N}. \tag{1.3}$$

It follows directly from Assumption that there exists two positive constants $B$ and $\bar{\gamma}$ such that

$$\|x\| \leqslant B, \|g(x)\| \leqslant \bar{\gamma}, \forall x \in \Omega.$$

# 2 Convergence analysis of the Algorithm

Since the conjugate gradient methods belong to the descent methods for solving unconstrained optimization problems, the new $\beta_k$ should be chosen such that $g_k^{\mathrm{T}} d_k < 0$ if a line search is used. Furthermore, due to the sufficient descent condition

$$g_k^{\mathrm{T}} d_k \leqslant -c \|g_k\|^2 \tag{1.4}$$

is a very nice and important property for conjugate gradient methods. The following theorem shows that Algorithm satisfies the sufficient descent condition(1.4).

**Theorem 2.1** Let $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm, there exists a constant $c > 0$ such that the new formula $\beta_k^{MLS}$ satisfies (1.4), $\forall k \geqslant 1$. Meanwhile $\beta_k^{MLS} \geqslant 0, \forall k \leqslant 2$.

**Proof** If $k$=1, (1.4) hold for $c$=1, since $d_1 = -g_1$. Now we suppose that $g_{k-1}^{\mathrm{T}} d_{k-1} \leqslant -c \|g_{k-1}\|^2 < 0$ holds, then we have from (0.2), (0.5), (1.2) and Cauchy-Schwarz inequality that

$$\begin{aligned}
\frac{g_k^{\mathrm{T}} d_k}{\|g_k\|^2} &= -1 + \frac{g_k^{\mathrm{T}} \bar{y}_{k-1}}{\mu \left|g_k^{\mathrm{T}} d_{k-1}\right| - g_{k-1}^{\mathrm{T}} d_{k-1}} \frac{g_k^{\mathrm{T}} d_{k-1}}{\|g_k\|^2} \\
&\leqslant -1 + \frac{\left| \|g_k\|^2 - \dfrac{\|g_k\|}{\|g_{k-1}\|} g_k^{\mathrm{T}} g_{k-1} \right|}{\mu \left|g_k^{\mathrm{T}} d_{k-1}\right| - g_{k-1}^{\mathrm{T}} d_{k-1}} \frac{\left|g_k^{\mathrm{T}} d_{k-1}\right|}{\|g_k\|^2} \\
&\leqslant -1 + \frac{2 \|g_k\|^2}{\mu \left|g_k^{\mathrm{T}} d_{k-1}\right| - g_{k-1}^{\mathrm{T}} d_{k-1}} \frac{\left|g_k^{\mathrm{T}} d_{k-1}\right|}{\|g_k\|^2} \\
&\leqslant -1 + \frac{\left|2 g_k^{\mathrm{T}} d_{k-1}\right|}{-g_{k-1}^{\mathrm{T}} d_{k-1}} \leqslant -1 + \frac{-2\sigma g_{k-1}^{\mathrm{T}} d_{k-1}}{-g_{k-1}^{\mathrm{T}} d_{k-1}} = -(1-2\sigma).
\end{aligned}$$

Now we choose $c = \max\{1, 1-2\sigma\} = 1$, then (1.4) holds for all $k \geqslant 1$. We have from Cauchy-Schwarz inequality and (1.4) that

$$g_k^{\mathrm{T}} \bar{y}_{k-1} \geqslant 0, \quad -g_{k-1}^{\mathrm{T}} d_{k-1} \geqslant 0,$$

so we get $\beta_k^{MLS} \geqslant 0, \forall k \geqslant 2$. This completes the proof.

According to Theorem 2.1, we present the main result of this section.

**Theorem 2.2** Suppose that Assumption holds. Let $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm, then we have

$$\sum_{k=1}^{\infty} \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2} < +\infty. \tag{1.5}$$

**Proof** From Theorem 2.1 we have $g_k^{\mathrm{T}} d_k < 0$ for all $k \geqslant 1$. We also have from (0.5) and (1.3) that

$$-(1-\sigma) g_k^{\mathrm{T}} d_k \leqslant (g_{k+1} - g_k)^{\mathrm{T}} d_k \leqslant L \alpha_k \|d_k\|^2 .$$

Thus

$$\alpha_k \geqslant -\frac{1-\sigma}{L} \frac{g_k^{\mathrm{T}} d_k}{\|d_k\|^2} \ , \tag{1.6}$$

which combining (0.3), we get

$$f(x_k)-f(x_{k+1})\geqslant -\delta\alpha_k g_k^{\mathrm{T}}d_k\geqslant \delta\frac{1-\sigma}{L}\frac{\left(g_k^{\mathrm{T}}d_k\right)^2}{\|d_k\|^2}. \tag{1.7}$$

Further, from Assumption (i) we have $\{f(x_k)\}$ is a decreasing sequence and has a bound below in $\Omega$, and shows $\lim\limits_{k\to\infty}f(x_{k+1})<+\infty$; this together with (1.7) shows

$$+\infty>f(x_1)-\lim_{k\to\infty}f(x_{k+1})=\sum_{k\geqslant 1}\left[f(x_k)-f(x_{k+1})\right]\geqslant\frac{\delta\,(1-\sigma)}{L}\sum_{k=1}^{\infty}\frac{\left(g_k^{\mathrm{T}}d_k\right)^2}{\|d_k\|^2}.$$

We can conclude that (1.5) holds.

Focus on the study of the conjugate gradient methods, Powell[12] suggested that $\beta_k$ should not be less than zero ,which is useful to the PRP method. Under the sufficient descent condition, Gilbert and Nocedal[13] proved that the modified PRP method $\beta_k^{+}=\max\left(0,\beta_k^{PRP}\right)$ was globally convergent with the Wolfe-Powell line search, and who introduced the following property(*) which pertains to the PRP method under the sufficient descent condition. In this paper the $\beta_k^{MLS}$ are always not less than zero. Now before we show that this property(*) pertains to the new method, we will give Property(*) and the following lemmas firstly.

**Property(*)** Consider a method of form (0.1) and (0.2). Suppose that

$$0<\gamma\leqslant\|g_k\|\leqslant\bar{\gamma}. \tag{1.8}$$

We say that the method has Property(*), if for all $k$, there exist constants $b>1,\lambda>0$, such that $\|\beta_k\|\leqslant b$ and if $\|s_{k-1}\|\leqslant\lambda$, we have $|\beta_k|\leqslant\dfrac{1}{2b}$, where $s_{k-1}=x_k-x_{k-1}$.

**Lemma 2.1** Suppose that Assumptions hold. Let $\{d_k\}$ be generated by the new Algorithm. If there exist a constant $\gamma>0$, such that $\|g_k\|\geqslant\gamma$ for all $k$, we have

$$\sum_{k\geqslant 2}\|\mu_k-\mu_{k-1}\|^2<\infty,\quad\text{where}\quad \mu_k=\frac{d_k}{\|d_k\|}.$$

**Lemma 2.2** Suppose that Assumption and (1.4) hold. Let $\{s_k\}$ and $\{d_k\}$ be generated by the new Algorithm. We have $\beta_k^{MLS}$ has Property(*), if there exist a constant $\gamma>0$, such that $\|g_k\|\geqslant\gamma$ for all k, then, for any $\lambda>0$, there exist $\Delta\in\mathrm{Z}^{+}$ and $k_0\in\mathrm{Z}^{+}$, for all $k\geqslant k_0$, such that

$$\left|\kappa_{K,\Delta}^{\lambda}\right|\geqslant\frac{\Delta}{2},$$

where $\kappa_{K,\Delta}^{\lambda}=\{i\in\mathrm{Z}^{+}:k\leqslant i\leqslant k+\Delta-1,\|s_{i-1}\|>\lambda\}$, $\left|\kappa_{K,\Delta}^{\lambda}\right|$ denotes the numbers of the $\kappa_{K,\Delta}^{\lambda}$. If (1.8) holds and the methods have Property(*), then, the small steplength should not be too many. The above lemma shows this property.

**Lemma 2.3** Suppose that Assumption and (1.8) hold. Let $\{x_k\}$ be generated by (0.1) and (0.2), $\alpha_k$ satisfies WWP, and $\beta_k\geqslant 0$ has Property(*). Then, $\lim\limits_{k\to\infty}\inf\|g_k\|=0$.
The proofs of Lemmas 2.1 , 2.2 and Lemmas 2.3 had been given in [14].

The following lemma shows that the new method has Property(*).

**Lemma 2.4** Suppose that Assumption and (1.8) hold. Consider $\{\beta_k^{MLS}\}$ that generated by the new Algorithm. We can get that the new formula has Property(*).

**Proof** Set $q = c$, thus $0 < q \leqslant 1$. Let $b = \dfrac{2\bar{\gamma}^2}{q\gamma^2}, \lambda = \dfrac{q\gamma^4}{8L\bar{\gamma}^3}$ which combining $0 < \gamma \leqslant \bar{\gamma}$ shows $b > 1, \lambda > 0$. We have from (0.5),(1.3),(1.2) and Cauchy-Schwarz inequality that

$$|\beta_k| = \frac{\|g_k\|^2 \left|1 - \dfrac{g_k^{\mathrm{T}} g_{k-1}}{\|g_k\| \|g_{k-1}\|}\right|}{\mu \left|g_k^{\mathrm{T}} d_{k-1}\right| - g_{k-1}^{\mathrm{T}} d_{k-1}} \leqslant \frac{2\|g_k\|^2}{q\|g_{k-1}\|^2} \leqslant \frac{2\bar{\gamma}^2}{q\gamma^2} = b,$$

and if $\|s_{k-1}\| \leqslant \lambda$, then

$$\begin{aligned} |\beta_k| &\leqslant \frac{\|g_k\| \left\| g_k - g_{k-1} + g_{k-1} - \dfrac{\|g_k\| g_{k-1}}{\|g_{k-1}\|} \right\|}{q\|g_{k-1}\|^2} \leqslant \frac{\|g_k\|}{q\|g_{k-1}\|^2} \left(\|g_k - g_{k-1}\| + \|g_k - g_{k-1}\|\right) \\ &\leqslant \frac{2\|g_k\| \|g_k - g_{k-1}\|}{q\|g_{k-1}\|^2} \leqslant \frac{2L\|g_k\| \|s_{k-1}\|}{q\|g_{k-1}\|^2} < \frac{2L\lambda\bar{\gamma}}{q\gamma^2} = \frac{1}{2b}. \end{aligned}$$

This completes the proof.

Now, it is sufficient to prove the global convergence result of the algorithm.

**Theorem 2.3** Consider the new algorithm in which $\beta_k = \beta_k^{MLS}$, and $\beta_k \geqslant 0$ for $k \geqslant 2$. Suppose that Assumption hold, then, the method has

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

This theorem is a immediate result of the above three lemmas, so the proof is omitted.

**Remark** By the way, the corresponding conclusion also holds for SWP line search or MSWP (modify strong Wolfe-Powell) line search. The new method has Property(*), and Theorem 2.3 shows that under some assumptions, the new formula with WWP (SWP) is globally convergent.

# 3 Numerical results

In this section, we report the detailed numerical results of a number of problems by Algorithm. In order to weigh the numerical effects of the different methods, we also test the following four CG methods.

PRP$^+$: the PRP formula (0.7) with nonnegative values $\beta_k = \max\left\{0, \beta_k^{PRP}\right\}$ and SWP conditions, where $\delta = 0.01$, $\sigma = 0.1$, the termination condition is $\|g_k\| \leqslant 10^{-5}$ or It-limit >9 999.

HS: the HS formula (0.8) with SWP conditions, where $\delta = 0.01$, $\sigma = 0.1$, the termination condition is $\|g_k\| \leqslant 10^{-5}$ or It-limit>9 999.

LS: the LS formula(1.0) with SWP conditions, where $\delta = 0.01$, $\sigma = 0.1$, the termination condition is $\|g_k\| \leqslant 10^{-5}$ or It-limit >9 999.

MLS: the Algorithm with SWP conditions, where $\delta = 0.01$, $\sigma = 0.1$, $\mu = 2.0$, the termination condition is $\|g_k\| \leqslant 10^{-5}$ or It-limit >9 999.

Where It-limit means the iterative limit.

In this part, the experiments were carried out on some famous test problems which can be obtained in [15]. We use MATLAB 7.0.1 to tested the chosen problems, and run it on a PC with one 1.73 GHz Genuine Intel (R) CPU processor and 1.00 GB RAM memory and the Windows Vista$^{TM}$ Home Basic system. We compare the PRP$^+$, HS and LS methods with MLS as shown in Tab. 1:

Tab. 1 Numerical results

| Problem | Dim | NI/NF/NG PRP$^+$ | HS | LS | MLS($\mu$=2.0) |
|---|---|---|---|---|---|
| BADSCP | 2 | 35/178/158 | 100/413/359 | 96/399/388 | 24/173/159* |
| HELIX | 3 | 65/181/156 | 48/136/114* | 79/215/184 | 50/148/124 |
| MEYER | 3 | 1/1/1 | 1/1/1 | 1/1/1 | 1/1/1 |
| GULF | 3 | 1/2/2 | 1/2/2 | 1/2/2 | 1/2/2 |
| BOX | 3 | 1/1/1 | 1/1/1 | 1/1/1 | 1/1/1 |
| SING | 4 | 67/215/183* | 92/303/259 | 139/433/373 | 114/384/327 |
| WOOD | 4 | 107/317/270 | 295/765/685 | 105/364/298* | 107/291/250 |
| KOWOSB | 4 | 97/266/235 | 92/278/247 | 60/174/154* | 87/255/225 |
| OSB1 | 5 | 1/1/1 | 1/1/1 | 1/1/1 | 1/1/1 |
| BIGGS | 6 | 149/465/411 | 165/484/424 | 143/358/320 | 62/218/187* |
| OSB2 | 11 | 414/1 000/926 | 235/581/522 | 204/498/448* | 260/635/577 |
| WATSON | 5 | 136/382/337 | 122/366/316 | 116/334/266 | 77/248/213* |
|  | 30 | 2 898/9 906/8 779* | 7 369/20 889/18 665 | 4 867/13 902/12 433 | 3 175/11 115/9 820 |
| SINGX | 100 | 102/331/285 | 92/303/259 | 88/287/246 | 62/204/174* |
|  | 500 | 120/393/343 | 94/307/262* | 118/395/342 | 96/309/268 |
| PEN2 | 100 | …… | 60/186/152* | …… | 63/187/155* |
|  | 500 | 1/1/1 | 1/1/1 | 1/1/1 | 1/1/1 |
| VAEDIM | 5 | 6/57/38* | 10/65/46 | 6/57/38* | 6/57/38* |
|  | 10 | 7/81/52* | 9/91/62 | 7/81/52* | 7/81/52* |
| TRIG | 100 | 50/104/98 | 51/116/109 | 49/109/100* | 52/118/110 |
|  | 500 | 51/107/101 | 45/90/86 | 47/107/101 | 44/98/92* |
| BV | 100 | 4 449/8 341/8 338 | 4 411/8 281/8 278 | 5 622/107 111/10 704 | 3 525/6 998/6 884* |
|  | 500 | 10/19/18* | 13/25/24 | 13/24/23 | 12/20/19 |
| IE | 100 | 5/11/6* | 6/13/7 | 6/13/7 | 5/11/6* |
|  | 500 | 5/11/6* | 6/13/7 | 6/13/7 | 6/13/7 |
| TRID | 100 | 31/73/67 | 31/73/67 | 31/73/67 | 31/73/67 |
|  | 500 | 31/70/66 | 31/70/65* | 31/70/66 | 31/70/65* |
| BAVD | 5 | 11/51/31* | 12/54/35 | 11/51/31* | 12/53/32 |
|  | 10 | 2/21/14 | 2/21/14 | 2/21/14 | 2/21/14 |

**Note** 'Problem' is the name of the test problem in MATLAB; 'Dim' the dimension of the problem; NI the number of iterations; NF the number of function evaluations; NG the number of gradient evaluations; * denotes that this result is the best one among these four methods. $\cdots\cdots$ means the iteration failed.

From the numerical results, we can show that the MLS method performed better than the other three methods. But for further research, we should study the convergence of the new methods with other line search rules, and more numerical experiments for large practical problems and for the choice of the constant $\mu$ should be done in the future.

## [ References ]


[1] SHI Z J. Convergence of line search methods for unconstrained optimization[J]. Applied Mathematics and Computation, 2004, 157: 393-405.

[2] FLETCHER R. REEVES C. Function minimization by conjugate gradients[J]. Comput J, 1964(7): 149-154.

[3] POLAK E. RIBIERE G. Note sur la convergence de directions conjuguées[J]. Rev Fr Inform Rech Oper, 1969, 16: 35-43.

[4] POLYAK B T. The conjugate gradient method in extreme problems[J]. USSR Comp Math Math Phys, 1969(9): 94-112.

[5] HESTENES M R. STIEFEL E L. Methods of conjugate gradients for solving linear systems[J]. J Res Natl Bur Stand Section B, 1952, 49: 409-432.

[6] DAI Y H, YUAN Y. A nonlinear conjugate gradient method with a strong global convergence property[J]. SIAM J Optimization, 1999(10): 177-182.

[7] LIU Y M, STOREY C S. Efficient generalized conjugate gradient algorithms Part 1: Theory[J]. J Optim Theory Appl, 1991, 69: 129-137.

[8] DAI Y H, YUAN Y. Nonlinear Conjugate Gradient Methods[M]. Shanghai: Shanghai Scientific and Technical Publishers, 1998: 37-48.

[9] ZHANG L. A new Liu-Storey type nonlinear conjugate gradient method for unconstrained optimization problems[J]. Journal of Computational and Applied Mathematics, 2008(10): 1016.

[10] TANG C M, WEI Z X, LI G Y. A new version of the Liu-Storey conjugate gradient method[J]. Applied Mathematics and Computation, 2007, 189: 302-313.

[11] SHI Z J, SHEN J. Convergence of Liu-Storey conjugate gradient method[J]. European Journal of Operational Research, 2007, 182: 552-560.

[12] POWELL M J D. Nonconvex minimization Calculations and the Conjugate Gradient Method[M]//Lecture Notes in Mathematics 1066. Berlin: Springer, 1984: 122-141.

[13] GILBERT J C, NOCEDAL J. Global convergence properties of conjugate gradient methods for optimization[J]. SIAM J Optimizat, 1992, 2(1): 21-42.

[14] FLETCHER R. Practical Method of Optimization[M]. 2nd ed. New York: Wiley, 1987.

[15] MORE J J, GARBOW B S, HILLSTROME K E. Testing unconstrained optimization software[J]. ACM Trans Math Software, 1981(7): 17-41.