DOI: 10.13656/j.cnki.gxkx.2009.02.008
广西科学 Guangxi Sciences 2009, 16(2): 101-104, 108

A New Nonmonotone Line Search Method*
一种新的非单调线搜索方法

WEI Zeng-xin, YANG Zhi-mei, QIN Wei-da
韦增欣,杨志梅,覃炜达

(College of Mathematics and Information Science, Guangxi University, Nanning, Guangxi, 530004, China)
(广西大学数学与信息科学学院,广西南宁 530004)

Abstract A new nonmonotone line search for the WYL conjugate gradient method is presented. The nonmonotone line search can guarantee the global convergence of WYL method. Numerical experiments show that WYL method with the nonmonotone line search is more available than Armijo method.

Key words conjugate gradient method, line search, global convergence

摘要:给出一种新的非单调线搜索方法,并用数值实验来验证其优越性.新方法能够确保 WYL共轭梯度法的全局收敛性,实验效果比 Armijo线搜索更好.

关键词:共轭梯度法 线搜索 全局收敛
中图法分类号: O241.7 文献标识码: A 文章编号: 1005-9164(2009)02-0101-04

We consider the general unconstrained optimization problem

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

where $f: R^n \to R$ is a continuously differentiable nonlinear function whose gradient is denoted by $g(x)$. The iteration of the gradient method is given by

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

In computing the steplength α_k, traditional line searches require the function value to decrease monotonically at every iteration. Namely

$$f(x_{k+1}) < f(x_k). \tag{3}$$

However, the nonmonotone line search does not impose the condition (3). In the already-existing nonmonotone line search methods, the following line searches are often used. The nonmonotone Armijo rule as follows:

For each k, let $m(k)$ satisfy: $m(0) = 0$ and $0 \leqslant m(k) \leqslant \min[m(k-1)+1, M]$ for $k \geqslant 1$, where M is a nonnegative integer.

Let $\alpha_k = \rho^{p_k} a$ and p_k be the smallest nonnegative integer p such that $f(x_k + \rho^p a d_k) \leqslant \max\limits_{0 \leqslant j \leqslant m(k)} [f(x_{k-j})] + \sigma \rho^p a g_k^T d_k$, where $a > 0$, $\sigma \in (0,1)$, and $\rho \in (0,1)$.

Similarly, the nonmonotone Goldstein rule can be defined as follows: $f(x_k + \alpha_k d_k) \leqslant \max\limits_{0 \leqslant j \leqslant m(k)} [f(x_{k-j})] + \delta_1 \alpha_k g_k^T d_k$, $f(x_k + \alpha_k d_k) \geqslant \max\limits_{0 \leqslant j \leqslant m(k)} [f(x_{k-j})] + \delta_2 \alpha_k g_k^T d_k$, where $0 < \delta_1 \leqslant \delta_2 < 1$. The nonmonotone Wolf rule can be described as follows: $f(x_k + \alpha_k d_k) \leqslant \max\limits_{0 \leqslant j \leqslant m(k)} [f(x_{k-j})] + \sigma_1 \alpha_k g_k^T d_k$, $g(x_k + \alpha_k d_k)^T d_k \geqslant \sigma_2 g_k^T d_k$, where $0 < \sigma_1 \leqslant \sigma_2 < 1$. Search direction d_k defined by

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -g_k + \beta_k d_{k-1}, & \text{if } k \geqslant 1, \end{cases} \tag{4}$$

where g_k denotes $g(x_k)$, β_k is computed by some well-known formulas[1-3], such as $\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}$, $\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2}$, $\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})}$.

Among them, PRP conjugate gradient method is regarded as the best one in practical computation. However, PRP conjugate gradient method has not global convergence in some conditions[4-6]. Some modified PRP conjugate gradient methods with global

收稿日期: 2008-11-17
作者简介:韦增欣(1963-),男,教授,主要从事最优化理论与方法的研究。
* Supported by Natural Science Foundation of China(No. 10761001).

convergence were proposed[7-9], such as the WYL conjugate gradient method[9], where the β_k is described as follows:

$$\beta_k^{WYL}=\frac{g_k^T\left(g_k-\frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{\|g_{k-1}\|^2}. \qquad (5)$$

In this paper, a new nonmonotone line search is proposed for the WYL conjugate gradient method. The nonmonotone line search can guarantee the global convergence of WYL method under some mild conditions.

1 New nonmonotone line search

The following two basic assumptions are often used in the studies of the conjugate gradient methods.

(H1) The objective function $f(x)$ is bounded from below on the level set $\Omega=\{x\in R^n: f(x)\leqslant f(x_0)\}$.

(H2) In some neighborhood N of Ω, f is continuously differentiable, and its gradient $g(x)$ is Lipschitz continuous, that is to say, for all $x,y\in N$, there exists a constant $L\geqslant 0$ such that $\|g(x)-g(y)\|\leqslant L\|x-y\|$.

Throughout this paper we suppose that the Lipschitz constant L of $g(x)$ is a known prior or easy to estimate in practical computation. There are some estimations L_k for the Lipschitz constant L[10].

Give $L_0>0$, in the k th iteration the sequence $\{L_k\}$ is taken by

$$L_k=\max\left(L_{k-1},\frac{\|y_{k-1}\|}{\|s_{k-1}\|}\right) k=1,2,\cdots, \qquad (6)$$

or

$$L_k=\max\left(L_{k-1},\frac{s_{k-1}^T y_{k-1}}{\|s_{k-1}\|}\right) k=1,2,\cdots, \qquad (7)$$

where $s_{k-1}=x_k-x_{k-1}$ and $y_{k-1}=g_k-g_{k-1}$.

New nonmonotone line search: given $\delta\in(0,\frac{1}{2})$, $\rho\in(0,1)$ and $c\in(\frac{1}{2},1)$. Set $s_k=\frac{1-c}{2L_k}\frac{\|g_k\|^2}{\|d_k\|^2}$ and α_k is the largest α in $\{s_k, s_k\rho, s_k\rho^2,\cdots\}$ such that $\max_{0\leqslant j\leqslant m(k)}[f(x_{k-j})]-f(x_k+\alpha d_k)\geqslant-\alpha\delta g_k^T d_k$, and satisfy the sufficient descent condition

$$g(x_k+\alpha d_k)^T d(x_k+\alpha d_k)\leqslant -c\|g(x_k+\alpha d_k)\|^2, \qquad (8)$$

where $m(0)=0$, $0\leqslant m(k+1)\leqslant\max(m(k)+1, m)$, for $k\geqslant 1$, and L_k is estimated by formula (6) or formula (7).

Algorithm 1

Step0 Choose $x_0\in R^n$ and set $d_0=-g_0$, $L_0>0$, $k:=0$.

Step1: If $\|g_k\|=0$ then stop, otherwise go to step 2.

Step2 Set $x_{k+1}=x_k+\alpha_k d_k$, where d_k is defined by formula (4), $\beta_k=\beta_k^{WYL}$ and α_k is defined by the new nonmonotone line search.

Step3 Set $k:=k+1$ and go to step 1.

Lemma 1.1 Assume that (H1) and (H2) hold. The sequence $\{x_k\}$ is generated by Algorithm 1, then

$$L_0\leqslant L_k\leqslant \max(L_0, L). \qquad (9)$$

Proof By formula (6) and formula (7), we have $L_k\geqslant L_{k-1}\geqslant\cdots\geqslant L_0$. By (H2), we have $\|y_{k-1}\|=\|g_k-g_{k-1}\|\leqslant\|x_k-x_{k-1}\|L=\|s_{k-1}\|L$, so $\frac{\|y_{k-1}\|}{\|s_{k-1}\|}\leqslant L$. Using Cauthy-Schwartz inequality, we get

$$\frac{s_{k-1}^T y_{k-1}}{\|s_{k-1}\|^2}\leqslant\frac{\|s_{k-1}^T y_{k-1}\|}{\|s_{k-1}\|^2}\leqslant\frac{\|s_{k-1}\|\|y_{k-1}\|}{\|s_{k-1}\|^2}=\frac{\|y_{k-1}\|}{\|s_{k-1}\|}\leqslant L,$$

therefore formula (9) holds.

Lemma 1.2 If (H1) and (H2) hold, then the new nonmonotone line search is suitable for the WYL conjugate gradient method.

Proof On the one hand, since

$$\lim_{\alpha\to 0}\frac{\max_{0\leqslant j\leqslant m(k)}[f_{k-j}]-f(x_k+\alpha d_k)}{\alpha}\geqslant$$

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

there is an α_k' such that $\frac{\max_{0\leqslant j\leqslant m(k)}[f_{k-j}]-f(x_k+\alpha d_k)}{\alpha}\geqslant-\delta g_k^T d_k$, $\forall\alpha\in[0,\alpha_k']$. Thus, let $\alpha_k''=\min(s_k,\alpha_k')$, yields

$$\frac{\max_{0\leqslant j\leqslant m(k)}[f_{k-j}]-f(x_k+\alpha d_k)}{\alpha}\geqslant-\delta g_k^T d_k, \forall\alpha\in[0,\alpha_k'']. \qquad (10)$$

On the other hand, we can obtain

$$g(x_k+\alpha d_k)^T d(x_k+\alpha d_k)\leqslant -c\|g(x_k+\alpha d_k)\|^2. \qquad (11)$$

By $d_{k+1}=-g_{k+1}+\beta_{k+1}d_k$, we get $g_{k+1}^T d_{k+1}=g_{k+1}^T(-g_{k+1}+\beta_{k+1}d_k)=-\|g_{k+1}\|^2+\beta_{k+1}g_{k+1}^T d_k$, by formula (11) we get $\beta_{k+1}g_{k+1}^T d_k\leqslant(1-c)\|g_{k+1}\|^2$, so formula (11) holds if and only if

$$\frac{g_{k+1}^T\left(g_{k+1}-\frac{\|g_{k+1}\|}{\|g_k\|}g_k\right)}{\|g_k\|^2}g_{k+1}^T d_k\leqslant(1-c)\|g_{k+1}\|^2. \qquad (12)$$

When $\alpha < \frac{1-c}{2L}\frac{\|g_k\|^2}{\|d_k\|^2}$, using Cauchy-Schwartz inequality and (H2), we have

$$\frac{g_{k+1}^T\left(g_{k+1}-\frac{\|g_{k+1}\|}{\|g_k\|}g_k\right)}{\|g_k\|^2}g_{k+1}^T d_k \leqslant$$

$$\frac{\|g_{k+1}\|^2\|d_k\|}{\|g_k\|^2}\left(\|g_{k+1}-g_k\|+\left\|g_k-\frac{\|g_{k+1}\|}{\|g_k\|}g_k\right\|\right)\leqslant$$

$$\frac{\|g_{k+1}\|^2\|d_k\|}{\|g_k\|^2}\left(\|g_{k+1}-g_k\|+\big|\|g_k\|-\|g_{k+1}\|\big|\right)\leqslant$$

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

Let $\bar{\alpha}_k=\min\left(\alpha_k',\frac{1-c}{2L}\frac{\|g_k\|^2}{\|d_k\|^2}\right)$, we can prove that the new nonmonotone line search is suitable for the WYL conjugate gradient method when $\alpha\in[0,\bar{\alpha}_k]$.

2 Global convergence

Lemma 2.1 Assume that (H1) and (H2) hold. The sequence $\{x_k\}$ is generated by Algorithm 1, then

$$\|d_k\|\leqslant\left(1+\frac{L(1-C)}{L_0}\right)\|g_k\| \text{ for } \forall k. \quad (13)$$

Proof For $k=0$, we have $\|d_k\|=\|g_k\|\leqslant\left(1+\frac{L(1-c)}{L_0}\right)\|g_k\|$. For $k\geqslant 1$, by Lemma 1.1, we have $\alpha_k\leqslant\frac{1-c}{2L_k}\frac{\|g_k\|^2}{\|d_k\|^2}\leqslant\frac{1-c}{2L_0}\frac{\|g_k\|^2}{\|d_k\|^2}$. By Cauchy-Schwartz inequality and the above inequality, noting the WYL formula and (H2), we have

$$\|d_{k+1}\|=\|-g_{k+1}+\beta_{k+1}^{WYL}d_k\|\|d_k\|\leqslant\|g_{k+1}\|+\frac{\left\|g_{k+1}\left(g_{k+1}-\frac{\|g_{k+1}\|}{\|g_k\|}g_k\right)\right\|}{\|g_k\|^2}\|d_k\|\leqslant\|g_{k+1}\|\left(1+\frac{\left\|g_{k+1}-\frac{\|g_{k+1}\|}{\|g_k\|}g_k\right\|}{\|g_k\|^2}\right)\leqslant\|g_{k+1}\|\left(1+2\alpha_k L\frac{\|d_k\|^2}{\|g_k\|^2}\right)\leqslant\left(1+\frac{L(1-c)}{L_0}\right)\|g_{k+1}\|.$$

Theorem 2.1 Assume that (H1) and (H2) hold. The sequence $\{x_k\}$ is generated by Algorithm 1, then there exists $\eta>0$ such that

$$\max_{0\leqslant j\leqslant m(k)}[f_{k-j}]-f_{k+1}\geqslant\eta\|g_k\|^2. \quad (14)$$

Proof Let $\eta_0=\inf_{\forall k}\{\alpha_k\}$. If $\eta_0>0$, then we have

$$\max_{0\leqslant j\leqslant m(k)}[f_{k-j}]-f_{k+1}\geqslant-\delta\alpha_k g_k^T d_k\geqslant\delta\eta_0 c\|g_k\|^2, \quad (15)$$

by letting $\eta=\delta\eta_0 c$ we can prove formula (14) holds.

For the contrary, assume that $\eta_0=0$, then there exists an infinite subset $K\subseteq\{0,1,2,\cdots\}$ such that

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

By Lemmas 1.1 and 2.1, we have $s_k=\frac{1-c}{2L_k}\frac{\|g_k\|^2}{\|d_k\|^2}\geqslant\frac{1-c}{2\max(L_0,L)}\left(1+\frac{L(1-c)}{L_0}\right)^{-2}>0$. Therefore, there is a k' such that $\frac{\alpha_k}{\rho}\leqslant s_k$, $\forall k\geqslant k'$, and $k\in K$. Let $\alpha=\frac{\alpha_k}{\rho}$, at least one of the following two inequalities

$$\max_{0\leqslant j\leqslant m(k)}[f_{k-j}]-f(x_k+\alpha d_k)\geqslant-\alpha\delta g_k^T d_k, \quad (17)$$

and

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

does not hold for $k\geqslant k'$ and $k\in K$.

If formula (17) does not hold, then we have $f_k-f(x_k+\alpha d_k)\leqslant\max_{0\leqslant j\leqslant m(k)}[f_{k-j}]-f(x_k+\alpha d_k)<-\delta\alpha g_k^T d_k$, where $k\geqslant k'$ and $k\in K$. Using the mean value theorem on the left-hand side of the above inequality, there is $\theta_k\in[0,1]$ such that $-\alpha g(x_k+\theta_k\alpha d_k)^T d_k<-\delta\alpha g_k^T d_k$, thus $g(x_k+\theta_k\alpha d_k)^T d_k>\delta g_k^T d_k$, where $k\geqslant k'$ and $k\in K$. By (H2), Cauthy-Schwartz inequality and the above inequality, we have $\alpha L\|d_k\|^2\geqslant\|g(x_k+\alpha d_k)-g_k\|\cdot\|d_k\|\geqslant(g(x_k+\alpha d_k)-g_k)^T d_k>-(1-\delta)g_k^T d_k\geqslant(1-\delta)c\|g_k\|^2$. By Lemma 2.1 and the above inequality, we have $\alpha_k\geqslant\frac{\rho(1-\delta)c}{L}\frac{\|g_k\|^2}{\|d_k\|^2}\geqslant\frac{\rho(1-\delta)c}{L\left(1+\frac{L(1-c)}{L_0}\right)^2}>0$, where $k\geqslant k'$ and $k\in K$, which contradicts to formula (16).

If formula (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 $\frac{g_{k+1}^T\left(g_{k+1}-\frac{\|g_{k+1}\|}{\|g_k\|}g_k\right)}{\|g_k\|^2}g_{k+1}^T d_k>(1-c)\|g_{k+1}\|^2$. By the above inequality, (H2) and Cauchy-Schwartz inequality, we can deduce $\frac{2\|g_{k+1}\|^2\|d_k\|^2\alpha L}{\|g_k\|^2}>(1-c)\|g_{k+1}\|^2$, so we have $\alpha L\frac{\|d_k\|^2}{\|g_k\|^2}>\frac{1-c}{2}$. By Lemma 2.1, we have

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

where $k\geqslant k'$ and $k\in K$, which also contradicts to formula (16), this shows that $\eta_0>0$ and formula (15) always holds. By letting $\eta=\delta\eta_0 c$, we can obtain formula(14).

Remark Theorem 2.1 shows that if $m = 0$, then the corresponding WYL conjugate gradient method reduces to a monotone descent method. In the sequel, we assume that $m \geqslant 1$.

Lemma 2.2 If the conditions of Theorem 2.1 hold, then

$$\max_{1 \leqslant j \leqslant m} [f(x_{ml+j})] \leqslant \max_{1 \leqslant i \leqslant m} [f(x_{m(l-1)+i})] - Z \min_{1 \leqslant j \leqslant m} \|g_{ml+j-1}\|^2, \quad (19)$$

and

$$\sum_{l=1}^{\infty} \min_{1 \leqslant j \leqslant m} \|g_{ml+j-1}\|^2 < +\infty. \quad (20)$$

Proof By (H1) and Theorem 2.1, it suffices to show that the following inequality holds for $j = 1, 2, \cdots, m$,

$$f(x_{ml+j}) \leqslant \max_{1 \leqslant i \leqslant m} [f(x_{m(l-1)+i})] - Z\|g_{ml+j-1}\|^2. \quad (21)$$

Theorem 2.1 implies that

$$f(x_{ml+j}) \leqslant \max_{1 \leqslant i \leqslant m} [f(x_{ml-i})] - Z\|g_{ml}\|^2, \quad (22)$$

which yields that formula (21) holds for $j = 1$. Suppose that formula (21) holds for any j: $1 \leqslant j \leqslant m - 1$. With the descent property of d_k, this implies that

$$\max_{1 \leqslant i \leqslant j} [f(x_{ml+i})] \leqslant \max_{1 \leqslant i \leqslant m} [f(x_{m(l-1)+i})]. \quad (23)$$

By the induction hypothesis, Theorem 2.1 and formula (23), we obtain

$$f(x_{ml+j+1}) \leqslant \max_{0 \leqslant i \leqslant m} [f(x_{ml+j-i})] - Z\|g_{ml+j}\|^2 \leqslant \max\{\max_{1 \leqslant i \leqslant m} f(x_{m(l-1)+i}), \max_{1 \leqslant i \leqslant j} f(x_{ml+i})\} - Z\|g_{ml+j}\|^2 \leqslant \max_{1 \leqslant i \leqslant m} [f(x_{m(l-1)+i})] - Z\|g_{ml+j}\|^2,$$

thus, formula (21) is also true for $j + 1$. By induction, formula (21) holds for $1 \leqslant j \leqslant m$, this shows that formula (19) holds. Since $f(x)$ is bounded from below by (H1), it follows that $\max_{1 \leqslant i \leqslant m} [f(x_{ml+i})] > -\infty$, by summing formula (19) over l, we can get $\sum_{l=1}^{\infty} \min_{1 \leqslant j \leqslant m} \|g_{ml+j-1}\|^2 < +\infty$. Therefore formula (20) holds.

Theorem 2.2 Assume that (H1) and (H2) hold. The sequence $\{x_k\}$ is generated by Algorithm 1, then $\lim_{k \to +\infty} \|g_k\| = 0$.

Proof By Lemma 2.2 and letting $\|g_{h(l)}\| = \min_{1 \leqslant j \leqslant m} \|g_{ml+j-1}\|$, we have $\sum_{l=1}^{\infty} \|g_{h(l)}\|^2 < +\infty$, which yields

$$\lim_{l \to \infty} \|g_{h(l)}\| = 0. \quad (24)$$

The new nonmonotone line search implies that $g_k^T d_k \leqslant -c\|g_k\|^2$, which results in

$$\|d_k\| \geqslant c\|g_k\|. \quad (25)$$

By using Cauthy-Schwartz inequality, Lemma 1.1 and 2.1, formula (25) and the new nonmonotone line search, we have

$$\|g_{k+1}\| = \|g_{k+1} - g_k + g_k\| \leqslant \|g_k\| + \|g_{k+1} - g_k\| \leqslant \|g_k\| + T_k L\|d_k\| \leqslant \|g_k\| + T_k L\left(1 + \frac{L(1-C)}{L_0}\right)\|g_k\| \leqslant \|g_k\|\left[1 + \frac{1-c}{2L_k} L\left(1 + \frac{L(1-c)}{L_0}\right)\frac{\|g_k\|^2}{\|d_k\|^2}\right] \leqslant \|g_k\|\left[1 + \frac{1-c}{2c^2 L_k} L\left(1 + \frac{L(1-c)}{L_0}\right)\right] \leqslant \|g_k\|\left[1 + \frac{1-c}{2c^2 L_0} L\left(1 + \frac{L(1-c)}{L_0}\right)\right] = \bar{m}\|g_k\|, \quad (26)$$

where $\bar{m} = 1 + \frac{1-c}{2c^2 L_0} L\left(1 + \frac{L(1-c)}{L_0}\right)$. Formula (26) implies that $\|g_{m(l+1)+j}\| \leqslant \bar{m}\|g_{m(l+1)+j-1}\| \leqslant \cdots \leqslant \bar{m}^m \|g_{h(l)}\|$ for $j = 1, 2, \cdots, m$, by formula (24) we obtain $\lim_{k \to +\infty} \|g_k\| = 0$.

3 Numerical experiments

Choose 20 numerical examples (http://www.ici.ro/camo/neculai/ansoft.htm.) to test the WYL conjugate gradient method with the new nonmonotone line search and compare the numerical results with that of the WYL conjugate gradient method with the nonmonotone Armijo line search.

In the new nonmonotone line search, we set $\_ = 0.38$, $d = 0.618$, $L_0 = 1$, $c = 0.618$ and $m = 3$. If L_k is estimated by formula (6) or formula (7) then the corresponding WYL conjugate gradient method is denoted by N-WYL1 or N-WYL2, respectively. WYL denotes the WYL conjugate gradient method with the nonmonotone Armijo line search. We denote the dimension of problems by "dim". The stop criteria is $\|g_k\| \leqslant 10^{-5}$. The numerical results are summarized in table 1.

As you can see in table 1, the new nonmonotone line search is available and efficient for the WYL conjugate gradient method, and the estimation formula (7) is superior to the estimation formula (6).

(下转第 108页 Continue on page 108)

index set K and a point x^* such that $(x^k, d^k, z_k) \to (x^*, 0, 0), k \in K$. So following the proof of Theorem 3.2 in Reference [1], we can conclude that x^* is a KKT point for problem (1).

Assumption 5 (i) $f_i, i \in \{0\} \cup I$, are the third-order continuously differentiable. (ii) The matrices $G_i^k, i \in \{0\} \cup I$, are chosen as $G_i^k = \nabla^2 f_i(x^k), i \in \{0\} \cup I$, if k is sufficiently large, and the parameter sequence $\{\epsilon_k\}$ satisfies $\lim_{k \to \infty} \epsilon_k = 0$. (iii) The sequence $\{G_0^k\}$ of matrices is uniformly positive definite, i.e., there exist two positive constants a and b such that $a\|d\|^2 \leqslant d^T G_0^k d \leqslant b\|d\|^2, \forall d \in R^n, \forall k$.

Using Corollary 1, similar to Theorem 4.2 in Reference [1], we can prove the following result.

Theorem 2 Suppose that Assumptions 2~5 hold. Then Algorithm A is superlinearly convergent, i.e., $\|x^{k+1} - x^*\| = o(\|x^k - x^*\|)$.

References

[1] Jian J B. New sequential quadratically constrained quadratic programming norm-relaxed method of feasible directions [J]. Journal of Optimization Theory and Applications, 2006, 129: 109-130.

[2] Solodov M V. On the sequential quadratically constrained quadratic programming methods [J]. Mathematics of Operations Research, 2004, 29: 64-79.

[3] Fukushima M, Luo Z Q, Tseng P. A sequential quadratically constrained quadratic programming method for differentiable convex minimization [J]. SIAM Journal on Optimization, 2003, 13: 1098-1119.

[4] Anitescu M. A superlinearly convergent sequential quadratically constrained quadratic programming algorithm for degenerate nonlinear programming [J]. SIAM Journal on Optimization, 2002, 12: 949-978.

[5] Liu T W. BFGS method and its applications in solving constrained optimization problems [D]. Changsha: Ph D Thesis, Hunan University, 2007.

[6] Lobo M S, Vandenberghe L, Boyd S, et al. Applications of second-order cone programming [J]. Linear Algebra and Its Appl, 1998, 284: 193-228.

(责任编辑: 尹　闯)

(上接第 104 页　Continue from page 104)

Table 1　Iterations\ function evaluations

Problem	dim	N-WYL1	N-WYL2	WYL
1	2	35\ 70	35\ 70	40\ 84
2	4	37\ 74	37\ 74	40\ 77
3	2	16\ 38	13\ 30	19\ 33
4	2	7\ 20	5\ 14	6\ 21
5	2	16\ 38	11\ 28	F
6	2	25\ 50	23\ 47	F
7	4	31\ 62	30\ 61	31\ 65
8	2	18\ 32	13\ 28	10\ 21
9	4	27\ 54	25\ 50	F
10	6	53\ 108	50\ 103	55\ 100
11	100	285\ 574	217\ 438	292\ 580
12	2	26\ 52	25\ 50	27\ 58
13	4	27\ 54	26\ 52	F
14	100	79\ 161	78\ 158	82\ 166
15	2	70\ 141	69\ 140	65\ 131
16	2	272\ 544	267\ 534	260\ 551
17	20	99\ 201	98\ 196	103\ 220
18	200	629\ 1261	647\ 1295	650\ 1302
19	5	73\ 146	72\ 144	F
20	11	75\ 150	74\ 146	74\ 148

4 Conclusion

In this paper, a nonmonotone line search has been proposed for guaranteeing the global convergence of WYL conjugate gradient method. It needs to estimate the Lipschitz constant but the estimation is easy and available in practical. In particular, if $m = 0$, then the new nonmonotone line search will reduce to a monotone line search and the WYL conjugate gradient method with the monotone line search has also global convergence. The Numerical experiments show that WYL method with the nonmonotone line search is available and efficient in practical computation.

References

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

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

[3] Hestenes M R, Stiefel E. Method of conjugate gradient for solving linear systems [J]. J Res, Nat Bur Stand, 1952, 49: 409-436.

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

[5] Goldstein A A. On steepest descent [J]. SIAM J Control, 1965(3): 147-151.

[6] Wolfe P. Convergence conditions for ascent methods [J]. SIAM Rev, 1969(11): 226-235.

[7] Grippo L, Lucidi S. A globally convergent version of the Polak-Ribiere conjugate gradient method [J]. Math Prog, 1997, 78: 375-391.

[8] SHI Z J. Restricted PR conjugate gradient method and its global convergence [J]. Adv Math, 2002, 31(1): 47-55.

[9] WEI Z X, YAO S W. The convergence properties of some new conjugate gradient methods [J]. Appl Math Comput, 2006, 183: 1341-1350.

[10] SHI Z J, Shen J. Convergence of PRP method with new nonmonotone line search [J]. Appl Math Comput, 2006, 181: 423-431.

(责任编辑: 尹　闯)