Journal of Mathematical Research & Exposition
Jan., 2009, Vol. 29, No. 1, pp. 65–75
DOI:10.3770/j.issn:1000-341X.2009.01.009
Http://jmre.dlut.edu.cn

A Modified Conjugate Gradient Method with Global Convergence Property

HONG Ling, MO Li Liu, WEI Zeng Xin
(College of Mathematics & Information Science, Guangxi University, Guangxi 530004, China)
(E-mail: hongling166@163.com)

Abstract A new conjugate gradient method is proposed in this paper. For any (inexact) line search, our scheme satifies the sufficient descent property. The method is proved to be globally convergent if the restricted Wolfe-Powell line search is used. Preliminary numerical result shows that it is efficient.

Keywords unconstrained optimization; conjugate gradient method; Wolfe-Powell line search; global convergence.

Document code A
MR(2000) Subject Classification 49M37; 65K05; 90C30
Chinese Library Classification O221.2

1. Introduction

We consider the unconstrained nonlinear optimization problem:

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

where $f: R^n \to R$ is a smooth, nonlinear function whose gradient will be denoted by g. Nonlinear conjugate gradient method is one of the effective methods for solving unconstrained nonlinear optimization problem (1.1), its iterative formula is given by

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.2}$$

$$d_k = \begin{cases} -g_k, & k = 1, \\ -g_k + \beta_k d_{k-1}, & k \geq 2, \end{cases} \tag{1.3}$$

where d_k is the search direction, α_k is a step-length which is computed by carrying out a line search. The main step-length rules are as follows:

(1) Amijo rule. Let m_k be the minimum integer $m > 0$ such that $\alpha_k = \beta^{m_k} s$ satisfies

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta\beta^m s g_k^{\mathrm{T}} d_k, \quad \beta, \delta \in (0,1). \tag{1.4}$$

(2) Amijo-Goldstein rule. Find an $\alpha_k > 0$ such that

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta\alpha_k g_k^{\mathrm{T}} d_k, \tag{1.5}$$

Received date: 2006-12-31; **Accepted date**: 2007-11-22
Foundation item: the National Natural Science Foundation of China (No. 10761001); the Natural Science Foundation of Guangxi Province (No. 0542043).

$$f(x_k+\alpha_k d_k)-f(x_k)\geq(1-\delta)\alpha_k g_k^{\mathrm{T}}d_k,\quad \delta\in(0,\frac{1}{2}). \tag{1.6}$$

(3) Weak Wolfe-Powell rule (WWP). Find an $\alpha_k>0$ satisfying (1.5) and

$$g(x_k+\alpha_k d_k)^{\mathrm{T}}d_k\geq\sigma g_k^{\mathrm{T}}d_k,\quad \sigma\in(\delta,1). \tag{1.7}$$

(4) Strong Wolfe-Powell rule (SWP). Find an $\alpha_k>0$ satisfying (1.5) and

$$|g(x_k+\alpha_k d_k)^{\mathrm{T}}d_k|\leq\sigma|g_k^{\mathrm{T}}d_k|,\quad \sigma\in(\delta,1). \tag{1.8}$$

Where β_k is a scalar and g_k denotes $g(x_k)$. There are some famous formulas for β_k such as β_k^{FR}, β_k^{PRP}, β_k^{HS}, β_k^{CD} etc.[1−7].

Recently, the authors in [8] proposed a new quasi-Newton formula $B_{k+1}s_k=y_k^*$, where $y_k^*=y_k+A_k(3)s_k$ and $y_k=g_{k+1}-g_k$, $A_k(3)$ is a positive definite matrix, $s_k=x_{k+1}-x_k$. They suggested that $A_k(3)=\frac{2[f(x_k)-f(x_{k+1})]+(g(x_{k+1})+g(x_k))^{\mathrm{T}}s_k}{\|s_k\|^2}I$, where I denotes the identity matrix. By using the new quasi-Newton equation, good numerical results are obtained.

In [9], the authors proposed a new formula for β_k as follows:

$$\bar{\beta}_k^N=\max\{\beta_k^N,\eta_k\},\quad \beta_k^N=\frac{1}{d_k^{\mathrm{T}}y_k}(y_k-2d_k\frac{\|y_k\|^2}{d_k^T y_k})^T g_{k+1},$$
$$\eta_k=\frac{-1}{\|d_k\|\min\{\eta,\|g_k\|\}}\quad(\eta>0). \tag{1.9}$$

The conjugate gradient method is proved to be globally convergent for general nonlinear function if the weak Wolfe-Powell line search is used, and preliminary numerical result shows that it is efficient.

Enlightened by the above ideas, we propose a new formula for β_k as follows:

$$\hat{\beta}_k^N=\max\{\tilde{\beta}_k^N,\eta_k\},\quad \tilde{\beta}_k^N=\frac{1}{d_k^{\mathrm{T}}y_k^*}(y_k^*-2d_k\frac{\|y_k^*\|^2}{d_k^{\mathrm{T}}y_k^*})^{\mathrm{T}}g_{k+1},$$
$$\eta_k=\frac{-1}{\|d_k\|\min\{\eta,\|g_k\|\}}\quad(\eta>0). \tag{1.10}$$

The new formula is proved to be globally convergent if Wolfe-Powell line search is used.

Our paper is organized as follows. In Section 2, we first present a new algorithm (Algorithm 2.1), the sufficient descent property (2.2) of Algorithm 2.1 is also proved in this section. In Section 3, we establish the global convergence of Algorithm 2.1. The preliminary numerical results are contained in Section 4.

2. Algorithm and its properties

Algorithm 2.1

Step 1. Choose an initial point $x_1\in R^n$, $\varepsilon\geq0$, set $d_1=-g$, $k=1$ if $\|g_1\|\leq\varepsilon$, then stop.

Step 2. Compute $\alpha_k>0$ by some line searches.

Step 3. Let $x_{k+1}=x_k+\alpha_k d_k$ and $g_{k+1}=g(x_{k+1})$. If $\|g_{k+1}\|\leq\varepsilon$, then stop.

Step 4. Compute $\beta_{k+1}=\hat{\beta}_{k+1}^N$ by the formula (1.10) and generate d_{k+1} by (1.3).

Step 5. Set $k := k + 1$, and go to Step 2.

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

Assumption A The level set $\Omega = \{x \in R^n \mid f(x) \leq f(x_1)\}$ is bounded.

Assumption B The function $g(x)$ is Lipschitz continuous in Ω, i.e., there exists a constant $L > 0$ such that $\| g(x) - g(y) \| \leq L \| x - y \|$ for any $x, y \in \Omega$.

An attractive feature of the conjugate gradient scheme which, we now establish, is that the search directions always yield descent when $d_k^{\mathrm{T}} y_k^* \neq 0$.

Property 2.1 *If* $d_k^{\mathrm{T}} y_k^* \neq 0$ *and*

$$d_{k+1} = -g_{k+1} + \tau d_k, \quad d_1 = -g_1, \tag{2.1}$$

for any $\tau \in [\tilde{\beta}_k^N, \max\{\tilde{\beta}_k^N, 0\}]$, *then*

$$g_{k+1}^{\mathrm{T}} d_{k+1} \leq -\frac{7}{8}\|g_{k+1}\|^2. \tag{2.2}$$

Proof Since $d_1 = -g_1$, we have $g_1^{\mathrm{T}} d_1 = -\|g_1\|^2$, which satisfies (2.2).

(1) Suppose $\tau = \tilde{\beta}_k^N \geq 0$. Multiplying (2.1) by g_{k+1}^{T}, we have

$$\begin{aligned} g_{k+1}^{\mathrm{T}} d_{k+1} &= -\|g_{k+1}\|^2 + \tilde{\beta}_k^N g_{k+1}^{\mathrm{T}} d_k \\ &= -\|g_{k+1}\|^2 + g_{k+1}^{\mathrm{T}} d_k \Big(\frac{y_k^{*\mathrm{T}} g_{k+1}}{d_k^{\mathrm{T}} y_k^*} - 2\frac{\|y_k^*\|^2 g_{k+1}^{\mathrm{T}} d_k}{(d_k^{\mathrm{T}} y_k^*)^2}\Big) \\ &= \frac{y_k^{*\mathrm{T}} g_{k+1} (d_k^{\mathrm{T}} y_k^*)(g_{k+1}^{\mathrm{T}} d_k) - \|g_{k+1}\|^2 (d_k^{\mathrm{T}} y_k^*)^2 - 2\|y_k^*\|^2 (g_{k+1}^{\mathrm{T}} d_k)^2}{(d_k^{\mathrm{T}} y_k^*)^2}. \end{aligned} \tag{2.3}$$

In the first term of the numerator of the last equality above, let $u = \frac{1}{2}(d_k^{\mathrm{T}} y_k^*) g_{k+1}$ and $v = 2(g_{k+1}^{\mathrm{T}} d_k) y_k^*$. Since

$$u^{\mathrm{T}} v \leq \frac{1}{2}(\|u\|^2 + \|v\|^2),$$

we have

$$\begin{aligned} y_k^{*\mathrm{T}} g_{k+1} (d_k^{\mathrm{T}} y_k^*)(g_{k+1}^{\mathrm{T}} d_k) &\leq \frac{1}{2}\Big(\|\frac{1}{2}(d_k^{\mathrm{T}} y_k^*) g_{k+1}\|^2 + \|2(g_{k+1}^{\mathrm{T}} d_k) y_k^*\|^2\Big) \\ &= \frac{1}{8}(d_k^{\mathrm{T}} y_k^*)^2 \|g_{k+1}\|^2 + 2(g_{k+1}^{\mathrm{T}} d_k)^2 \|y_k^*\|^2, \end{aligned}$$

which implies $g_{k+1}^{\mathrm{T}} d_{k+1} \leq -\frac{7}{8}\|g_{k+1}\|^2$ by considering the last equality of (2.3).

(2) If $\tau \neq \tilde{\beta}_k^N$, then $\tilde{\beta}_k^N \leq \tau \leq 0$. After multiplying (2.1) by g_{k+1}^{T}, we have

$$g_{k+1}^{\mathrm{T}} d_{k+1} = -\|g_{k+1}\|^2 + \tau g_{k+1}^{\mathrm{T}} d_k.$$

If $g_{k+1}^{\mathrm{T}} d_k \geq 0$, then (2.2) follows immediately since $\tau \leq 0$. If $g_{k+1}^{\mathrm{T}} d_k < 0$, then

$$g_{k+1}^{\mathrm{T}} d_{k+1} = -\|g_{k+1}\|^2 + \tau g_{k+1}^{\mathrm{T}} d_k \leq -\|g_{k+1}\|^2 + \tilde{\beta}_k^N g_{k+1}^{\mathrm{T}} d_k.$$

By $\tilde{\beta}_k^N \leq \tau \leq 0$ and analysis of (1), (2.2) follows from our previous analysis.

Property 2.2 *For any given* k, *if* $(\alpha_k, x_{k+1}, g_{k+1}, d_{k+1})$ *is generated by Algorithm 2.1 and*

$A_k = A_k(3)$, we consider the following restricted Wolfe-Powell step-size (RWP)

$$f(x_k+\alpha_k d_k)-f(x_k)\leq\delta\alpha_k g_k^{\mathrm{T}}d_k, \tag{2.4}$$

$$g_{k+1}^{\mathrm{T}}d_k\geq\sigma g_k^{\mathrm{T}}d_k, \tag{2.5}$$

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

$$\alpha_k\geq\frac{(1-\sigma)}{L}\frac{|g_k^{\mathrm{T}}d_k|}{\|d_k\|^2}; \tag{2.6}$$

$$y_k^{*\mathrm{T}}d_k>0; \tag{2.7}$$

$$\|y_k^*\|\leq 2L\|s_k\|. \tag{2.8}$$

Proof (1) Subtracting $g_k^{\mathrm{T}}d_k$ from both sides of (2.5) using the Lipschitz condition gives

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

Since d_k is a descent direction and $\sigma<\frac{1}{2}$, (2.6) follows immediately.

(2) By the restricted Wolfe condition (2.4)–(2.5), we have

$$\begin{aligned}
y_k^{*\mathrm{T}}d_k &= (y_k+A_k(3)s_k)^{\mathrm{T}}d_k\\
&= (y_k+\frac{2[f(x_k)-f(x_{k+1})]+(g(x_{k+1})+g(x_k))^{\mathrm{T}}s_k}{\|s_k\|^2}s_k)^{\mathrm{T}}d_k\\
&\geq(g_{k+1}^{\mathrm{T}}d_k-g_k^{\mathrm{T}}d_k)+\frac{-2\delta g_k^{\mathrm{T}}s_k+g_{k+1}^{\mathrm{T}}s_k+g_k^{\mathrm{T}}s_k}{\|s_k\|^2}s_k^{\mathrm{T}}d_k\\
&=(g_{k+1}^{\mathrm{T}}d_k-g_k^{\mathrm{T}}d_k)+\frac{-2\delta g_k^{\mathrm{T}}d_k+g_{k+1}^{\mathrm{T}}d_k+g_k^{\mathrm{T}}d_k}{\|s_k\|^2}s_k^{\mathrm{T}}s_k\\
&=(g_{k+1}^{\mathrm{T}}d_k-g_k^{\mathrm{T}}d_k)-2\delta g_k^{\mathrm{T}}d_k+g_{k+1}^{\mathrm{T}}d_k+g_k^{\mathrm{T}}d_k\\
&\geq-2(\delta-\sigma)g_k^{\mathrm{T}}d_k,
\end{aligned} \tag{2.9}$$

where $\sigma\in(0,\delta)$, $s_k=x_{k+1}-x_k=\alpha_k d_k$. From (2.9) and $g_k^{\mathrm{T}}d_k<0$, (2.7) follows immediately.

(3) By using the Mean Value Theorem, we obtain $\theta\in[0,1]$, such that

$$\begin{aligned}
|\,2(f_k-f_{k+1})+(g_{k+1}+g_k)^{\mathrm{T}}s_k\,| &\leq |\,[-2g(x_k+\theta s_k)+g_{k+1}+g_k]^{\mathrm{T}}s_k\,|\\
&\leq\|\,s_k\,\|\,[\|g_{k+1}-g(x_k+\theta s_k)\|+\|g_k-g(x_k+\theta s_k)\|]\\
&\leq\|\,s_k\,\|\,[L(1-\theta)\,\|\,s_k\,\|+L\theta\,\|\,s_k\,\|]=L\,\|\,s_k\,\|^2\,.
\end{aligned}$$

Thus from the definition of y_k^* and Assumption B, we have

$$\|\,y_k^*\,\|=\|\,y_k+A_k(3)s_k\,\|\leq L\,\|\,s_k\,\|+\frac{L\,\|\,s_k\,\|^2}{\|\,s_k\,\|}=2L\,\|\,s_k\,\|\,. \tag{2.10}$$

Hence (2.8) holds.

3. Convergence analysis for general nonlinear functions

In this section, we study the convergence properties of our new method. First we give our Lemma 3.1.

Lemma 3.1 *If assumptions A and B hold, then for scheme (1.3), (1.10) and a line search that satisfies the RWP conditions (2.4)–(2.5), we have*

$$d_k \neq 0 \ \ \text{for each } k \ \ \text{and} \ \ \sum_{k=0}^{\infty} \|u_{k+1} - u_k\|^2 < \infty,$$

where u_k *denotes* $\frac{d_k}{\|d_k\|}$, *whenever* $\inf\{\|g_k\| : k \geq 0\} > 0$.

Proof Define $\gamma = \inf\{\|g_k\| : k \geq 0\}$. Since $\gamma > 0$ by assumption, it follows from the descent property, property 2.1, that $d_k \neq 0$ for any k. Since Ω is bounded, f is bounded from below, and by (2.4) and (2.6), the following Zoutendijk condition[10] holds:

$$\sum_{k=0}^{\infty} \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2} < \infty.$$

Again, the descent property yields

$$\gamma^4 \sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} \leq \sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} \leq \frac{64}{49} \sum_{k=0}^{\infty} \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2} < \infty. \tag{3.1}$$

Define the quantities

$$\beta_k^+ = \max\{\hat{\beta}_k^N, 0\}, \quad \beta_k^- = \min\{\hat{\beta}_k^N, 0\}, \quad r_k = \frac{-g_k + \beta_{k-1}^- d_{k-1}}{\|d_k\|}, \quad \delta_k = \beta_{k-1}^+ \frac{\|d_{k-1}\|}{\|d_k\|}.$$

By (1.3) and (1.10), we have

$$u_k = \frac{d_k}{\|d_k\|} = \frac{-g_k + (\beta_{k-1}^+ + \beta_{k-1}^-) d_{k-1}}{\|d_k\|} = r_k + \delta_k u_{k-1}.$$

Since the u_k are unit vectors, then

$$\|r_k\| = \|u_k - \delta_k u_{k-1}\| = \|\delta_k u_k - u_{k-1}\|.$$

Since $\delta_k > 0$, it follows that

$$\begin{aligned} \|u_k - u_{k-1}\| &\leq \|(1+\delta_k)(u_k - u_{k-1})\| \leq \|u_k - \delta_k u_{k-1}\| + \|\delta_k u_k - u_{k-1}\| \\ &= 2\|r_k\|. \end{aligned} \tag{3.2}$$

By the definition of β_k^- and the fact that $\eta_k < 0$ and $\hat{\beta}_k^N \geq \eta_k$ in (1.10), we have the following bound for the numerator of r_k:

$$\begin{aligned} \| -g_k + \beta_{k-1}^- d_{k-1}\| &\leq \|g_k\| - \min\{\hat{\beta}_{k-1}^N, 0\} \|d_{k-1}\| \\ &\leq \|g_k\| - \eta_{k-1} \|d_{k-1}\| \\ &\leq \|g_k\| + \frac{1}{\|d_{k-1}\| \min\{\eta, \gamma\}} \|d_{k-1}\| \\ &\leq \Gamma + \frac{1}{\min\{\eta, \gamma\}}, \end{aligned} \tag{3.3}$$

where

$$\Gamma = \max_{x \in \Omega} \| \bigtriangledown f(x) \|. \tag{3.4}$$

Let c denote the expression $c = \Gamma + \frac{1}{\min\{\eta,\gamma\}}$ in (3.3). This bound for the numerator of r_k coupled with (3.2) gives

$$\|u_k - u_{k-1}\| \le 2\|r_k\| \le \frac{2c}{\|d_k\|}. \tag{3.5}$$

Finally, by squaring (3.5), summing over k, and utilizing (3.1), we complete the proof. □

Theorem 3.1 *If assumptions A and B hold, then for scheme (1.3), (1.10) and a line search that satisfies the RWP conditions (2.4)–(2.5), either $g_k = 0$ for some k, or*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{3.6}$$

Proof We suppose that both $g_k \ne 0$ for all k and $\liminf_{k\to\infty} \|g_k\| > 0$. In the following, we obtain a contradiction. Defining $\gamma = \inf\{\|g_k\| : k \ge 0\}$, we have $\gamma > 0$ due to (3.6) and the fact that $g_k \ne 0$ for all k. The proof is divided into the following three steps:

(1) A bound for $\hat\beta_k^N$. By the Wolfe condition $f(x_{k+1}) - f(x_k) \le \delta\alpha_k g_k^{\mathrm T} d_k$, $g_{k+1}^{\mathrm T} d_k \ge \sigma g_k^{\mathrm T} d_k$, and Property 2.1, we have $-g_k^{\mathrm T} d_k \ge \frac{7}{8}\|g_k\|^2 \ge \frac{7}{8}\gamma^2$. Combining this with (2.9) gives

$$y_k^{*\mathrm T} d_k \ge \frac{7(\delta-\sigma)}{4}\gamma^2. \tag{3.7}$$

Again, the restricted Wolfe condition (2.4)–(2.5) gives

$$\begin{aligned}
g_{k+1}^{\mathrm T} d_k &= y_k^{*\mathrm T} d_k + g_k^{\mathrm T} d_k - (A_k(3)s_k)^{\mathrm T} d_k \\
&= y_k^{*\mathrm T} d_k + g_k^{\mathrm T} d_k - \Big(\frac{2[f(x_k) - f(x_{k+1})] + (g_{k+1}+g_k)^{\mathrm T} s_k}{\|s_k\|^2} s_k\Big)^{\mathrm T} d_k \\
&\le y_k^{*\mathrm T} d_k + g_k^{\mathrm T} d_k + \Big(\frac{2\delta g_k^{\mathrm T} s_k - (g_{k+1}+g_k)^{\mathrm T} s_k}{\|s_k\|^2} s_k\Big)^{\mathrm T} d_k \\
&= y_k^{*\mathrm T} d_k + 2\delta g_k^{\mathrm T} d_k - g_{k+1}^{\mathrm T} d_k.
\end{aligned}$$

Hence, we have

$$g_{k+1}^{\mathrm T} d_k \le \frac{1}{2}(y_k^{*\mathrm T} d_k + 2\delta g_k^{\mathrm T} d_k) \le \frac{1}{2} y_k^{*\mathrm T} d_k. \tag{3.8}$$

Also, observe that

$$\begin{aligned}
g_k^{\mathrm T} d_k &= g_{k+1}^{\mathrm T} d_k - y_k^{*\mathrm T} d_k + (A_k(3)s_k)^{\mathrm T} d_k \\
&= g_{k+1}^{\mathrm T} d_k - y_k^{*\mathrm T} d_k + \Big(\frac{2[f(x_k) - f(x_{k+1})] + (g_{k+1}+g_k)^{\mathrm T} s_k}{\|s_k\|^2} s_k\Big)^{\mathrm T} d_k \\
&\ge g_{k+1}^{\mathrm T} d_k - y_k^{*\mathrm T} d_k + \Big(\frac{-2\delta g_k^{\mathrm T} s_k + (g_{k+1}+g_k)^{\mathrm T} s_k}{\|s_k\|^2} s_k\Big)^{\mathrm T} d_k \\
&= g_{k+1}^{\mathrm T} d_k - y_k^{*\mathrm T} d_k + (-2\delta+1) g_k^{\mathrm T} d_k + g_{k+1}^{\mathrm T} d_k.
\end{aligned}$$

Hence, we have

$$g_k^{\mathrm T} d_k \ge \frac{1}{\delta} g_{k+1}^{\mathrm T} d_k - \frac{1}{2\delta} y_k^{*\mathrm T} d_k. \tag{3.9}$$

Again, The restricted Wolfe condition (2.5) gives

$$g_{k+1}^{\mathrm T} d_k \ge \sigma g_k^{\mathrm T} d_k \ge \frac{\sigma}{\delta} g_{k+1}^{\mathrm T} d_k - \frac{\sigma}{2\delta} y_k^{*\mathrm T} d_k \Rightarrow \frac{\delta-\sigma}{\delta} g_{k+1}^{\mathrm T} d_k \ge -\frac{\sigma}{2\delta} y_k^{*\mathrm T} d_k.$$

Hence, we have

$$g_{k+1}^{\mathrm{T}} d_k \geq -\frac{\sigma}{2(\delta-\sigma)} y_k^{*\mathrm{T}} d_k.$$

Combining this lower bound for $g_{k+1}^{\mathrm{T}} d_k$ with the upper bound (3.8) yields

$$\left|\frac{g_{k+1}^{\mathrm{T}} d_k}{y_k^{*\mathrm{T}} d_k}\right| \leq \max\{\frac{\sigma}{2(\delta-\sigma)}, \frac{1}{2}\}. \tag{3.10}$$

By the definition of $\hat{\beta}_k^N$ in (1.10), we have

$$\hat{\beta}_k^N = \tilde{\beta}_k^N \text{ if } \tilde{\beta}_k^N \geq 0 \text{ and } 0 \geq \hat{\beta}_k^N \geq \tilde{\beta}_k^N \text{ if } \tilde{\beta}_k^N < 0.$$

Hence, $|\hat{\beta}_k^N| \leq |\tilde{\beta}_k^N|$ for each k. We now insert the upper bound (3.10) for $|g_{k+1}^{\mathrm{T}} d_k|/|y_k^{*\mathrm{T}} d_k|$, the lower bound (3.7) for $y_k^{*\mathrm{T}} d_k$, and the Lipschitz estimate (2.8) for y_k^* into the expression (1.10) to obtain

$$\begin{aligned} |\hat{\beta}_k^N| \leq |\tilde{\beta}_k^N| &\leq \frac{1}{|d_k^{\mathrm{T}} y_k^*|}(|y_k^{*\mathrm{T}} g_{k+1}| + 2\|y_k^*\| \frac{|g_{k+1}^{\mathrm{T}} d_k|}{|y_k^{*\mathrm{T}} d_k|}) \\ &\leq \frac{4}{7(\delta-\sigma)\gamma^2}(2L\Gamma\|s_k\| + 8L^2\|s_k\|^2 \max\{\frac{\sigma}{2(\delta-\sigma)}, \frac{1}{2}\}) \\ &\leq C\|s_k\|, \end{aligned} \tag{3.11}$$

where Γ is defined in (3.4), and where C is defined as follows:

$$C = \frac{4}{7(\delta-\sigma)\gamma^2}(2L\Gamma + 8L^2 D \max\{\frac{\sigma}{2(\delta-\sigma)}, \frac{1}{2}\}).$$
$$D = \max\{\|y-z\| : y, z \in \Omega\}. \tag{3.12}$$

Here D is the diameter of Ω.

(2) A bound on the steps s_k. Observe that for any $l \geq k$,

$$x_l - x_k = \sum_{j=k}^{l-1}(x_{j+1} - x_j) = \sum_{j=k}^{l-1}\|s_j\|u_j = \sum_{j=k}^{l-1}\|s_j\|u_k + \sum_{j=k}^{l-1}\|s_j\|(u_j - u_k). \tag{3.13}$$

By the triangle inequality,

$$\sum_{j=k}^{l-1}\|s_j\| \leq \|x_l - x_k\| + \sum_{j=k}^{l-1}\|s_j\|\|u_j - u_k\| \leq D + \sum_{j=k}^{l-1}\|s_j\|\|u_j - u_k\|. \tag{3.14}$$

Let Δ be a positive integer, chosen large enough so that

$$\Delta \geq 4CD, \tag{3.15}$$

where C and D appear in (3.12) and (3.14). Choose k_0 large enough such that

$$\sum_{i \geq k_0} \|u_{i+1} - u_i\|^2 \leq \frac{1}{4\Delta}. \tag{3.16}$$

By Lemma 3.1, k_0 can be chosen in this way. If $j > k \geq k_0$ and $j - k \leq \Delta$, then by (3.16) and the Cauchy-Schwarz inequality, we have

$$\|u_j - u_k\| \leq \sum_{i=k}^{j-1}\|u_{i+1} - u_i\| \leq \sqrt{j-k}(\sum_{i=k}^{j-1}\|u_{i+1} - u_i\|^2)^{1/2} \leq \sqrt{\Delta}(\frac{1}{4\Delta})^{1/2} = \frac{1}{2}.$$

Combining this with (3.14) yields

$$\sum_{j=k}^{l-1}\|s_j\| \le 2D, \tag{3.17}$$

when $l > k \ge k_0$ and $l - k \le \Delta$.

(3) A bound on the directions d_l. By (1.3) and the bound on $\hat{\beta}_k^N$ given in step I, we have

$$\|d_l\|^2 \le (\|g_l\| + |\hat{\beta}_{l-1}^N|\|d_{l-1}\|)^2 \le 2\Gamma^2 + 2C^2\|s_{l-1}\|^2\|d_{l-1}\|^2,$$

where Γ is the bound on the gradient given in (3.4). Defining $S_i = 2C^2\|s_i\|^2$, we conclude that for $l > k_0$,

$$\|d_l\|^2 \le 2\Gamma^2(\sum_{i=k_0+1}^{l}\prod_{j=i}^{l-1}S_j) + \|d_{k_0}\|^2\prod_{j=k_0}^{l-1}S_j. \tag{3.18}$$

Above, the product is defined to be 1 whenever the index range is vacuous. Let us consider as follows a product of Δ consecutive S_j, where $k \ge k_0$:

$$\prod_{j=k}^{k+\Delta-1}S_j = \prod_{j=k}^{k+\Delta-1}2C^2\|s_j\|^2 = (\prod_{j=k}^{k+\Delta-1}\sqrt{2}C\|s_j\|)^2$$
$$\le (\frac{\sum_{j=k}^{k+\Delta-1}\sqrt{2}C\|s_j\|}{\Delta})^{2\Delta} \le (\frac{2\sqrt{2}CD}{\Delta})^{2\Delta} \le \frac{1}{2^\Delta}.$$

The first inequality above is the arithmetic-geometric mean inequality, the second is due to (3.17), and the third comes from (3.15). Since the product of Δ consecutive S_j is bounded by $1/2^\Delta$, it follows that the sum in (3.18) is bounded, and the bound is independent of l. This bound for $\|d_l\|$, independent of $l > k_0$, contradicts (3.1). Hence, $\gamma = \liminf_{k\to\infty}\|g_k\| = 0$.

4. Numerical experiment

In this section, we will test the following five methods:

PRPSWP the PRP formula with the SWP condition, where $\delta = 0.01$, $\sigma = 0.1$, the termination condition is $\|g_k\| \le 10^{-5}$.

PRP$^+$SWP the PRP formula with nonnegative values $\beta_k = \max\{0, \beta_k^{\mathrm{PRP}}\}$ and the SWP condition, where $\delta = 0.01$, $\sigma = 0.1$, the termination condition is $\|g_k\| \le 10^{-5}$.

CGSWP the CG formula with values $\beta_k = \bar{\beta}_k^N$ defined by (1.9) and the SWP condition, where $\delta = 0.01$, $\sigma = 0.1$, $\eta = 0.01$, the termination condition is $\|g_k\| \le 10^{-5}$.

NCGRWP the NCG formula with values $\beta_k = \hat{\beta}_k^N$ defined by (1.10) and the RWP condition, where $\delta = 0.1$, $\sigma = 0.099$, $\eta = 0.01$, the termination condition is $\|g_k\| \le 10^{-5}$.

NCGWWP the NCG formula with values $\beta_k = \hat{\beta}_k^N$ defined by (1.10) and the WWP condition, where $\delta = 0.1$, $\sigma = 0.9$, $\eta = 0.01$, the termination condition is $\|g_k\| \le 10^{-5}$.

The experiments were carried out on some famous test problems which can be obtained on net. In the following tables, the numerical results are written in the form NI/NF/NG, where NI, NF, NG denote the number of iterations, function evaluations and gradient evaluations, respectively. Problem denotes the name of the problem in MATLAB. Dim denotes the dimension of the test problems. "-" denotes that this method failed to yield a solution for the problem.

Table 1 Numerical results

		PRPSWP	PRP$^+$SWP	CGSWP	NCQRWP	NCQWWP
Problem	Dim	NI/NF/NG	NI/NF/NG	NI/NF/NG	NI/NF/NG	NI/NF/NG
Rose	2	29/502/65	22/394/60	25/375/64	29/447/49	45/145/59
Froth	2	12/30/20	10/28/20	11/75/18	11/215/14	30/50/34
Gauss	3	4/57/6	4/57/6	4/9/5	4/9/5	4/9/5
Meyer	3	–	–	–	–	–
Gulf	3	1/2/2	1/2/2	1/2/2	1/2/2	1/2/2
Sing	4	231/756/378	49/155/79	152/597/244	124/628/176	84/112/88
Badscp	2	–	46/474/114	42/499/101	–	99/330/129
Badscb	2	13/80/22	11/123/22	11/73/19	11/76/17	–
Beale	2	9/126/21	9/173/20	15/229/24	11/167/15	24/40/28
Jeasam	3	–	–	11/77/19	11/120/16	17/30/18
Helix	3	49/255/83	32/265/55	41/330/72	49/582/66	84/117/91
Bard	3	23/98/37	27/152/43	31/67/53	35/401/42	43/65/48
Wood	4	337/2125/599	101/549/195	105/407/174	54/441/74	151/368/171
Kowosb	4	62/361/105	51/249/79	106/306/170	124/759/168	136/184/148
Bd	4	–	–	70/389/110	47/379/113	–
Osb1	5	1/51/2	1/51/2	1/51/2	1/51/2	1/51/2
Biggs	6	121/495/197	–	101/503/172	10/308/12	–
Osb2	11	293/1372/480	250/1011/412	321/1394/540	52/773/66	833/943/857
Watson	20	990/2773/1567	2143/5780/3396	2169/5476/3434	1663/8386/2345	4325/5678/4783
Rosex	50	31/533/76	24/492/60	24/283/66	28/631/40	50/105/68
Singx	4	231/756/378	49/155/79	152/597/244	124/128/176	84/112/88
Pen1	2	5/18/12	6/20/14	5/18/12	4/107/7	7/156/8
Pen2	4	12/134/28	12/136/27	13/151/33	12/142/24	29/58/36
	50	906/4075/1585	136/898/282	686/3023/1219	243/1541/353	306/478/367
Vardim	2	3/9/7	3/9/7	3/9/7	3/55/6	7/14/9
	50	10/52/36	10/52/36	10/52/36	9/44/19	19/32/24
Trig	50	41/279/72	41/230/72	39/368/69	46/510/52	50/69/52
	100	46/342/87	46/341/85	50/440/95	46/704/53	61/82/64
Bv	3	12/25/16	12/25/16	9/67/11	8/64/9	11/20/12
	10	75/241/117	75/241/117	43/83/66	52/477/64	337/380/347
ie	3	5/12/7	6/14/8	5/12/7	6/61/7	6/13/7
	100	6/13/8	6/13/8	6/61/8	7/62/8	7/14/8
	200	6/13/8	6/13/8	6/13/8	8/16/9	8/16/9
	500	6/13/8	6/13/8	6/13/8	7/63/8	7/14/8
Trid	100	30/67/36	30/67/36	30/67/36	27/104/29	26/50/27
	200	30/66/36	30/66/36	30/160/36	29/152/31	26/49/27
Band	3	9/68/13	10/23/17	8/21/14	9/67/11	10/21/12
	50	18/183/24	16/331/25	21/773/30	20/523/24	22/44/24
	100	18/183/24	16/373/26	21/778/30	2/101/14	23/45/24
	200	19/283/27	17/340/27	21/822/28	3/102/5	22/43/24
Lin	2	1/3/3	1/3/3	1/3/3	1/3/3	3/5/4
	50	1/3/3	1/3/3	1/3/3	1/3/3	1/3/3
	500	1/3/3	1/3/3	1/3/3	1/3/3	1/3/3
	1000	1/3/3	1/3/3	1/3/3	1/3/3	1/3/3
Lin1	2	1/51/2	1/51/2	1/51/2	1/51/2	2/4/3
	10	1/3/3	1/3/3	1/3/3	1/3/3	3/5/4
Lino	4	1/3/3	1/3/3	1/3/3	1/51/2	2/4/3

In order to rank these methods, we compute the total number of function and gradient evaluations by the following formula

$$N_{\text{total}} = \text{NF} + 5 * \text{NG}. \tag{4.1}$$

In this part, we compare the PRP$^+$SWP, CQSWP, NCQRWP and the NCQWWP with the PRPSWP as follows: for each testing example i, compute the total numbers of function evaluation and gradient evaluations required by the evaluated method $j(\text{EM}(j))$ and the PRPSWP method by the formula (4.1), and denote them by $N_{\text{total},i}(\text{EM}(j))$ and $N_{\text{total},i}(\text{PRP})$; then calculate the radio

$$r_i(\text{EM}(j)) = \frac{N_{\text{total},i}(\text{EM}(j))}{N_{\text{total},i}(\text{PRP})}. \tag{4.2}$$

If $\text{EM}(j_0)$ does not work for example i_0, but EM(PRP) works for example i_0, we replace the $r_{i_0}(\text{EM}(j_0))$ by a constant τ_1 which is defined as follows:

$$\tau_1 = \max\{r_i(EM(j_0)) : \ (i, j_0) \notin S_1\},$$

where $S_1 = \{(i, j_0)$: method j_0 does not work for example $i\}$.

If EM(PRP) does not work for example i_0, but $\text{EM}(j_0)$ works for example i_0, we replace the $r_{i_0}(\text{EM}(j_0))$ by a constant τ_2 which is defined as follows:

$$\tau_2 = \min\{r_i(\text{EM}(j_0)) : \ (i, j_0) \notin S_1\}.$$

If EM(PRP) and $\text{EM}(j_0)$ do not work for example i_0, then we define $r_{i_0}(\text{EM}(j_0)) = 1$. The geometric mean of these ratios for method j over all the test problems is defined by:

$$r(\text{EM}(j)) = (\prod_{i \in S} r_i(\text{EM}(j)))^{\frac{1}{|S|}}, \tag{4.3}$$

where S denotes the set of the test problems and $\mid S \mid$ the number of elements in S. One advantage of the above rule is that, the comparison is relative and hence is not dominated by a few problems for which the method requires a great deal of function evaluations and gradient functions.

According to the above rule, it is clear that r(PRPSWP)= 1. The values of $r(\text{PRP}^+SWP)$, r(CGSWP), r(NCGRWP) and r(NCGWWP) are listed in the following.

Table 2 Relative efficiency of numerical results

PRPSWP	PRP^+SWP	CGSWP	NCGRWP	NCGWWP
1	0.9337	0.8951	0.9372	0.7594

From above numerical analysis, we can see that the PRPSWP, PRP$^+$SWP and the NCGRWP methods are similar. Also, the numerical results of the NCGWWP method is very efficient, but we can not prove its global convergence theoretically.

Remark The NCGWWP algorithm may have the global convergence and this might be an important topic of further research.

References

[1] FLETCHER R, REEVES C M. *Function minimization by conjugate gradients* [J]. Comput. J., 1964, **7**: 149–154.

[2] POLAK E, RIBIÈRE G. *Note sur la convergence de méthodes de directions conjuguées* [J]. Rev. Francaise Informat. Recherche Opérationnelle, 1969, **16**: 35–43. (in French)

[3] POLYAK B T. *The conjugate gradient method in extreme problems* [J]. UUSR Comput. Math. and Math. Phys., 1969, **9**: 94–112.

[4] HESTENES M R, STIEFEL E. *Methods of conjugate gradients for solving linear systems* [J]. J. Research Nat. Bur. Standards, 1952, **49**: 409–436.

[5] FLETCHER R. *Practial Methods of Optimization* [M]. John Wiley & Sons, Ltd., Chichester, 1987.

[6] DAI Yuhong, YUAN Yaxiang. *Nonlinear Conjugate Gradient Methods* [M]. Shanghai: Shanghai Science and Technology Publisher, 2000. (in Chinese)

[7] YUAN Yaxiang, SUN Wenyu. *Optimization Theories and Methods* [M]. Beijing: Science Press, 1997. (in Chinese)

[8] WEI Zengxin, LI Guoyin, QI Liqun. *New quasi-Newton methods for unconstrained optimization problems* [J]. Appl. Math. Comput., 2006, **175**(2): 1156–1188.

[9] HAGER W W, ZHANG Hongchao. *A new conjugate gradient method with guaranteed descent and an efficient line search* [J]. SIAM J. Optim., 2005, **16**(1): 170–192.

[10] ZOUTENDIJK G. *Nonlinear Programming, Computational Methods* [M]. North-Holland, Amsterdam, 1970.

[11] GILBERT J C, NOCEDAL J. *Global convergence properties of conjugate gradient methods for optimization* [J]. SIAM J. Optim., 1992, **2**(1): 21–42.

[12] DAI Yuhong, HAN Jiye, LIU Guanghui. et al. *Convergence properties of nonlinear conjugate gradient methods* [J]. SIAM J. Optim., 2000, **10**(2): 345–358.

[13] DAI Yuhong, LIAO Lizhi. *New conjugacy conditions and related nonlinear conjugate gradient methods* [J]. Appl. Math. Optim., 2001, **43**(1): 87–101.

[14] DAI Yuhong, YUAN Yaxiang. *A nonlinear conjugate gradient method with a strong global convergence property* [J]. SIAM J. Optim., 1999, **10**(1): 177–182.

[15] DAI Yuhong, YUAN Yaxiang. *An efficient hybrid conjugate gradient method for unconstrained optimization* [J]. Ann. Oper. Res., 2001, **103**: 33–47.

[16] YABE H, TAKANO M. *Global convergence properties of nonlinear conjugate gradient methods with modified secant condition* [J]. Comput. Optim. Appl., 2004, **28**(2): 203–225.

[17] WEI Zengxin, QI Liqun, CHEN Xiaojun. *An SQP-type method and its application in stochastic programs* [J]. J. Optim. Theory Appl., 2003, **116**(1): 205–228.