Wuhan University
Journal of Natural Sciences

2024, Vol.29 No.5, 403-411

Article ID 1007-1202(2024)05-0403-09 DOI https://doi.org/10.1051/wujns/2024295403

Cite this article: ZHU Tiefeng. A New Two-Parameter Family of Nonlinear Conjugate Gradient Method Without Line Search for Unconstrained Optimization Problem[J]. *Wuhan Univ J of Nat Sci*, 2024, 29(5): 403-411.

A New Two-Parameter Family of Nonlinear Conjugate Gradient Method Without Line Search for Unconstrained Optimization Problem

□ **ZHU Tiefeng**

School of Statistics and Mathematics, Inner Mongolia University of Finance and Economics, Hohhot 010070, Inner Mongolia Autonomous Region, China

Abstract: This paper puts forward a two-parameter family of nonlinear conjugate gradient (CG) method without line search for solving unconstrained optimization problem. The main feature of this method is that it does not rely on any line search and only requires a simple step size formula to always generate a sufficient descent direction. Under certain assumptions, the proposed method is proved to possess global convergence. Finally, our method is compared with other potential methods. A large number of numerical experiments show that our method is more competitive and effective.

Key words: unconstrained optimization; conjugate gradient method without line search; global convergence

CLC number: O221.2

0 Introduction

Unconstrained optimization methods are widely used in the fields of nonlinear

dynamic systems and engineering computation to obtain the numerical solution of the optimal control problem[1,2]. In this paper, we consider the following unconstrained optimization problem:

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

where $f: \mathbb{R}^n \to \mathbb{R}$ is sufficiently smooth high-dimensional function. The nonlinear conjugate gradient (CG) methods are highly useful for solving this kind of problem because they can avoid storing and remembering matrices[3,4]. In general, the iterative formula of the CG method is obtained by

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k, \tag{2}$$

where α_k is step size along the search direction $\boldsymbol{d}_k$, and $\boldsymbol{d}_k$ defined by

$$\boldsymbol{d}_k = \begin{cases} -\boldsymbol{g}_k & ,k=1, \\ -\boldsymbol{g}_k + \beta_k \boldsymbol{d}_{k-1} & ,k \geqslant 2, \end{cases} \tag{3}$$

where $\boldsymbol{g}_k$ denotes the gradient $\nabla f(\boldsymbol{x}_k)$. β_k is scalar and it is chosen differently so that there are different CG methods corresponding to the different β_k. Some well-known formulas for β_k are given by

Received date: 2023-11-28

Foundation item: Supported by 2023 Inner Mongolia University of Finance and Economics, General Scientific Research for Universities directly under Inner Mongolia, China (NCYWT23026), and 2024 High-quality Research Achievements Cultivation Fund Project of Inner Mongolia University of Finance and Economics, China (GZCG2479)

Biography: ZHU Tiefeng, male, Associate professor, research direction: optimization theory and methods, reliability statistics. E-mail: tfzhu2016@163.com

$\beta_k^{\mathrm{HS}} = \dfrac{\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{y}_{k-1}}{\boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1}}$ (Hestenes-Stiefel (HS) method[5]),

$\beta_k^{\mathrm{FR}} = \dfrac{\|\boldsymbol{g}_k\|^2}{\|\boldsymbol{g}_{k-1}\|^2}$ (Fletcher-Reeves (FR) method[6]),

$\beta_k^{\mathrm{PRP}} = \dfrac{\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{y}_{k-1}}{\|\boldsymbol{g}_{k-1}\|^2}$ (Polak-Ribiere-Polyak (PRP) method[7]),

$\beta_k^{\mathrm{CD}} = -\dfrac{\|\boldsymbol{g}_k\|^2}{\boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{g}_{k-1}}$ (Conjugate Descent (CD) method[8]),

$\beta_k^{\mathrm{LS}} = \dfrac{\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{y}_{k-1}}{-\boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{g}_{k-1}}$ (Liu -Storey (LS) method[9]),

and

$\beta_k^{\mathrm{DY}} = \dfrac{\|\boldsymbol{g}_k\|^2}{\boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1}}$ (Dai and Yuan (DY) method[10]),

where $\|\cdot\|$ is Euclidean norm and "T" stands for the transpose, $\boldsymbol{y}_{k-1} = \boldsymbol{g}_k - \boldsymbol{g}_{k-1}$. In general, DY and FR methods based on inexact line search have good convergence performance, but the computational efficiency is not as good as that of PRP method. In order to establish methods with both good numerical performance and convergence properties, many scholars have proposed improved conjugate gradient methods in recent years[11-14]. They showed that the method is globally convergent if the following strong Wolfe line search conditions for α_k are satisfied,

$$\begin{cases} f(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) - f(\boldsymbol{x}_k) \leqslant \rho \alpha_k \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{d}_k \\ \left| g(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k)^{\mathrm{T}} \boldsymbol{d}_k \right| \leqslant -\sigma \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{d}_k \end{cases}, \tag{4}$$

where $0 < \rho < \sigma < 1$.

The line search in the conjugate gradient method is usually chosen by a strong Wolfe line search, however, the line search skill usually brings computational burden, especially in solving large-scale nonlinear unconstrained optimization problem. To overcome this problem, Sun and Zhang[3] introduced five CG methods respectively without line search in which the line search step is replaced by a simple step size formula α_k:

$$\alpha_k = \frac{-\delta \boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{d}_k}{\|\boldsymbol{d}_k\|_{\boldsymbol{Q}_k}^2}, \tag{5}$$

where $\|\boldsymbol{d}_k\|_{\boldsymbol{Q}_k} = \sqrt{\boldsymbol{d}_k^{\mathrm{T}} \boldsymbol{Q}_k \boldsymbol{d}_k}$, $\delta \in (0, v_{\min}/\mu)$ is chosen such that $\delta\mu/v_{\min} < 1$, μ is Lipchitz constant defined in Assumption 1 below, and $\{\boldsymbol{Q}_k\}$ is a sequence of positive definite matrices satisfying for positive constants $v_{\min} > 0$ and $v_{\max} > 0$ such that $\forall \boldsymbol{d} \in \mathbb{R}^n$, $v_{\min} \boldsymbol{d}^{\mathrm{T}} \boldsymbol{d} \leqslant \boldsymbol{d}^{\mathrm{T}} \boldsymbol{Q}_k \boldsymbol{d} \leqslant v_{\max} \boldsymbol{d}^{\mathrm{T}} \boldsymbol{d}$.

Chen and Sun[4] extended this technique to two-parameter family CG method, which can be deemed as generalization of the HS and LS without line search. Yu[15] and Narushima[16] applied this method to memory gradient method and obtained their global convergence, respectively. Yin and Chen[17] showed that three-term CG methods without line search are globally convergence. Other modified CG methods without line search reader may see Refs.[18-20].

As supplementary of these results, in this paper, we continue their research and would show that the proposed CG method without line search in which the line search step is replaced by fixed step-length formula (5), is global convergence for following new β_k:

$$\beta_k = \frac{\|\boldsymbol{g}_k\|^2}{\omega_k \|\boldsymbol{g}_{k-1}\|^2 + \mu_k \boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1}}, \tag{6}$$

where $\omega_k \geqslant 0$, $\mu_k \geqslant 0$. This motivation mainly comes from Refs. [3, 4]. Our main aim is to develop a new method and obtain better property of the new method while keeping its simple structure. It should be noted that the formula (6) includes some important special sub-classes. For example, when $\omega_k = 0$, $\mu_k = 1$, the proposed CG method can be deemed as DY method without line search; when $\omega_k = 1$, $\mu_k = 0$, the proposed CG method can be deemed as FR method without line search. Hence, the proposed CG method (6) without line search can be deemed as generalization of the DY and FR.

The rest of this paper is organized as follows. In Section 1, we first give some preliminary results on the CG method without line search and discuss our method sufficient descent property. Furthermore, we prove the global convergence of the proposed method without line search and present algorithm frame. A large amount of numerical experiments are given for illustrative purposes in Section 2. Conclusions appear in Section 3.

1 Analysis of Global Convergence and Algorithm Frame

1.1 Analysis of Global Convergence

We adopt the following assumption on function of which is commonly used in the literature.

Assumption 1[3] The objective function f is LC^1 in a neighborhood Ω of the level set $L := \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) \leqslant f(\boldsymbol{x}_1)\}$ and L is bounded. Here, by LC^1 we mean that the gradient $\nabla f(\boldsymbol{x})$ is Lipschitz continuous with modulus μ, i.e., there exist $\mu > 0$ such that

$$\left\|\nabla f\left(\boldsymbol{x}_{k+1}\right)-\nabla f\left(\boldsymbol{x}_{k}\right)\right\| \leqslant \mu\left\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_{k}\right\|$$

for any $\boldsymbol{x}_{k+1}$, $\boldsymbol{x}_{k} \in \Omega$.

Assumption 2[3] The function f is LC^1 and strongly convex on Ω. In other words, there exists $\lambda>0$ such that

$$\left[\nabla f\left(\boldsymbol{x}_{k+1}\right)-\nabla f\left(\boldsymbol{x}_{k}\right)\right]^{\mathrm{T}}\left(\boldsymbol{x}_{k+1}-\boldsymbol{x}_{k}\right) \geqslant \lambda\left\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_{k}\right\|^{2}$$

for any $\boldsymbol{x}_{k+1}$, $\boldsymbol{x}_{k} \in \Omega$.

Lemma 1 Suppose that $\boldsymbol{x}_k$ is given by (2), (3) and (5). Then

$$\boldsymbol{g}_{k+1}{}^{\mathrm{T}} \boldsymbol{d}_{k}=\rho_{k} \boldsymbol{g}_{k}{}^{\mathrm{T}} \boldsymbol{d}_{k} \tag{7}$$

holds for all k, where

$$\rho_{k}=1-\frac{\delta \phi_{k}\left\|\boldsymbol{d}_{k}\right\|^{2}}{\left\|\boldsymbol{d}_{k}\right\|_{Q_{k}}^{2}} \tag{8}$$

and

$$\phi_{k}=\begin{cases} 0, & \text{for } \alpha_{k}=0, \\ \dfrac{\left(\boldsymbol{g}_{k+1}-\boldsymbol{g}_{k}\right)^{\mathrm{T}}\left(\boldsymbol{x}_{k+1}-\boldsymbol{x}_{k}\right)}{\left\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_{k}\right\|^{2}}, & \text{for } \alpha_{k} \neq 0. \end{cases} \tag{9}$$

Proof See Lemma 1 in Ref. [3].

Lemma 2 Suppose that Assumption 1 holds and that $\boldsymbol{x}_k$ is given by (2), (3) and (5). Then

$$\liminf_{k \rightarrow \infty}\left\|\boldsymbol{g}_{k}\right\| \neq 0 \text{ implies } \sum_{\boldsymbol{d}_{k} \neq \mathbf{0}} \frac{\left\|\boldsymbol{g}_{k}\right\|^{4}}{\left\|\boldsymbol{d}_{k}\right\|^{2}}<\infty . \tag{10}$$

Proof See Lemma 5 in Ref. [3].

Lemma 3 Suppose that Assumption 2 holds and β_k is given by (6). Then $\boldsymbol{g}_{k}{}^{\mathrm{T}} \boldsymbol{d}_{k} \leqslant-\left\|\boldsymbol{g}_{k}\right\|^{2}$.

Proof We prove Lemma 3 by induction. For $k=1$, it is easy to obtain that $\boldsymbol{g}_{1}^{\mathrm{T}} \boldsymbol{d}_{1}=-\left\|\boldsymbol{g}_{1}\right\|^{2} \leqslant-\left\|\boldsymbol{g}_{1}\right\|^{2}$. Suppose $\boldsymbol{g}_{k-1}^{\mathrm{T}} \boldsymbol{d}_{k-1} \leqslant-\left\|\boldsymbol{g}_{k-1}\right\|^{2}$ holds, now we prove that $\boldsymbol{g}_{k}^{\mathrm{T}} \boldsymbol{d}_{k} \leqslant-\left\|\boldsymbol{g}_{k}\right\|^{2}$ holds.

By (3), (6) and Lemma 1, we have

$$\begin{aligned}
\boldsymbol{g}_{k}{}^{\mathrm{T}} \boldsymbol{d}_{k} &=\boldsymbol{g}_{k}{}^{\mathrm{T}}\left(-\boldsymbol{g}_{k}+\beta_{k} \boldsymbol{d}_{k-1}\right) \\
&=\boldsymbol{g}_{k}{}^{\mathrm{T}}\left(-\boldsymbol{g}_{k}+\frac{\left\|\boldsymbol{g}_{k}\right\|^{2}}{\omega_{k}\left\|\boldsymbol{g}_{k-1}\right\|^{2}+\mu_{k} \boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1}} \boldsymbol{d}_{k-1}\right) \\
&=-\left\|\boldsymbol{g}_{k}\right\|^{2}+\frac{\left\|\boldsymbol{g}_{k}\right\|^{2}}{\omega_{k}\left\|\boldsymbol{g}_{k-1}\right\|^{2}+\mu_{k} \boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1}} \boldsymbol{g}_{k}{}^{\mathrm{T}} \boldsymbol{d}_{k-1} \\
&=-\left\|\boldsymbol{g}_{k}\right\|^{2}+\left\|\boldsymbol{g}_{k}\right\|^{2} \frac{\rho_{k-1} \boldsymbol{g}_{k-1}{}^{\mathrm{T}} \boldsymbol{d}_{k-1}}{\omega_{k}\left\|\boldsymbol{g}_{k-1}\right\|^{2}+\mu_{k} \boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1}} .
\end{aligned}$$

From Corollary 3 in Ref. [3], we know $0<\rho_{k-1} \leqslant 1$. From the induction hypothesis, we have $\rho_{k-1} \boldsymbol{g}_{k-1}^{\mathrm{T}} \boldsymbol{d}_{k-1} \leqslant -\rho_{k-1}\left\|\boldsymbol{g}_{k-1}\right\|^{2} \leqslant 0$. Furthermore, one has

$$\begin{aligned}
\mu_{k} \boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1} &=\mu_{k} \boldsymbol{d}_{k-1}^{\mathrm{T}}\left(\boldsymbol{g}_{k}-\boldsymbol{g}_{k-1}\right) \\
&=\mu_{k}\left(\boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{g}_{k}-\boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{g}_{k-1}\right) \\
&=\mu_{k}\left(\rho_{k-1} \boldsymbol{g}_{k-1}^{\mathrm{T}} \boldsymbol{d}_{k-1}-\boldsymbol{g}_{k-1}^{\mathrm{T}} \boldsymbol{d}_{k-1}\right) \\
&=\mu_{k}\left(\rho_{k-1}-1\right) \boldsymbol{g}_{k-1}^{\mathrm{T}} \boldsymbol{d}_{k-1} \geqslant 0,
\end{aligned}$$

which is due to $\mu_{k} \geqslant 0$, $\rho_{k-1}-1<0$ and $\boldsymbol{g}_{k-1}^{\mathrm{T}} \boldsymbol{d}_{k-1} \leqslant 0$. From above analysis, we have

$$\left\|\boldsymbol{g}_{k}\right\|^{2} \frac{\rho_{k-1} \boldsymbol{g}_{k-1}{}^{\mathrm{T}} \boldsymbol{d}_{k-1}}{\omega_{k}\left\|\boldsymbol{g}_{k-1}\right\|^{2}+\mu_{k} \boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1}} \leqslant 0,$$

where ω_k and μ_k are not equal to zero at the same time. Thus, we obtain $\boldsymbol{g}_{k}{}^{\mathrm{T}} \boldsymbol{d}_{k}=-\left\|\boldsymbol{g}_{k}\right\|^{2}+\left\|\boldsymbol{g}_{k}\right\|^{2} \frac{\rho_{k-1} \boldsymbol{g}_{k-1}{}^{\mathrm{T}} \boldsymbol{d}_{k-1}}{\omega_{k}\left\|\boldsymbol{g}_{k-1}\right\|^{2}+\mu_{k} \boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1}} \leqslant-\left\|\boldsymbol{g}_{k}\right\|^{2}$. This finishes our proof.

If there exists a constant $c>0$ such that $\boldsymbol{g}_{k}{}^{\mathrm{T}} \boldsymbol{d}_{k} \leqslant -c\left\|\boldsymbol{g}_{k}\right\|^{2} (k \geqslant 1)$, then we say the search direction $\boldsymbol{d}_k$ of the method satisfies the sufficient descent condition. Lemma 3 means that the search direction $\boldsymbol{d}_k$ is the sufficient descent direction, which is often required in the convergence analysis of CG method.

Theorem 1 Suppose that Assumptions 1, 2 hold and that β_k is given by (6). Then

$$\frac{\left\|\boldsymbol{d}_{k}\right\|^{2}}{\left\|\boldsymbol{g}_{k}\right\|^{4}} \leqslant \frac{\left\|\boldsymbol{d}_{k-1}\right\|^{2}}{\omega_{k}\left\|\boldsymbol{g}_{k-1}\right\|^{4}}+\frac{H}{\left\|\boldsymbol{g}_{k}\right\|^{2}},$$

where H is a positive constant.

Proof By (3), we have

$$\boldsymbol{d}_{k}=-\boldsymbol{g}_{k}+\beta_{k} \boldsymbol{d}_{k-1},$$

squaring both sides of it, we obtain

$$\left\|\boldsymbol{d}_{k}\right\|^{2}=\left\|-\boldsymbol{g}_{k}+\beta_{k} \boldsymbol{d}_{k-1}\right\|^{2} .$$

By Lemma 1 and substituting the expression (6) into above formula, we have

$$\begin{aligned}
\left\|\boldsymbol{d}_{k}\right\|^{2} &=\left\|-\boldsymbol{g}_{k}+\frac{\left\|\boldsymbol{g}_{k}\right\|^{2}}{\omega_{k}\left\|\boldsymbol{g}_{k-1}\right\|^{2}+\mu_{k} \boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1}} \boldsymbol{d}_{k-1}\right\|^{2} \\
&=\left\|\boldsymbol{g}_{k}\right\|^{2}-\frac{2\left\|\boldsymbol{g}_{k}\right\|^{2}}{\omega_{k}\left\|\boldsymbol{g}_{k-1}\right\|^{2}+\mu_{k} \boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1}} \boldsymbol{g}_{k}{}^{\mathrm{T}} \boldsymbol{d}_{k-1} \\
&\quad+\frac{\left\|\boldsymbol{g}_{k}\right\|^{4}\left\|\boldsymbol{d}_{k-1}\right\|^{2}}{\left[\omega_{k}\left\|\boldsymbol{g}_{k-1}\right\|^{2}+\mu_{k} \boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1}\right]^{2}} \\
&=\left\|\boldsymbol{g}_{k}\right\|^{2}-\frac{2 \rho_{k-1} \boldsymbol{g}_{k-1}{}^{\mathrm{T}} \boldsymbol{d}_{k-1}}{\omega_{k}\left\|\boldsymbol{g}_{k-1}\right\|^{2}+\mu_{k} \boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1}}\left\|\boldsymbol{g}_{k}\right\|^{2} \\
&\quad+\frac{\left\|\boldsymbol{g}_{k}\right\|^{4}\left\|\boldsymbol{d}_{k-1}\right\|^{2}}{\left[\omega_{k}\left\|\boldsymbol{g}_{k-1}\right\|^{2}+\mu_{k} \boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1}\right]^{2}} \\
&\leqslant\left\|\boldsymbol{g}_{k}\right\|^{2}\left|1+\frac{2 \rho_{k-1} \boldsymbol{g}_{k-1}^{\mathrm{T}} \boldsymbol{d}_{k-1}}{\omega_{k}\left\|\boldsymbol{g}_{k-1}\right\|^{2}+\mu_{k} \boldsymbol{d}_{k-1}^{\mathrm{T}} \boldsymbol{y}_{k-1}}\right|
\end{aligned}$$

$$+\frac{\left\|\boldsymbol{g}_k\right\|^4\left\|\boldsymbol{d}_{k-1}\right\|^2}{\left[\omega_k\left\|\boldsymbol{g}_{k-1}\right\|^2+\mu_k\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}\right]^2}$$

$$\leqslant\left\|\boldsymbol{g}_k\right\|^2\left|1+\frac{2\rho_{k-1}\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\omega_k\left\|\boldsymbol{g}_{k-1}\right\|^2}\right|+\frac{\left\|\boldsymbol{g}_k\right\|^4\left\|\boldsymbol{d}_{k-1}\right\|^2}{\left[\omega_k\left\|\boldsymbol{g}_{k-1}\right\|^2\right]^2},$$

which is due to $\mu_k\boldsymbol{d}_{k-1}^{\mathrm{T}}\boldsymbol{y}_{k-1}>0$ that is deduced in Lemma 3. By $\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1}\leqslant-\left\|\boldsymbol{g}_{k-1}\right\|^2$ in Lemma 3 , we have

$$\frac{2\rho_{k-1}\boldsymbol{g}_{k-1}^{\mathrm{T}}\boldsymbol{d}_{k-1}}{\omega_k\left\|\boldsymbol{g}_{k-1}\right\|^2}\leqslant-\frac{2\rho_{k-1}\left\|\boldsymbol{g}_{k-1}\right\|^2}{\omega_k\left\|\boldsymbol{g}_{k-1}\right\|^2}=-\frac{2\rho_{k-1}}{\omega_k}.$$

Now by $1-\delta\mu/v_{\min}\leqslant\rho_{k-1}$ that is deduced in Ref. [3], and the two inequalities above, we obtain

$$\left\|\boldsymbol{d}_k\right\|^2\leqslant\left|1-\frac{2\rho_{k-1}}{\omega_k}\right|\left\|\boldsymbol{g}_k\right\|^2+\frac{\left\|\boldsymbol{g}_k\right\|^4\left\|\boldsymbol{d}_{k-1}\right\|^2}{\left[\omega_k\left\|\boldsymbol{g}_{k-1}\right\|^2\right]^2}$$

$$\leqslant\left|1-\frac{2(1-\delta\mu/v_{\min})}{\omega_k}\right|\left\|\boldsymbol{g}_k\right\|^2+\frac{\left\|\boldsymbol{g}_k\right\|^4\left\|\boldsymbol{d}_{k-1}\right\|^2}{\omega_k{}^2\left\|\boldsymbol{g}_{k-1}\right\|^4}.$$

Let $H=\left|1-\frac{2\left(1-\delta\mu/v_{\min}\right)}{\omega_k}\right|$, we obtain

$$\left\|\boldsymbol{d}_k\right\|^2\leqslant H\left\|\boldsymbol{g}_k\right\|^2+\frac{\left\|\boldsymbol{g}_k\right\|^4\left\|\boldsymbol{d}_{k-1}\right\|^2}{\omega_k\left\|\boldsymbol{g}_{k-1}\right\|^4},$$

$$\frac{\left\|\boldsymbol{d}_k\right\|^2}{\left\|\boldsymbol{g}_k\right\|^4}\leqslant\frac{\left\|\boldsymbol{d}_{k-1}\right\|^2}{\omega_k\left\|\boldsymbol{g}_{k-1}\right\|^4}+\frac{H}{\left\|\boldsymbol{g}_k\right\|^2}.$$

The proof is completed. There holds $1-\delta\mu/v_{\min}\leqslant\rho_{k-1}$ for all k if Assumption 1 is valid.

Theorem 2 Suppose that Assumptions 1, 2 hold and that sequence $\{\boldsymbol{x}_k\}$ is given by (2), (3), (5) and (6). Then $\liminf_{k\to\infty}\left\|\boldsymbol{g}_k\right\|=0$.

Proof Suppose, by contradiction, that the stated conclusion is not true. Then, in view of $\left\|\boldsymbol{g}_k\right\|>0$, there exists a constant $\varepsilon>0$, such that $\left\|\boldsymbol{g}_k\right\|\geqslant\varepsilon$ for all k, and by Theorem 1 we have $\frac{\left\|\boldsymbol{d}_k\right\|^2}{\left\|\boldsymbol{g}_k\right\|^4}\leqslant\frac{\left\|\boldsymbol{d}_{k-1}\right\|^2}{\omega_k\left\|\boldsymbol{g}_{k-1}\right\|^4}+\frac{H}{\varepsilon^2}$.

Thus, we have a recursive equation which leads to

$$\frac{\left\|\boldsymbol{d}_k\right\|^2}{\left\|\boldsymbol{g}_k\right\|^4}\leqslant\frac{\left\|\boldsymbol{d}_{k-1}\right\|^2}{\omega_k\left\|\boldsymbol{g}_{k-1}\right\|^4}+\frac{H}{\varepsilon^2}\leqslant\frac{\left\|\boldsymbol{d}_{k-2}\right\|^2}{\omega_k\left\|\boldsymbol{g}_{k-2}\right\|^4}+\frac{2H}{\varepsilon^2}$$

$$\leqslant\cdots\leqslant\frac{\left\|\boldsymbol{d}_1\right\|^2}{\omega_2\left\|\boldsymbol{g}_1\right\|^4}+\frac{(k-1)H}{\varepsilon^2}. \tag{11}$$

Thus, equation (11) means that

$$\frac{\left\|g_k\right\|^4}{\left\|d_k\right\|^2}\leqslant\frac{1}{ab+(k-1)c} \tag{12}$$

where $a=\frac{1}{\omega_2}$, $b=\frac{\left\|\boldsymbol{d}_1\right\|^2}{\left\|\boldsymbol{g}_1\right\|^4}$ and $c=\frac{H}{\varepsilon^2}$, a, b and c are constants. From equation (12) we have $\sum_{d_k\neq0}\frac{\left\|\boldsymbol{g}_k\right\|^2}{\left\|\boldsymbol{d}_k\right\|^2}=+\infty$, which is contradictory to Lemma 2. Hence we obtain $\liminf_{k\to\infty}\left\|\boldsymbol{g}_k\right\|=0$.

Theorem 2 denotes that our proposed CG method without line search possesses global convergence.

1.2 Algorithm Frame

Based on the discussion above, now we can describe the algorithm frame for solving the unconstrained optimization problem (1) as follows:

Step 0 Given an initial point $\boldsymbol{x}_0\in\mathbb{R}^n$, constants $\varepsilon_0>0$, $\rho\in(0,1)$, $\sigma\in(0,1)$, $\omega_k\geqslant0$, $\mu_k\geqslant0$ and let $k:=0$.

Step 1 If a stopping criterion $\left\|\boldsymbol{g}_k\right\|<\varepsilon_0$ is satisfied, then stop; otherwise, go to Step 2.

Step 2 Compute a step size α_k by formula (6).

Step 3 Let $\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k\boldsymbol{d}_k$, compute $\boldsymbol{d}_k$ and β_k by (3) and (6).

Step 4 Let $k:=k+1$ and go to Step 1.

From above algorithm, we note that the step-length α_k depends on the Lipschitz constant, but in general, the Lipschitz constant is not known in previously. It should be noted that if $f(\boldsymbol{x})$ is a twice continuous differentiable strictly convex function, the Lipschitz constant can be replaced by a positive constant[15]. So, we choose $\delta=1$.

2 Numerical Experiments

In this section, in order to show the performance of the given method, we test our proposed method with fixed step-length (5) (denoted by OMFSL), method (6) with the strong Wolfe line search (4) (denoted by SSWLS), and some other considered algorithms without line search such as DY and FR method[3] and Chen and Sun's method[4] via 24 test problems from Ref.[21]. We use the same convergence criteria in Ref.[21].

The parameters are chosen as $\rho=0.04$, $\sigma=0.5$, $\omega_k=0.7$ and $\mu_k=0.3$. All codes are written in MATLAB 7.5 and ran on Lenovo with 1.90 GHz CPU processor, 2.43 GB RAM memory, and Windows XP operating system. In our numerical experiments the no-line-search method with $\boldsymbol{Q}_k=\boldsymbol{I}$ (the unit matrix) show bad convergence behavior. Thus we consider BFGS updates for $\boldsymbol{Q}_k$ in formula (5).

The test results are shown in Tables 1 and 2. The re-

Table 1 Numerical results for the tested problems

No.	Functions/n	k/NG/NF				
		OMFSL	SSWLS	DY[10]	FR[3]	Chen and Sun[4]
1	Sphere/100	2/2/2	7/20/25	2/2/2	2/2/2	2/2/2
2	Rastrigin/10	2/2/2	15/164/297	2/2/2	2/2/2	2/2/2
3	Froth/2	59/59/59	1 144/16 759/31 229	1 996/1 996/1 996	50/50/50	F/F/F
4	Pqd/5	41/41/41	67/209/283	14/14/14	13/13/13	13/13/13
5	Ewh/10	284/284/284	76/998/1847	F/F/F	406/406/406	F/F/F
6	Raydan1/2	8/8/8	29/30/30	8/8/8	8/8/8	8/8/8
7	Raydan2/100	7/7/7	5/9/7	8/8/8	8/8/8	8/8/8
8	Etri/10	124/124/124	F/F/F/F	61/61/61	71/71/71	67/67/67
9	Epow/4	127/27/27	159/1 765/3211	28/28/28	23/23/23	18/18/18
10	Wood/4	213/213/213	509/2 720/4 993	1 512/1 512/1 512	296/296/2 996	F/F/F
11	Ewood/4	268/268/268	F/F/F	1 033/1 033/1 033	248/248/248	F/F/F
12	Perq/3	21/21/21	21/67/87	27/27/27	18/18/18	23/23/23
13	Etri1/34	11/11/11	F/F/F	43/43/43	45/45/45	50/50/50
14	Emic/8	290/290/290	F/F/F	F/F/F	F/F/F	F/F/F
15	Erosen/20	1 072/1 072/1 072	54/732/1357	403/403/403	1 121/1 121/1 121	62/62/62
16	Grosen/3	9/9/9	2 141/6 137/11 293	383/383/383	202/202/202	776/776/776
17	QUARTC/50	2/2/2	F/F/F	2/2/2	2/2/2	2/2/2
18	LIARWHD/10	57/57/57	F/F/F	78/78/78	59/5/59	72/72/72
19	Staircase1/4	14/14/14	34/143/217	23/23/23	14/14/14	14/14/14
20	Staircase2/4	14/14/14	29/123/187	23/23/23	14/14/14	13/13/13
21	POWER/20	44/44/44	F/F/F	102/102/102	117/117/117	53/53/53
22	Diagonal4/4	2/2/2	57/435/755	7/7/7	2/2/2	2/2/2
23	EBD1/2	14/14/14	18/82/127	F/F/F	F/F/F	13/13/13
24	CUBE/2	F/F/F	196/2 529/4 665	F/F/F	F/F/F	F/F/F

Note: F denotes that the method fails to convergence within 5 000 iterations or there exists numerical overflow

Table 2 Results of $f(\bar{x})$ and T_{cpu} for the tested problems

No.	Functions/n	$f(\bar{x})/T_{cpu}$				
		OMFSL	SSWLS	DY[10]	FR[3]	Chen and Sun[4]
1	Sphere/100	0/0.062 5	1.924 7E−11/0.312 5	2.169 4E−30/0.171 9	2.169 4E−30/0.078 1	2.169 4E−30/0.265 6
2	Rastrigin/10	0.0/0.004	3.836 9E−13/0.187 5	0/0.015 6	0/0.005	0/0.007
3	Froth/2	7.854 2E−16/0.015 6	2.583 1E−10/2.187 5	1.148 3E−12/0.046 9	5.928 8E−13/0.015 6	F/F
4	Pqd/5	2.707 2E−9/0.046 9	2.001 3E−8/0.078 1	7.629 2E−8/0.015 6	1.6000E−8/0.031 3	9.231 9E−8/0.031 3
5	Ewh/10	2.578 8E−11/0.025 0	1.978 4E−9/0.593 8	F/F	1.008 8E−9/0.031 3	F/F
6	Raydan1/2	0.3/0.015 6	0.3/0.046 9	0.3/0.015 6	0.3/0.015 6	0.3/0.015 6
7	Raydan2/100	9/0.015 6	9/0.062 5	9/0.015 6	9/0.015 6	9/0.015 6
8	Etri/10	2.300 1E−11/0.015 6	F/F	1.304 4E−10/0.015 6	7.070 4E−10/0.015 6	6.3120 E−10/0.015 6
9	Epow/4	1.169 2E−13/0.015 6	7.092 5E−12/0.140 6	4.052 5E−13/0.015 6	9.708 0E−13/0.046 9	8.836 6E−11/0.078 1
10	Wood/4	3.999 3E−9/0.078 1	9.306 7E−10/0.234 4	7.0341 E−10/0.109 4	8.009 0E−11/0.031 3	F/F
11	Ewood/4	2.966 6E−10/0.048 1	F/F	3.848 0E−100.078 1	5.499 0E−10/0.062 5	F/F
12	Perq/3	3.548 7E−10/0.0156	3.870 9E−10/0.015 6	2.891 0E−10/0.015 6	5.533 2E−10/0.031 3	2.783 2E−11/0.078 1
13	Etri1/34	4.211 3E−14/0.031 3	F/F	5.408 7E−10/0.031 3	3.246 5E−10/0.031 3	6.731 8E−11/0.031 3
14	Emic/8	0.108 9/0.156 3	F/F	F/F	F/F	F/F
15	Erosen/20	1.045 3E−9/0.354 9	1.456 5E−11/2.031 3	2.845 6E−9/0.015 6	1.321 7E−9/0.046 9	4.224 8E−9/0.031 3
16	Grosen/3	5.245 9E−10/0.015 6	8.216 8E−10/0.484 4	4.786 7E−9/0.031 3	3.265 7E−9/0.015 6	3.725 3E−12/0.078 1
17	QUARTC/50	3.111 5E−60/0.015 6	F/F	0/0.015 6	0/0.015 6	0/0.015 6
18	LIARWHD/10	4.331 6E−14/0.015 6	F/F	6.953 2E−10/0.015 6	1.923 9E−10/0.015 6	6.228 2E−11/0.015 6
19	Staircase1/4	4.476 1E−9/0.015 6	3.536 8E−10/0.031 3	5.005 0E−9/0.015 6	4.595 9E−9/0.015 6	6.299 0E−9/0.015 6
20	Staircase2/4	1.119 0E−9/0.015 6	1.888 4E−9/0.015 6	7.512 6E−8/0.015 6	6.4898 E−8/0.015 6	9.150 1E−9/0.015 6
21	POWER/20	1.976 9E−10/0.015 6	F/F	1.206 6E−9/0.046 9	3.272 9E−10/0.031 3	7.501 9E−10/0.015 6
22	Diagonal4/4	3.841 6E−14/0.015 6	1.122 0E−10/0.062 5	2.555 0E−14/0.015 6	3.841 6E−14/0.015 6	3.841 6E−14/0.015 6
23	EBD1/2	9.253 8E−13/0.015 6	1.624 6E−11/0.015 6	F/F	F/F	2.601 2E−10/0.015 6
24	CUBE/2	F/F	1.161 1E−8/0.265 6	F/F	F/F	F/F

Note: F denotes that the method fails to convergence within 5 000 iterations or there exists numerical overflow.

sults for each problem by each method are shown in the form of k/NG/NF/$f(\bar{\boldsymbol{x}})$/T_{cpu}, where k, NG, NF, $f(\bar{\boldsymbol{x}})$, T_{cpu} denote the number of the iteration at the terminate iteration, the total number of calculations of the gradient value of the objective function, the total number of computations of the objective function value, the value of function at the terminate iteration, the CPU time in seconds, respectively. n indicates the variable dimension of the test problems.

From Tables 1 and 2, it is easy to see that OMFSL performs better than other considering methods. Meanwhile, we presented the Dolan and Moré[22] performance profiles for OMFSL and other considered methods. Note that the performance ratio $p(\tau)$ is the probability for a solver for the tested problems with the factor τ of the smallest cost. In Figs. 1-5, the performance of OMFSL with fixed step-length (5) and other considered methods are presented.

As we can see from Fig.1, OMFSL is superior to all other considering methods for the absolute errors of $f(\bar{\boldsymbol{x}})$ versus $f(\boldsymbol{x}^*)$ (the function value at the optimal solution). Figures 2-5 shows that OMFSL is better than all other considered methods for the number of iteration, the number of gradient evaluations (NG), the number of function value evaluations (NF) and CPU time (T_{cpu}). In conclusion, we can see that OMFSL without line search is very competitive for the test problems, and OMFSL is alternative for solving nonlinear unconstrained optimization problems.

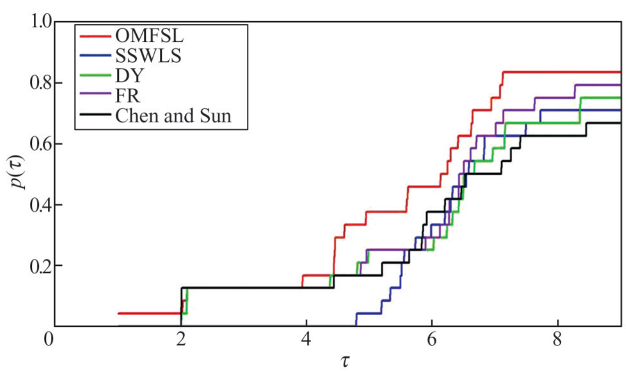

Fig. 1 Performance profile on the absolute errors of $f(\bar{\boldsymbol{x}})$ versus $f(\boldsymbol{x}^*)$

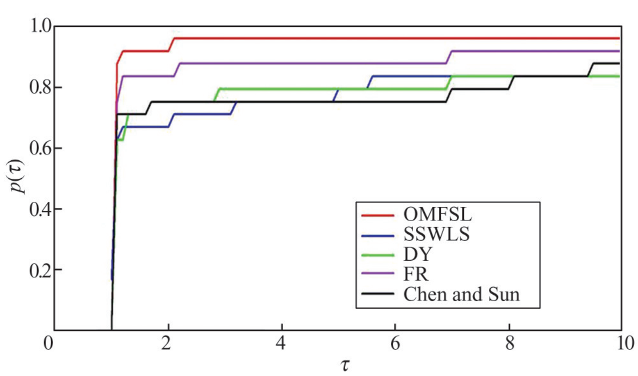

Fig. 2 Performance profile on the number of iteration

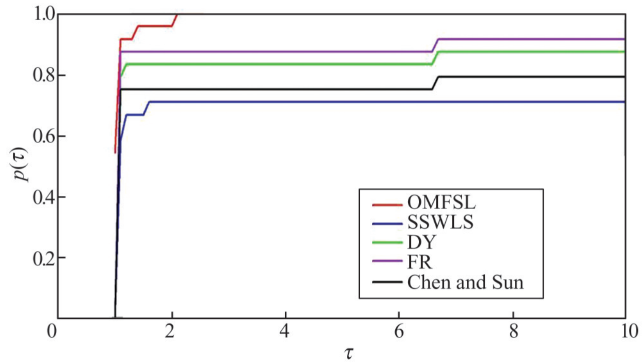

Fig. 3 Performance profile on the NG

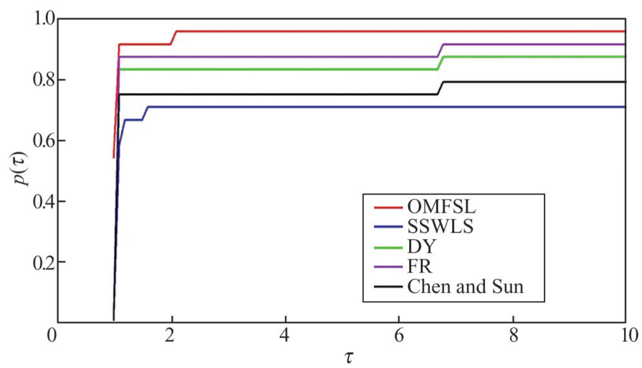

Fig. 4 Performance profile on the NF

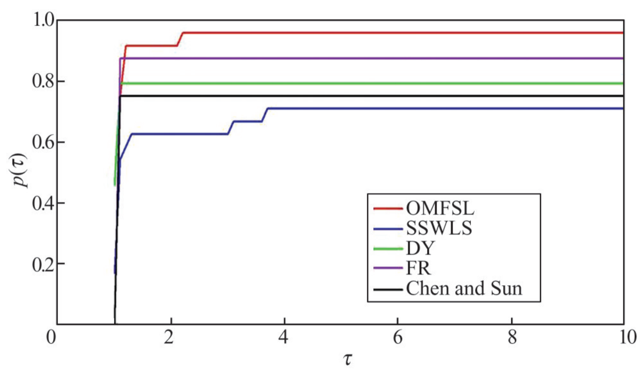

Fig. 5 Performance profile on the T_{cpu}

3 Conclusion

In this paper, we have combined a two-parameter conjugate gradient method defined by formula (6) with

Sun-Zhang's step-length formula defined by formula (5) and have proved its global convergence property under the appropriate assumptions. This step-length formula (5) might be practical in case that the line search is expensive or hard. We allow certain flexibility in selecting the sequence $\{\boldsymbol{Q}_k\}$ in practical computation, and we also can consider other updates for $\boldsymbol{Q}_k$. Our proofs require that the function is at least and the level set is bounded. Reducing the requirement of optimization function and finding completely constant step-length might be topic of further research.

References

[1] Zhang L M, Gao H T, Chen Z Q, *et al*. Multi-objective global optimal parafoil homing trajectory optimization via Gauss pseudo spectral method[J]. *Nonlinear Dynamics*, 2013, **72**(1): 1-8.

[2] Jiang X Z, Jian J B. A sufficient descent Dai-Yuan type nonlinear conjugate gradient method for unconstrained optimization problems[J]. *Nonlinear Dynamics*, 2013, **72**(1): 101-112.

[3] Sun J, Zhang J P. Global convergence of conjugate gradient methods without line search[J]. *Annals of Operations Research*, 2001, **103**(1): 161-173.

[4] Chen X D, Sun J. Global convergence of a two-parameter family of conjugate gradient methods without line search[J]. *Journal of Computational and Applied Mathematics*, 2002, **146**(1): 37-45.

[5] Hestenes M R, Stiefel E. Methods of conjugate gradients for solving linear systems[J]. *Journal of Research of the National Bureau of Standards*, 1952, **49**(6): 409-436.

[6] Fletcher R, Reeves C M. Function minimization by conjugate gradients[J]. *The Computer Journal*, 1964, **7**(2): 149-154.

[7] Polyak B T. The conjugate gradient method in extremal problems[J]. *USSR Computational Mathematics & Mathematical Physics*, 1969, **9**(4): 94-112.

[8] Flecher R. *Practical Methods of Optimization*, *Vol*1: *Unconstrained Optimization*[M]. New York: John Wiley & Sons, 1987.

[9] Liu Y, Storey C. Efficient generalized conjugate gradient algorithms[J]. *Journal of Optimization Theory and Application*, 1991, **69**(1): 129-137.

[10] Dai Y H, Yuan Y X. A nonlinear conjugate gradient with a strong global convergence property[J]. *SIAM Journal on Optimization*, 1999, **10**(1): 177-182.

[11] Deepho J, Abubakar A B, Malik M, *et al*. Solving unconstrained optimization problems via hybrid CD-DY conjugate gradient methods with applications[J]. *Journal of Computational and Applied Mathematics*, 2022, **405**: 113823.

[12] Abubakar A B, Kumam P, Malik M, *et al*. A hybrid conjugate gradient based approach for solving unconstrained optimization and motion control problems[J]. *Mathematics and Computers in Simulation*, 2022, **201**: 640-657.

[13] Goncalves M L N, Lima F S, Prudente L F. A study of Liu-Storey conjugate gradient methods for vector optimization [J]. *Applied Mathematics and Computation*, 2022, **425**: 127099.

[14] Chen C L, Luo L L, Han C H, *et al*. Global convergence of an extended descent algorithm without line search for unconstrained optimization[J]. *Journal of Applied Mathematics and Physics*, 2018, **6**(1): 130-137.

[15] Yu Z S. Global convergence of a memory gradient method without line search[J]. *Journal of Applied Mathematics & Computing*, 2008, **26**(1): 545-553.

[16] Narushima Y S. A memory gradient method without line search for unconstrained optimization[J]. *SUT Journal of Mathematics*, 2006, **42**(2): 191-206.

[17] Yin L, Chen X D. Global convergence of two kinds of three-term conjugate gradient methods without line search[J]. *Asia Pacific Journal of Operational Research*, 2013, **30**(1):1-10.

[18] Li X, Chen X D. Global convergence of shortest-residual family of conjugate gradient methods without line search[J]. *Asia Pacific Journal of Operational Research*, 2005, **22**(4): 529-538.

[19] Du S Q, Chen Y Y. Global convergence of a modified spectral FR conjugate gradient method[J]. *Applied Mathematics and Computation*, 2008, **202**(2): 766-770.

[20] Zhu H, Chen X D. Global convergence of a special case of the Dai-Yuan family without line search[J]. *Asia-Pacific Journal of Operational Research*, 2008, **25**(3): 411-420.

[21] Zhu T F, Yan Z Z, Peng X Y. A modified nonlinear conjugate gradient method for engineering computation[J]. *Mathematical Problems in Engineering*, 2017, **2017**(1): 1425857.

[22] Dolan E D, Moré J J. Benchmarking optimization software with performance profiles[J]. *Mathematical Programming*, 2002, **91**(2): 201-213.

无约束优化问题的一种新的无线搜索的两参数族非线性共轭梯度法

朱铁锋

内蒙古财经大学 统计与数学学院,内蒙古自治区 呼和浩特 010070

摘要：针对无约束优化问题，提出了一种无需线搜索的两参数族非线性共轭梯度法。该方法的主要特点是不依赖于任何线搜索，仅需要一个简单的步长公式总能产生充分下降的方向。在一定的假设条件下，证明了该方法具有全局收敛性。最后，我们的方法与其他数值效果较好的方法进行了比较。大量的数值实验表明，该方法更具有竞争力和有效性。

关键词：无约束优化；无线搜索共轭梯度法；全局收敛性

□