Journal of Mathematical Research with Applications
Sept., 2018, Vol. 38, No. 5, pp. 541–550
DOI:10.3770/j.issn:2095-2651.2018.05.011
Http://jmre.dlut.edu.cn

Global Convergence of Conjugate Gradient Methods without Line Search

Cuiling CHEN[1,2,*], **Yu CHEN**[1]

1. *College of Mathematics and Statistics, Guangxi Normal University, Guangxi* 541004*, P. R. China;*
2. *School of Computing and Information, University of Pittsburgh, Pittsburgh* 15238*, U. S. A.*

Abstract In this paper, a new steplength formula is proposed for unconstrained optimization, which can determine the step-size only by one step and avoids the line search step. Global convergence of the five well-known conjugate gradient methods with this formula is analyzed, and the corresponding results are as follows: (1) The DY method globally converges for a strongly convex LC^1 objective function; (2) The CD method, the FR method, the PRP method and the LS method globally converge for a general, not necessarily convex, LC^1 objective function.

Keywords unconstrained optimization; conjugate gradient method; line search; global convergence

MR(2010) Subject Classification 90C30; 65K05

1. Introduction

The conjugate gradient method is very useful in large-scale unconstrained optimization. For a general unconstrained problem

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

the method takes the following form:

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

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

where $g_k = \nabla f(x_k)$, α_k is a positive steplength determined by a line search, d_k is a search direction, and β_k is a scalar given by different formulae which result in distinct conjugate gradient methods. Several well-known formulae for β_k are given by [1]

$$\beta_k^{CD} = \frac{\|g_k\|^2}{-g_{k-1}^T d_{k-1}} \quad \text{(The Conjugate Descent Method)}, \tag{1.4}$$

Received October 27, 2017; Accepted June 6, 2018
Supported by the National Natural Science Foundation of China (Grant No. 11761014), the Natural Science Foundation of Guangxi Zhuang Autonomous Region (Grant No. 2017GXNSFAA198243), Guangxi Basic Ability Improvement Project for the Middle-Aged and Young Teachers of Colleges and Universities (Grant Nos. 2017KY0068; KY2016YB069), Guangxi Higher Education Undergraduate Course Teaching Reform Project (Grant No. 2017JGB147).
* Corresponding author
E-mail address: mathchen@163.com (Cuiling CHEN); chenyu4660@163.com (Yu CHEN)

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \quad \text{(Fletcher-Reeves)}, \tag{1.5}$$

$$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2} \quad \text{(Polak-Ribière-Polyak)}, \tag{1.6}$$

$$\beta_k^{LS} = -\frac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T d_{k-1}} \quad \text{(Liu-Storey)}, \tag{1.7}$$

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T(g_k - g_{k-1})} \quad \text{(Dai-Yuan)}, \tag{1.8}$$

where $\|\cdot\|$ is the Euclidean norm and "T" stands for the transpose. For ease of presentation we call the methods corresponding to (1.4)–(1.8) the CD method, the FR method, the PRP method, the LS method and the DY method, respectively. The global convergence of these methods has been studied.

As we all know that a key factor of global convergence is how to select the steplength α_k. The commonly-used line search rules are Armijo rule, Goldstein rule and Wolfe rule, and many authors investigated the global convergence of related line search methods [1–3]. It is obvious that any line search rule is a procedure for finding α_k. This certainly adds the number of evaluations for objective functions and gradients. So Dixon [4] proposed a conjugate gradient method without line search. And then Sun and Zhang [5], Chen and Sun [6] investigated some conjugate gradient methods without line search. In [5], the steplength formula is

$$\alpha_k = -\frac{\delta g_k^T d_k}{\|d_k\|_{Q_k}^2}, \tag{1.9}$$

where δ is a parameter and $\|d_k\|_{Q_k} = \sqrt{d_k^T Q_k d_k}$ in which $\{Q_k\}$ is a sequence of positive definite matrices. Evidently, according to (1.9) the steplength is obtained only by one step. And with (1.9), global convergence results are derived for well-known conjugate gradient methods, such as the FR method, the LS method, the DY method, the PRP method and the CD method. In [7], Shi and Shen proposed a new descent method without line search, in which

$$\alpha_k = -\frac{g_k^T d_k}{L_k\|d_k\|^2}, \tag{1.10}$$

or

$$\alpha_k = -\frac{g_k^T d_k}{M_k\|d_k\|^2}, \tag{1.11}$$

where L_k or M_k is a parameter required to be estimated. Under mild conditions, global convergence of the relating algorithms was analyzed.

We note that in the papers without line search, the formula for α_k generally takes (1.9) or (1.10) (see [5–8]). It should be worth researching further whether the formula for α_k can take other forms or not. This motivates us to design a new steplength formula. And with this formula we study global convergence of the five well-known conjugate gradient methods: the CD method, the FR method, the PRP method, the LS method and the DY method. The results show that the new steplength formula can guarantee the global convergence of them. The next section concerns the global convergence.

2. Analysis of global convergence

In order to establish the global convergence, we assume that

Assumption 2.1 The function f is LC^1 in a neighborhood $\mathcal{N}$ of the level set $\mathcal{L} = \{x \in \Re^n | f(x) \leq f(x_1)\}$ and $\mathcal{L}$ is bounded. Here LC^1 means that the gradient g is Lipschitz continuous, i.e., there exists $\mu > 0$ such that $\|g(x) - g(y)\| \leq \mu \|x - y\|$ for any $x, y \in \mathcal{N}$.

Remark 2.2 Since $\mathcal{L}$ is bounded, both $\{x_k\}$ and $\{g_k\}$ are all bounded for the five well-known conjugate gradient methods.

Under Assumption 2.1, it is easy to obtain the global convergence of the CD method, the FR method, the PRP method and the LS method, but it seems not easy to derive the global convergence of the DY method. Thus, for the method we impose the following stronger assumption.

Assumption 2.3 The function f is LC^1 and strongly convex on $\mathcal{N}$. That is to say, there exists $\lambda > 0$ such that $[g(x) - g(y)]^T (x - y) \geq \lambda \|x - y\|^2$ for any $x, y \in \mathcal{N}$.

Remark 2.4 Note that Assumption 2.3 implies Assumption 2.1 since a strongly convex function has bounded level sets.

First we present the new steplength formula as follows:

$$\alpha_k = -\frac{\delta g_k^T d_k}{\|g_k\|^2 + \|d_k\|^2}, \tag{2.1}$$

where $0 < \delta < \min\{\frac{1}{\mu}, \frac{1}{\lambda}\}$. Combining this formula, we now analyze the global convergence of the five well-known conjugate gradient methods.

Lemma 2.5 *Suppose that x_k is given by (1.2), (1.3) and (2.1). Then*

$$g_{k+1}^T d_k \leq \rho_k g_k^T d_k \tag{2.2}$$

and

$$|g_{k+1}^T d_k| \leq \sigma_k |g_k^T d_k| \tag{2.3}$$

hold for all k, where

$$\rho_k = 1 - \delta \phi_k, \ \ \sigma_k = 1 + \delta \phi_k \tag{2.4}$$

and

$$\phi_k = \begin{cases} 0, & \text{if } \alpha_k = 0, \\ \frac{|(g_{k+1} - g_k)^T (x_{k+1} - x_k)|}{\|x_{k+1} - x_k\|^2}, & \text{if } \alpha_k \neq 0. \end{cases} \tag{2.5}$$

Proof The case of $\alpha_k = 0$ implies that $\rho_k = 1$, $\sigma_k = 1$ and $g_{k+1} = g_k$. Hence (2.2) and (2.3) hold. In the following, we consider the case of $\alpha_k \neq 0$. From (1.2) and (2.1) we have

$$\begin{aligned} g_{k+1}^T d_k &= g_k^T d_k + (g_{k+1} - g_k)^T d_k = g_k^T d_k + \alpha_k^{-1}(g_{k+1} - g_k)^T (x_{k+1} - x_k) \\ &\leq g_k^T d_k + \alpha_k^{-1}|(g_{k+1} - g_k)^T (x_{k+1} - x_k)| = g_k^T d_k + \alpha_k^{-1}\phi_k \|x_{k+1} - x_k\|^2 \\ &= g_k^T d_k + \alpha_k \phi_k \|d_k\|^2 = g_k^T d_k - \frac{\delta g_k^T d_k}{\|g_k\|^2 + \|d_k\|^2}\phi_k \|d_k\|^2 \\ &= (1 - \frac{\delta \phi_k \|d_k\|^2}{\|g_k\|^2 + \|d_k\|^2}) g_k^T d_k \leq (1 - \delta\phi_k) g_k^T d_k = \rho_k g_k^T d_k. \end{aligned}$$

Similarly, we have

$$\begin{aligned} |g_{k+1}^T d_k| &\leq (1 - \frac{\delta \phi_k \|d_k\|^2}{\|g_k\|^2 + \|d_k\|^2})|g_k^T d_k| \leq (1 + \frac{\delta \phi_k \|d_k\|^2}{\|g_k\|^2 + \|d_k\|^2})|g_k^T d_k| \\ &\leq (1 + \delta\phi_k)|g_k^T d_k| = \sigma_k |g_k^T d_k|. \end{aligned}$$

The proof is completed. □

Lemma 2.6 *Suppose that Assumption 2.1 holds. Then for all k there hold*

$$0 < 1 - \mu\delta \leq \rho_k \leq 1 + \mu\delta, \quad 1 \leq \sigma_k \leq 1 + \mu\delta; \tag{2.6}$$

and suppose Assumption 2.3 holds, then for all k it holds that

$$0 < 1 - \mu\delta \leq \rho_k \leq 1 - \lambda\delta. \tag{2.7}$$

Proof By (2.4) and (2.5) we have

$$\begin{aligned} 1 - \delta \frac{\|g_{k+1} - g_k\| \|x_{k+1} - x_k\|}{\|x_{k+1} - x_k\|^2} \leq \rho_k &= 1 - \delta \frac{|(g_{k+1} - g_k)^T (x_{k+1} - x_k)|}{\|x_{k+1} - x_k\|^2} \\ &\leq 1 + \delta \frac{\|g_{k+1} - g_k\| \|x_{k+1} - x_k\|}{\|x_{k+1} - x_k\|^2}, \end{aligned}$$

$$\sigma_k = 1 + \delta\phi_k = 1 + \delta \frac{|(g_{k+1} - g_k)^T (x_{k+1} - x_k)|}{\|x_{k+1} - x_k\|^2} \leq 1 + \delta \frac{\|g_{k+1} - g_k\| \|x_{k+1} - x_k\|}{\|x_{k+1} - x_k\|^2}.$$

From Assumption 2.1 and $0 < \delta < \min\{\frac{1}{\mu}, \frac{1}{\lambda}\}$, we have $0 < 1 - \mu\delta \leq \rho_k \leq 1 + \mu\delta$. Further, noting $\phi_k \geq 0$, we have $1 \leq \sigma_k \leq 1 + \mu\delta$. Therefore, (2.6) holds.

On the other hand, by Assumption 2.3 we get

$$\rho_k = 1 - \delta \frac{|(g_{k+1} - g_k)^T (x_{k+1} - x_k)|}{\|x_{k+1} - x_k\|^2} = 1 - \delta \frac{(g_{k+1} - g_k)^T (x_{k+1} - x_k)}{\|x_{k+1} - x_k\|^2} \leq 1 - \lambda\delta. \tag{2.8}$$

Since Assumption 2.3 implies Assumption 2.1 and $0 < \delta < \min\{\frac{1}{\mu}, \frac{1}{\lambda}\}$, then (2.7) holds. □

Lemma 2.7 *Suppose that Assumption 2.1 holds and that x_k is given by (1.2), (1.3) and (2.1). Then*

$$\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|g_k\|^2 + \|d_k\|^2} < +\infty. \tag{2.9}$$

Proof By the mean-value theorem, the Cauchy-Schwartz inequality, (1.2), (2.1) and Assumption 2.1, we obtain

$$f(x_{k+1}) - f(x_k) = \bar{g}^T (x_{k+1} - x_k) = g_k^T (x_{k+1} - x_k) + (\bar{g} - g_k)^T (x_{k+1} - x_k)$$

$$
\begin{aligned}
&\leq g_k^T(x_{k+1}-x_k)+\|\bar{g}-g_k\|\|x_{k+1}-x_k\| \leq g_k^T(x_{k+1}-x_k)+\mu\|x_{k+1}-x_k\|^2 \\
&= \alpha_k g_k^T d_k+\mu\alpha_k^2\|d_k\|^2 = \alpha_k g_k^T d_k-\frac{\mu\delta\alpha_k g_k^T d_k\|d_k\|^2}{\|g_k\|^2+\|d_k\|^2} = \alpha_k g_k^T d_k(1-\frac{\mu\delta\|d_k\|^2}{\|g_k\|^2+\|d_k\|^2}) \\
&\leq \alpha_k g_k^T d_k(1-\frac{\mu\delta\|d_k\|^2}{\|d_k\|^2}) = (1-\mu\delta)\alpha_k g_k^T d_k = -\delta(1-\mu\delta)\frac{(g_k^T d_k)^2}{\|g_k\|^2+\|d_k\|^2}, \qquad (2.10)
\end{aligned}
$$

where $\bar{g}=\nabla f(\bar{x})$ for some $\bar{x}\in[x_k,x_{k+1}]$. Evidently, (2.10) shows $f(x_{k+1})\leq f(x_k)$. It follows by Assumption 2.1 that $\lim_{k\to\infty} f(x_k)$ exists. Thus from (2.10) we obtain

$$
\frac{(g_k^T d_k)^2}{\|g_k\|^2+\|d_k\|^2} \leq \frac{1}{\delta(1-\mu\delta)}[f(x_k)-f(x_{k+1})].
$$

This completes the proof. □

Lemma 2.8 *Suppose that Assumption 2.1 holds and x_k is given by (1.2), (1.3) and (2.1). Then*

$$
\|x_{k+1}-x_k\| \to 0 \qquad (2.11)
$$

as $k\to\infty$. Further, for the FR, the PRP and the LS methods $\{\|d_k\|\}$ is uniformly bounded.

Proof From (1.2), (2.1) and Lemma 2.7, we have

$$
\begin{aligned}
\sum_{k\geq 1}\|x_{k+1}-x_k\|^2 &= \sum_{k\geq 1}\|\alpha_k d_k\|^2 = \sum_{k\geq 1}\frac{\delta^2(g_k^T d_k)^2}{(\|g_k\|^2+\|d_k\|^2)^2}\|d_k\|^2 \\
&\leq \delta^2\sum_{k\geq 1}\frac{(g_k^T d_k)^2(\|g_k\|^2+\|d_k\|^2)}{(\|g_k\|^2+\|d_k\|^2)^2} = \delta^2\sum_{k\geq 1}\frac{(g_k^T d_k)^2}{\|g_k\|^2+\|d_k\|^2} < +\infty.
\end{aligned}
$$

Hence (2.11) holds.

In addition, from (1.5), (1.6), (1.7) and (2.11), it is evident that

$$
|\beta_k^{FR}| = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \to 1, \qquad (2.12)
$$

$$
|\beta_k^{PRP}| = |\frac{g_k^T(g_k-g_{k-1})}{\|g_{k-1}\|^2}| \to 0 \qquad (2.13)
$$

and

$$
|\beta_k^{LS}| \leq \frac{|g_k^T(g_k-g_{k-1}|)}{\frac{1}{4}\|g_{k-1}\|^2} \to 0 \qquad (2.14)
$$

as $k\to\infty$. Noting Remark 2.2 and using

$$
\|d_k\| \leq \|g_k\|+|\beta_k|\|d_{k-1}\|, \qquad (2.15)
$$

we easily know that $\{\|d_k\|\}$ is uniformly bounded for the FR, the PRP and the LS methods. This completes the proof. □

Lemma 2.9 *(1) Suppose that Assumption 2.1 holds and x_k is given by (1.2), (1.3) and (2.1). Then the CD and the FR methods satisfy*

$$
(g_k^T d_k)^2 > \|g_k\|^4, \qquad (2.16)
$$

and for large k the PRP and the LS methods satisfy

$$(g_k^T d_k)^2 \geq \frac{\|g_k\|^4}{4}. \tag{2.17}$$

(2) Suppose that Assumption 2.3 holds and x_k is given by (1.2), (1.3) and (2.1). Then the DY method satisfies

$$(g_k^T d_k)^2 \geq \|g_k\|^4. \tag{2.18}$$

Proof (1) First, for the CD method, from (1.4), (2.2) and (2.6) we have

$$\begin{aligned} g_k^T d_k &= g_k^T(-g_k + \frac{\|g_k\|^2}{-g_{k-1}^T d_{k-1}} d_{k-1}) = -\|g_k\|^2(1 + \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}}) \\ &\leq -\|g_k\|^2(1 + \frac{\rho_{k-1} g_{k-1}^T d_{k-1}}{g_{k-1}^T d_{k-1}}) = -(1+\rho_{k-1})\|g_k\|^2 < -\|g_k\|^2. \end{aligned}$$

Therefore (2.16) holds.

Now we consider the FR method. It follows from (1.5), (2.2) that

$$g_k^T d_k = g_k^T(-g_k + \frac{\|g_k\|^2}{\|g_{k-1}\|^2} d_{k-1}) = (\frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} - 1)\|g_k\|^2 \leq (\frac{\rho_{k-1} g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} - 1)\|g_k\|^2.$$

By the recursive principle we obtain the following inequalities

$$\begin{aligned} \frac{g_k^T d_k}{\|g_k\|^2} &\leq \frac{\rho_{k-1} g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} - 1 \leq \rho_{k-1}(\frac{\rho_{k-2} g_{k-2}^T d_{k-2}}{\|g_{k-2}\|^2} - 1) - 1 \leq \cdots \\ &\leq \rho_{k-1}\rho_{k-2}\cdots\rho_1 \frac{g_1^T d_1}{\|g_1\|^2} - \rho_{k-1}\rho_{k-2}\cdots\rho_2 - \cdots - \rho_{k-1}\rho_{k-2} - \rho_{k-1} - 1. \end{aligned}$$

Noting $\frac{g_1^T d_1}{\|g_1\|^2} = -1$, then

$$\frac{g_k^T d_k}{\|g_k\|^2} \leq -\rho_{k-1}\rho_{k-2}\cdots\rho_1 - \rho_{k-1}\rho_{k-2}\cdots\rho_2 - \cdots - \rho_{k-1}\rho_{k-2} - \rho_{k-1} - 1.$$

By (2.6) we have $\frac{g_k^T d_k}{\|g_k\|^2} < -1$, i.e., (2.16) holds.

Next we prove that (2.17) is valid for the PRP method. From Lemma 2.8 and (2.13), we conclude that for large k, $\|\beta_k^{PRP} d_{k-1}\| \leq \frac{\|g_k\|}{2}$, which leads to

$$|g_k^T d_k| = |g_k^T(-g_k + \beta_k^{PRP} d_{k-1})| \geq \|g_k\|^2 - |\beta_k^{PRP}|\|g_k\|\|d_{k-1}\| \geq \frac{\|g_k\|^2}{2}.$$

Then it follows that (2.17) holds.

Finally, as to the LS method, from (1.7), the Cauchy-Schwartz inequality, (2.3) and (2.6), we obtain

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 - \frac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} \leq -\|g_k\|^2 + \frac{\|g_k\|\|g_k - g_{k-1}\|}{|g_{k-1}^T d_{k-1}|} \cdot |g_k^T d_{k-1}| \\ &\leq -\|g_k\|^2 + \frac{\|g_k\|\|g_k - g_{k-1}\|}{|g_{k-1}^T d_{k-1}|} \cdot \sigma_{k-1}|g_{k-1}^T d_{k-1}| = -\|g_k\|^2 + \sigma_{k-1}\|g_k\|\|g_k - g_{k-1}\| \\ &\leq -\|g_k\|^2 + (1+\mu\delta)\|g_k\|\|g_k - g_{k-1}\|. \end{aligned}$$

By (2.11) we have $g_k - g_{k-1} \to 0$ as $k \to \infty$, then $\|g_k - g_{k-1}\| \le \frac{\|g_k\|}{2(1+\mu\delta)}$. Hence

$$g_k^T d_k \le -\|g_k\|^2 + (1+\mu\delta)\|g_k\| \cdot \frac{\|g_k\|}{2(1+\mu\delta)} = (-1+\frac{1}{2})\|g_k\|^2 = -\frac{1}{2}\|g_k\|^2,$$

i.e., (2.17) holds.

(2) For the DY method, from (2.2) we know $g_k^T d_{k-1} < 0$, and under Assumption 2.3 it holds that

$$0 < d_{k-1}^T(g_k - g_{k-1}) < -g_{k-1}^T d_{k-1},$$

which, together with (1.8) and (2.7), leads to

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \frac{\|g_k\|^2}{d_{k-1}^T(g_k - g_{k-1})} g_k^T d_{k-1} < -\|g_k\|^2 + \frac{\|g_k\|^2}{-g_{k-1}^T d_{k-1}} g_k^T d_{k-1} \\ &\le -\|g_k\|^2 + \frac{\rho_{k-1} g_{k-1}^T d_{k-1}}{-g_{k-1}^T d_{k-1}} \|g_k\|^2 = -(1+\rho_{k-1})\|g_k\|^2 \le -\|g_k\|^2. \end{aligned}$$

This means that (2.18) holds. □

Lemma 2.10 *(1) Suppose that Assumption 2.1 holds and x_k is given by (1.2), (1.3) and (2.1). Then the CD and the FR methods satisfy*

$$\sum_{k\ge 1} \frac{\|g_k\|^4}{\|g_k\|^2 + \|d_k\|^2} < +\infty, \tag{2.19}$$

and for large k (2.19) is also true for the PRP and the LS methods.

(2) Suppose that Assumption 2.3 holds and x_k is given by (1.2), (1.3) and (2.1). Then the same conclusion holds for the DY method.

Proof (2.19) is evident from Lemmas 2.7 and 2.9. □

Lemma 2.11 *Suppose that Assumption 2.1 holds and x_k is given by (1.2), (1.3) and (2.1). Then for the CD method we have*

$$\frac{\|g_k\|^2 + \|d_k\|^2}{\|g_k\|^4} \le \frac{\|g_{k-1}\|^2 + \|d_{k-1}\|^2}{\|g_{k-1}\|^4} + \frac{\Omega}{\|g_k\|^2}, \tag{2.20}$$

where Ω is a positive constant.

Proof From Lemmas 2.5, 2.6 and 2.9, we have

$$\begin{aligned} \|d_k\|^2 &= \|-g_k + \beta_k^{CD} d_{k-1}\|^2 = \|-g_k + \frac{\|g_k\|^2}{-g_{k-1}^T d_{k-1}} d_{k-1}\|^2 \\ &= \|g_k\|^2 + \frac{\|g_k\|^4}{(g_{k-1}^T d_{k-1})^2}\|d_{k-1}\|^2 + \frac{2\|g_k\|^2}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} \\ &\le \|g_k\|^2 + \frac{\|g_k\|^4}{(g_{k-1}^T d_{k-1})^2}\|d_{k-1}\|^2 + \frac{2\sigma_{k-1}|g_{k-1}^T d_{k-1}|}{|g_{k-1}^T d_{k-1}|}\|g_k\|^2 \\ &= (1+2\sigma_{k-1})\|g_k\|^2 + \frac{\|g_k\|^4}{(g_{k-1}^T d_{k-1})^2}\|d_{k-1}\|^2 \\ &\le [1+2(1+\mu\delta)]\|g_k\|^2 + \frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_{k-1}\|^2. \end{aligned}$$

Then

$$\|g_k\|^2+\|d_k\|^2 \le [2+2(1+\mu\delta)]\|g_k\|^2+\frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_{k-1}\|^2 = 2(2+\mu\delta)\|g_k\|^2+\frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_{k-1}\|^2.$$

Let $\Omega := 2(2+\mu\delta)$. Then

$$\frac{\|g_k\|^2+\|d_k\|^2}{\|g_k\|^4} \le \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4}+\frac{\Omega}{\|g_k\|^2} \le \frac{\|g_{k-1}\|^2+\|d_{k-1}\|^2}{\|g_{k-1}\|^4}+\frac{\Omega}{\|g_k\|^2}.$$

The proof is completed. □

Theorem 2.12 *Suppose that Assumption 2.1 holds and x_k is given by (1.2), (1.3) and (2.1). Then the CD method will generate a sequence $\{x_k\}$ such that*

$$\liminf_{k\to\infty}\|g_k\| = 0. \tag{2.21}$$

Likewise, (2.21) also holds for the FR, the PRP and the LS methods under Assumption 2.1.

Proof If $\liminf_{k\to\infty}\|g_k\| \ne 0$, then there exists $\gamma > 0$ such that $\|g_k\| \ge \gamma$ for all k. We first consider the CD method. From Lemma 2.11 we have

$$\begin{aligned}\frac{\|g_k\|^2+\|d_k\|^2}{\|g_k\|^4} &\le \frac{\|g_{k-1}\|^2+\|d_{k-1}\|^2}{\|g_{k-1}\|^4}+\frac{\Omega}{\gamma^2} \le \frac{\|g_{k-2}\|^2+\|d_{k-2}\|^2}{\|g_{k-2}\|^4}+\frac{2\Omega}{\gamma^2}\\ &\le \cdots \le \frac{\|g_1\|^2+\|d_1\|^2}{\|g_1\|^4}+\frac{(k-1)\Omega}{\gamma^2} = \frac{2}{\|g_1\|^2}+\frac{(k-1)\Omega}{\gamma^2} \le \frac{k\Omega-\Omega+2}{\gamma^2}.\end{aligned}$$

Let $a=\frac{\Omega}{\gamma^2}$, $b=\frac{2}{\gamma^2}$. Then $\frac{\|g_k\|^2+\|d_k\|^2}{\|g_k\|^4} \le ka-a+b$, further it follows that

$$\frac{\|g_k\|^4}{\|g_k\|^2+\|d_k\|^2} \ge \frac{1}{ka-a+b}.$$

Therefore,

$$\sum_{k\ge 1}\frac{\|g_k\|^4}{\|g_k\|^2+\|d_k\|^2} = +\infty,$$

which is contradictory to Lemma 2.10. Then (2.21) holds and the proof for the CD method is completed.

Next is for the FR method. From Lemma 2.8, Remark 2.2 and (2.16), we have

$$\frac{(g_k^Td_k)^2}{\|g_k\|^2(\|g_k\|^2+\|d_k\|^2)} > \frac{\|g_k\|^2}{\|g_k\|^2+\|d_k\|^2}.$$

Combining $\|g_k\| \ge \gamma$, then there exists $\epsilon > 0$ such that

$$\frac{(g_k^Td_k)^2}{\|g_k\|^2(\|g_k\|^2+\|d_k\|^2)} \ge \epsilon,$$

i.e.,

$$\sum_{k\ge 1}\frac{(g_k^Td_k)^2}{\|g_k\|^2+\|d_k\|^2} = \sum_{k\ge 1}\|g_k\|^2\cdot\frac{(g_k^Td_k)^2}{\|g_k\|^2(\|g_k\|^2+\|d_k\|^2)} \ge \epsilon\sum_{k\ge 1}\|g_k\|^2 \ge \epsilon\sum_{k\ge 1}\gamma^2 = +\infty. \tag{2.22}$$

This contradicts Lemma 2.7. Hence the proof for the FR method is completed.

The proofs of the PRP and the LS methods are similar to that of the FR method, and are omitted here.

Hence the whole proof is completed. □

Theorem 2.13 *Suppose that Assumption 2.3 holds and x_k is given by (1.2), (1.3) and (2.1). Then the DY method will generate a sequence $\{x_k\}$ such that (2.21) holds.*

Proof By $\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T(g_k-g_{k-1})} = \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}$, we have

$$\|g_k\|^2 + \|d_k\|^2 + 2g_k^T d_k = (\beta_k^{DY})^2\|d_{k-1}\|^2 = \frac{(g_k^T d_k)^2}{(g_{k-1}^T d_{k-1})^2}\|d_{k-1}\|^2. \tag{2.23}$$

Then

$$\frac{\|g_k\|^2 + \|d_k\|^2}{(g_k^T d_k)^2} + \frac{2}{g_k^T d_k} = \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} \le \frac{\|g_{k-1}\|^2 + \|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2}.$$

Let $w_k = \frac{\|g_k\|^2+\|d_k\|^2}{(g_k^T d_k)^2}$, then $w_k \le w_{k-1} - \frac{2}{g_k^T d_k}$. From Lemma 2.9, $-g_k^T d_k \ge \|g_k\|^2$, i.e., $\frac{1}{-g_k^T d_k} \le \frac{1}{\|g_k\|^2}$. Thus we get

$$w_k \le w_{k-1} + \frac{2}{\|g_k\|^2}.$$

Further we have

$$\begin{aligned} w_k &\le w_{k-1} + \frac{2}{\|g_k\|^2} \le w_{k-2} + \frac{2}{\|g_{k-1}\|^2} + \frac{2}{\|g_k\|^2} \\ &\le w_{k-3} + \frac{2}{\|g_{k-2}\|^2} + \frac{2}{\|g_{k-1}\|^2} + \frac{2}{\|g_k\|^2} \\ &\le w_1 + \frac{2}{\|g_2\|^2} + \frac{2}{\|g_3\|^2} + \cdots + \frac{2}{\|g_{k-1}\|^2} + \frac{2}{\|g_k\|^2}. \end{aligned}$$

Note $w_1 = \frac{\|g_1\|^2+\|d_1\|^2}{\|g_1\|^4} = \frac{2}{\|g_1\|^2}$, then

$$w_k \le \sum_{i=1}^{k} \frac{2}{\|g_i\|^2}.$$

From $\|g_k\| \ge \gamma$ we have

$$w_k \le \sum_{i=1}^{k} \frac{2}{\gamma^2} = \frac{2k}{\gamma^2},$$

that is $\frac{1}{w_k} \ge \frac{\gamma^2}{2k}$. Then $\frac{(g_k^T d_k)^2}{\|g_k\|^2+\|d_k\|^2} \ge \frac{\gamma^2}{2k}$, thus

$$\sum_{k\ge 1} \frac{(g_k^T d_k)^2}{\|g_k\|^2 + \|d_k\|^2} \ge \frac{\gamma^2}{2}\sum_{k\ge 1}\frac{1}{k} = +\infty,$$

which contradicts Lemma 2.10. Therefore (2.21) holds. □

3. Final remarks

The new steplength formula for α_k in this paper guarantees the global convergence of the five well-known conjugate gradient methods, which indicates that proposing diverse forms of the steplength formula is a feasible and meaningful topic. However, the drawback of the steplength formula proposed here by us is that it cannot establish the global convergence of the HS conjugate gradient methods. Hence if one can present another steplength formula for α_k, which can guarantee that the six well-known conjugate gradient methods are all globally convergent, then it is a significant work. This is a topic needing to research further.

References

[1] Yuhong DAI, Yaxiang YUAN. *Nonlinear Conjugate Gradient Methods*. Shanghai Science and Technology Press, Shanghai, 2000. (in Chinese)

[2] J. NOCEDAL. *Theory of Algorithms for Unconstrained Optimization*. Cambridge Univ. Press, Cambridge, 1992.

[3] Zhenjun SHI. *Convergence of line search methods for unconstrained optimization*. Appl. Math. Comput., 2004, **157**(2): 393–405.

[4] L. C. W. DIXON. *Conjugate directions without line searches*. J. Inst. Math. Appl., 1973, **11**: 317–328.

[5] Jie SUN, Jiapu ZHANG. *Global convergence of conjugate gradient methods without line search*. Ann. Oper. Res., 2001, **103**: 161–173.

[6] Xiongda CHEN, Jie SUN. *Global convergence of a two-parameter family of conjugate gradient methods without line search*. J. Comput. Appl. Math., 2002, **146**(1): 37–45.

[7] Zhenjun SHI, Jie SHEN. *Convergence of descent method without line search*. Appl. Math. Comput., 2005, **167**(1): 94–107.

[8] Guangming ZHOU. *A descent algorithm without line search for unconstrained optimization*. Appl. Math. Comput., 2009, **215**(7): 2528–2533.