METHODOLOGY

CrossMark

New hybrid conjugate gradient methods with the generalized Wolfe line search

Xiao Xu[*] and Fan-yu Kong

*Correspondence:
xinkexuxiao@163.com
College of science,
Nanjing University
of Science and Technology,
Nanjing 210094, Jiangsu,
China

Abstract

The conjugate gradient method was an efficient technique for solving the unconstrained optimization problem. In this paper, we made a linear combination with parameters β_k of the DY method and the HS method, and putted forward the hybrid method of DY and HS. We also proposed the hybrid of FR and PRP by the same mean. Additionally, to present the two hybrid methods, we promoted the Wolfe line search respectively to compute the step size a_k of the two hybrid methods. With the new Wolfe line search, the two hybrid methods had descent property and global convergence property of the two hybrid methods that can also be proved.

Keywords: Conjugate gradient, Unconstrained optimization, Line search, Descent property, Global convergence

Background

The conjugate gradient method

The conjugate gradient method is a very important and efficient technique for solving large scale minimization problems, due to it can complete with lower storage and simple computation (Birgin and Martinez 2001).

Consider an unconstrained minimization problem.

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

where R^n denotes an n-dimensional Euclidean space and $f : R^n \longrightarrow R^1$ is a continuously differentiable function. We denote its gradient $\nabla f(x_k)$ by g_k. We are concerned with the conjugate gradient methods for solving (1). The iterative process of the conjugate gradient method is given by.

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 1, 2, \ldots \tag{2}$$

where x_1 is the initial point and $x_k \in R^n$ is the k-th approximation to a solution, α_k is a positive step size, and $d_k \in R^n$ is a search direction defined by the following:

$$d_k = \begin{cases} -g_k, & k = 1 \\ -g_k + \beta_k d_{k-1}, & k \geq 2 \end{cases} \tag{3}$$

where β_k is a parameter.

Selection of step size

For selection of step size α_k in the iteration formula, it is generally determined by the direct method (0.618 method), the analytical method (successive difference approximation method), exact line search and inexact line search methods and so on. What we usually used in nonlinear conjugate gradient method is the exact line search and inexact line search. In this paper, we mainly use the Wolfe line search as following

Given $\mu \in \left(0, \frac{1}{2}\right)$, $\sigma \in (\mu, 1)$ and $\alpha_k > 0$ which satisfies

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\mu \alpha_k \nabla f(x_k)^{\mathrm{T}} d_k \tag{4}$$

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\sigma \alpha_k \nabla f(x_k)^{\mathrm{T}} d_k \tag{5}$$

Generally, the smaller the value of σ, the line search will be more precise, but the greater the amount of calculation. So we set $\mu = 0.1$, $\sigma \in [0.6, 0.8]$. In order to facilitate analysis, the inexact line search should convert to the exact line search, but in the formula (5), when $\sigma \to 0$, the search is not the exact line search. For this purpose, Fletcher proposed the stronger condition to take the place of (5):

$$\left| g(x_k + \alpha_k d_k)^{\mathrm{T}} d_k \right| \leq \sigma \left| g_k^{\mathrm{T}} d_k \right| \tag{6}$$

We note (4) and (6) as the strong Wolfe line search.

Selection of search direction

The search direction d_k is generally required to satisfy

$$d_k^{\mathrm{T}} g_k < 0 \tag{7}$$

which guarantees that d_k is a descent direction of $f(x)$ at x_k (Yuan 1993). In order to maintain the global convergence property, we sometimes require d_k to satisfy a sufficient descent condition

$$g_k^T d_k \leq -c \|g_k\|^2$$

where $c > 0$ is a constant.

Selection of the parameter β_k

Different parameters β_k of different conjugate gradient methods as follows (see (i) Fletcher and Reeves 1964, (ii) Polyak 1969; Polak et al. 1968, (iii) Hestenes and Stiefel 1952, (iv) Vincent 1983, (v), (vi) Dai and Yuan 2000).

$$\text{(i) } \beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \qquad \text{(ii) } \beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \quad \text{(iii) } \beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}$$

$$\text{(iv) } \beta_k^{CD} = -\frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}}, \qquad \text{(v) } \beta_k^{LS} = \frac{g_k^T y_{k-1}}{-d_{k-1}^T g_{k-1}}, \quad \text{(vi) } \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}$$

For positive quadratic function, if we adopt the exact line search, several conjugate gradient methods in the above are equivalent, which implies that the conjugate gradient directions generated by several methods are equivalent. In practical application, the FR method and PRP method are the most common methods.

The FR method is the earliest nonlinear conjugate gradient method, under the exact line search, Powell (1977) pointed out the FR method could continuously produce small steps and have the global efficiency property. Zoutendijk (1970) proved the FR method are always convergence for general non-convex functions; Under the inexact line search, Al-Baali (1985); Liu et al. (1995) proved when $\sigma \leq \frac{1}{2}$, the FR method has the global convergence property with strong Wolfe line search. But the shortcoming is the FR method has slow convergence speed.

The PRP method is considered to be the best conjugate gradient method in numerical experiment results at present. Once produce a small step size, the next search direction generated by the PRP algorithm will approach the negative gradient direction automatically. It's good to avoid the shortcoming that the FR method could continuously produce small steps. Under the exact line search, the PRP method has global convergence property for uniformly convex functions, but it is not established for general non-convex functions (Powell 1984). Under the inexact line search, if search direction is descent direction and the objective function is uniformly convex function, Yuan (1995) proved the PRP method has the global convergence property with Wolfe line search.

The characteristics of the HS method are similar to the PRP method, but compared with PRP method, an important feature of the HS method is that no matter whether the line precision is exist, the conjugate relation $d_{k+1}^T y_k = 0$ is always formed.

The CD method is very similar with the FR method. Under the exact line search, $\beta_k^{CD} = \beta_k^{FR}$; Under the inexact line search, an important feature of the CD algorithm is that as long as the parameter $\sigma < 1$ in strong Wolfe line search, the CD method would generate a descent search direction in each iteration, but its global convergence property is not good.

Under the exact line search, the LS method is equivalent to the PRP method.

The DY method can always generate a descent direction in each iteration with Wolfe line search, but the disadvantage of this method is the bad numerical experiment results.

In the paper, we learn and analyze the above methods, then proposed two hybrid nonlinear conjugate gradient method, namely, the hybrid method of DY and HS and the hybrid method of FR and PRP. We also have a research on both each methods.

Hybrid conjugate gradient method of DY and HS

Preliminaries of the new conjugate gradient method

To the hybrid conjugate gradient method of DY and HS, we promote the Wolfe line search in our paper. The standard Wolfe line search (5) is revised as the following:

If $d_k^T g_{k+1} \geq 0$, then

$$\sigma_1 d_k^T g_k \leq d_k^T g(x_k + \alpha_k d_k) \leq -\sigma_2 d_k^T g_k \tag{8}$$

where $-\sigma_2 d_k^T g_k < -\sigma_2 d_k^T (g_k - g_{k+1})$

If $d_k^T g_{k+1} < 0$, then

$$\sigma_1 d_k^T g_k \leq d_k^T g(x_k + \alpha_k d_k) \leq -\sigma_2 d_k^T (g_k - g_{k+1}) \tag{9}$$

where $-\sigma_2 d_k^T (g_k - g_{k+1}) < -\sigma_2 d_k^T g_k$.

The parameters β_k of the hybrid conjugate gradient method is formulized as

$$\beta_k^{(1)} = \begin{cases} a_1\beta_k^{DY} + a2\beta_k^{HS} & if \ \left\|gk\right\|^2 > \left|g_k^T g_{k-1}\right| \\ 0 & else \end{cases}$$

i.e.

$$\beta_k^{(1)} = \begin{cases} \frac{a_1\|g_k\|^2}{d_{k-1}^T(g_k-g_{k-1})} + \frac{a_2 g_k^T(g_k-g_{k-1})}{d_{k-1}^T(g_k-g_{k-1})} & if \ \left\|g_k\right\|^2 > \left|g_k^T g_{k-1}\right| \\ 0 & else \end{cases} \tag{10}$$

where a_1, a_2 are nonnegative numbers and at least one are not zero, and that they are required to satisfy

$$0 < a_1 + 2a_2 < \frac{1}{1+\sigma_2} < 1 \tag{11}$$

A new Wolfe line search is proposed, which makes the hybrid conjugate gradient method keep the global convergence property and the descent property in this paper.

When $\beta_k = 0$, the new hybrid method will degenerate into the steepest descent method.

Algorithm (*The new hybrid method with the new Wolfe line search*)
Step 1: Choose an initial point $x_1 \in R^n$. Give the precision value $\varepsilon > 0$. Compute g_1, if $\|g_1\| < \varepsilon$, then stop, x_1 is the optimal point; otherwise go to Step 2
Step 2: Set $d_1 = -g_1$. Let $k = 1$
Step 3: Set $x_{k+1} = x_k + \alpha_k d_k$, α_k is defined by the new generalized Wolfe line search (4) (8) (9)
Step 4: Compute g_{k+1}, if $\|g_{k+1}\| < \varepsilon$, then stop; otherwise go to Step 5
Step 5: Set $k := k + 1$; Set $d_k = -g_k + \beta_k^1 d_{k-1}$, $\beta_k^{(1)}$ is defined by the formula (10), then go to Step 3

The descent property
Assumption H
H1: The objective function $f(x)$ is a continuously differentiable function. The level set $L_1 = \{x \in R^n : f(x) \le f(x_1)\}$ at x_1 is bounded (x_1 is the initial point); namely, there exists a constant $a > 0$ such that

$$\|x\| \le a \quad \text{for all} \quad x \in L_1$$

H2: In any neighborhood N of L_1, f is continuously differentiable, and its gradient $g(x)$ is Lipschitz continuous with Lipschitz constant $L > 0$; i.e.,

$$\left\|g(x) - g(y)\right\| \le \left\|Lx - y\right\| \ \text{for all } x, y \in N$$

Lemma 1 *Suppose that the objective function satisfies Assumption H. Consider the method (2), (3), where α_k satisfies the new Wolfe line search (4), (8), (9) and $\beta_k^{(1)}$ satisfies the formula (10), then the following holds:*

$g_k^T d_k < 0$ *for all k*

Proof For $k = 1$, we have $g_1^T d_1 = g_1^T g_1 = - \|g_1\|^2 < 0$ according to $d_1 = - g_1$.

For $k > 1$, suppose that $g_k^T d_k < 0$ holds at the k-th step, then we prove this inequality also holds at the k + 1-th step.

From $\|g_k\|^2 > |g_k^T g_{k-1}|$ and (11), we have that

$$
\begin{aligned}
g_{k+1}^T d_{k+1} &= -\|g_{k+1}\|^2 + \beta_{k+1}^{(1)} g_{k+1}^T d_k \\
&= -\|g_{k+1}\|^2 + \left[\frac{a_1 \|g_{k+1}\|^2}{d_k^T (g_{k+1} - g_k)} + \frac{a_2 g_{k+1}^T (g_{k+1} - g_k)}{d_k^T (g_{k+1} - g_k)} \right] \cdot g_{k+1}^T d_k \\
&\leq -\|g_{k+1}\|^2 + \frac{(a_1 + a_2)\|g_{k+1}\|^2 - a_2 g_{k+1}^T g_k}{d_k^T (g_{k+1} - g_k)} (-\sigma_2 d_k^T (g_k - g_{k+1})) \\
&\leq -\|g_{k+1}\|^2 + \frac{(a_1 + 2a_2)\|g_{k+1}\|^2}{d_k^T (g_{k+1} - g_k)} \cdot (-\sigma_2 d_k^T (g_k - g_{k+1})) \\
&= [(a_1 + 2a_2)\sigma_2 - 1]\|g_{k+1}\|^2
\end{aligned}
$$

Then, by $0 < a_1 + 2a_2 < \frac{1}{1+\sigma_2} < 1$ and $0 \leq \sigma_2 < 1$, we have $g_{k+1}^T d_{k+1} < 0$.

Therefore, according to the mathematical induction, Lemma 1 is proved, which implies that the new hybrid method has the descent property.

Global convergence

Lemma 2 *Assume that (H) hold, Consider the method* (2), (3), *where d_k is a descent direction and α_k satisfies the new Wolfe line search* (4), (8), (9), *$\beta_k^{(1)}$ satisfies the formula* (10). *Then we have that* $\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty$

Proof By Lemma 1, we have $g_k^T d_k < 0$, so the sequence $\{f(x_k)\}$ is bounded and has monotone descending property, which implies that $\{f(x_k)\}$ is a convergent sequence. From (8), (9), we have

$$
0 < (\sigma_1 - 1)g_k^T d_k \leq d_k^T (g_{k+1} - g_k) \leq -(1 + \sigma_2)g_k^T d_k \tag{12}
$$

From the Assumption H2, we have

$$
d_k^T (g_{k+1} - g_k) \leq L\alpha_k \|d_k\|^2 \tag{13}
$$

From (12), (13), we have

$$
\alpha_k \geq \frac{g_k^T d_k (\sigma_1 - 1)}{L \|d_k\|^2} \tag{14}
$$

From (8), (14), we have

$$
f_k - f_{k+1} \geq -\mu\alpha_k g_k^T d_k \geq \frac{-\mu(g_k^T d_k)^2(\sigma_1 - 1)}{L\|d_k\|^2} = \frac{\mu(1 - \sigma_1)}{L} \cdot \frac{(g_k^T d_k)^2}{\|d_k\|^2}
$$

By summing this formula, we have

$$\sum_{k \geq 1} (f_k - f_{k+1}) = f_1 - \lim_{k \to \infty} f_k \geq \sum_{k \geq 1} \frac{\mu(1 - \sigma_1)}{L} \cdot \frac{(g_k^T d_k)^2}{\|d_k\|^2}$$

Then

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

The proof is completed.

Theorem 3 *Suppose that Assumption H1 and H2 are satisfied. Consider the method* (2), (3), *where α_k satisfies the new Wolfe line search* (4), (8), (9), $\beta_k^{(1)}$ *satisfies the formula* (10). *Then the following holds:*

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

Proof If $\lim_{k \to \infty} \inf \|g_k\| = 0$ is not true, there exists a constant $c > 0$ such that

$$\|g_k\|^2 > c, \quad \text{for all k} \tag{15}$$

Therefore, from $d_k = -g_k + \beta_k^{(1)} d_{k-1}$, multiplying with g_k on the both sides, we have $g_k^T d_k = -\|g_k\|^2 + \beta_k^{(1)} g_k^T d_{k-1}$. Thus (8) and (9) yield

$$-\sigma_2 < \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} < \sigma_1$$

Then

$$\frac{\sigma_1}{\sigma_{1-1}} < \frac{g_k^T d_{k-1}}{d_{k-1}^T (g_k - g_{k-1})} < \frac{\sigma_2}{\sigma_2 + 1} \tag{16}$$

On the other hand, from $\|g_k\|^2 > |g_k^T g_{k-1}|$, we have

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

$$= 1 - \left[\frac{a_1 \|g_k\|^2}{d_{k-1}^T (g_k - g_{k-1})} + \frac{a_2 g_k^T (g_k - g_{k-1})}{d_{k-1}^T (g_k - g_{k-1})} \right] \cdot \frac{g_k^T d_{k-1}}{\|g_k\|^2}$$

If $g_k^T d_{k-1} \geq 0$, then

$$\frac{-g_k^T d_k}{\|g_k\|^2} \geq 1 - \frac{(a_1 + 2a_2) \|g_k\|^2}{\|g_k\|^2} \cdot \frac{g_k^T d_{k-1}}{d_{k-1}^T (g_k - g_{k-1})}$$

$$= 1 - (a_1 + 2a_2) \cdot \frac{g_k^T d_{k-1}}{d_{k-1}^T (g_k - g_{k-1})}$$

If $g_k^T d_{k-1} < 0$, then

$$\frac{-g_k^T d_k}{\left\|g_k\right\|^2} \leq 1 - \frac{(a_1 + 2a_2)\left\|g_k\right\|^2}{\left\|g_k\right\|^2} \cdot \frac{g_k^T d_{k-1}}{d_{k-1}^T(g_k - g_{k-1})}$$

$$= 1 - (a_1 + 2a_2) \cdot \frac{g_k^T d_{k-1}}{d_{k-1}^T(g_k - g_{k-1})}$$

Therefore, by (16), we have

$$1 - (a_1 + 2a_2)\frac{\sigma_2}{\sigma_2 + 1} \leq \frac{-g_k^T d_k}{\left\|g_k\right\|^2} \leq 1 - (a_1 + 2a_2) \cdot \frac{\sigma_1}{\sigma_1 - 1} \tag{17}$$

And, by squaring both sides of $d_k + g_k = \beta_k^{(1)} d_{k-1}$, we have

$$\left\|d_k\right\|^2 = -\left\|g_k\right\|^2 - 2g_k^T d_k + (\beta_k^{(1)})^2\left\|d_{k-1}\right\|^2$$

Then by multiplying with $(g_k^T d_k)^2$ on the both sides, we have

$$\frac{\left\|d_k\right\|^2}{(g_k^T d_k)^2} = \frac{(\beta_k^{(1)})^2\left\|d_{k-1}\right\|^2}{(g_k^T d_k)^2} - \frac{2}{g_k^T d_k} - \frac{\left\|g_k\right\|^2}{(g_k^T d_k)^2}$$

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

$$\leq \frac{(\beta_k^{(1)})^2\left\|d_{k-1}\right\|^2}{(g_k^T d_k)^2} + \frac{1}{\left\|g_k\right\|^2}$$

However, because of

$$(\beta_k^{(1)})^2 = \left[\frac{a_1\left\|g_k\right\|^2}{d_{k-1}^T(g_k - g_{k-1})} + \frac{a_2 g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})}\right]^2 \leq \frac{(a_1 + 2a_2)^2\left\|g_k\right\|^4}{(d_{k-1}^T(g_k - g_{k-1}))^2},$$

and thus from (17), we have

$$\frac{\left\|d_k\right\|^2}{(g_k^T d_k)^2} \leq \frac{(a_1 + 2a_2)^2\left\|g_k\right\|^4}{(d_k^T g_k)^2} \cdot \frac{\left\|d_{k-1}\right\|^2}{(d_{k-1}^T(g_k - g_{k-1}))^2} + \frac{1}{\left\|g_k\right\|^2}$$

$$\leq \frac{(a_1 + 2a_2)^2}{[1 - (a_1 + 2a_2)\frac{\sigma_2}{\sigma_2+1}]^2} \cdot \frac{\left\|d_{k-1}\right\|^2}{(d_{k-1}^T(g_k - g_{k-1}))^2} + \frac{1}{\left\|g_k\right\|^2}$$

$$\triangleq m \cdot \frac{\left\|d_{k-1}\right\|^2}{(d_{k-1}^T(g_k - g_{k-1}))^2} + \frac{1}{\left\|g_k\right\|^2}(m > 0)$$

By using the recurrence method on the left-hand side of the above inequality, there exists a constant $T > 0$ such that

$$\frac{\left\|d_k\right\|^2}{(g_k^T d_k)^2} \leq \sum_{i=1}^{k} \frac{m^{k-i}}{\left\|g_k\right\|^2} \leq \frac{1}{c}\sum_{i=1}^{k} m^{k-i} \leq T$$

By summing the above inequality, we have $\sum_{k\geq1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \sum_{k\geq1} 1/T = +\infty$, which contradicts Lemma 2. So the proof is complete.

Hybrid conjugate gradient method of FR and PRP

As the hybrid conjugate gradient method of DY and HS, we promote the Wolfe line search as well in our paper. The standard Wolfe line search (5) is revised as the following:

If $-\|g_k\|^2 \leq d_k^T g_k^T$ then

$$\sigma_1 d_k^T g_k^T \leq d_k^T g_{k+1}^T \leq -\sigma_2 d_k^T g_k^T \tag{18}$$

If $-\|g_k\|^2 > d_k^T g_k^T$ then

$$-\sigma_1 \|g_k\|^2 \leq d_k^T g_{k+1}^T \leq \sigma_2 \|g_k\|^2 \tag{19}$$

The parameters β_k of the hybrid conjugate gradient method of FR and PRP is formulized as

$$\beta_k^{(2)} = \begin{cases} a_1 \beta_k^{FR} + a_2 \beta_k^{PRP} & \text{if } \|g_k\|^2 > |g_k^T g_{k-1}| \\ 0 & \text{else} \end{cases} \tag{20}$$

where a_1, a_2 are nonnegative parameters and also at least one is not zero, and they are required to satisfy

$$0 < a_1 + 2a_2 < \frac{1}{1+\sigma_2} < 1 \tag{21}$$

The new Wolfe line search also can make the hybrid conjugate gradient method of FR and PRP keep the global convergence property and the descent property in this paper.

Both the properties can be proofed by the same process as the hybrid method of DY and HS.

Numerical experiments

In this section, we report some preliminary numerical experiments. We chose 15 test problems (problems 21–35) with the dimension n = 10,000 and initial points from the literature (More et al. 1981) to implement the two hybrid methods with the new line search with a portable computer. The stop criterion is $\|g_k\| \leq 10^{-6}$ and we set the parameters as $a_1 = 0.2$, $a_2 = 0.2$, $\sigma_1 = \sigma_2 = 0.6$ and $\mu = 0.4$. Four conjugate gradient algorithms (DY, Hybrid conjugate gradient method of DY and HS, PRP, Hybrid conjugate gradient method of FR and PRP) are compared in numerical performance and the numerical results are given in Table 1.

In Table 1, CPU denotes the CPU time (seconds) for solving all the 15 test problems. A pair numbers means the number of iterations and the number of functional evaluations. It can be seen from Table 1 that two hybrid methods with the new Wolfe line search is effective for solving some large scale problems. In particular, the Hybrid conjugate gradient method of FR and PRP seems to be the best one among the four algorithms because it uses the least number of iterations and functional evaluations when the algorithms reach the same precision.

Table 1 Number of iterations and number of functional evaluations

p	n	DY	DY and HS	PRP	FR and PRP
21	10^4	72/222	70/212	58/179	50/189
22	10^4	74/532	70/520	63/429	58/409
23	10^4	62/241	60/241	41/198	40/176
24	10^4	89/304	89/300	82/297	82/302
25	10^4	51/153	44/143	46/127	44/107
26	10^4	55/202	56/202	52/198	46/188
27	10^4	45/246	46/238	52/239	42/209
28	10^4	91/425	84/415	84/337	80/312
29	10^4	58/245	55/240	45/211	39/183
30	10^4	53/308	48/298	43/281	43/277
31	10^4	74/268	74/258	76/281	70/265
32	10^4	77/457	74/447	87/421	82/425
33	10^4	48/172	38/152	37/136	33/126
34	10^4	82/396	80/383	58/385	56/354
35	10^4	76/342	74/322	72/276	68/256
CPU	–	355 s	312 s	252 s	231 s

Conclusion

In this paper, we have proposed two hybrid conjugate gradient methods, respectively are the hybrid method of DY and HS, the hybrid method of FR and PRP. Moreover, we have proposed the corresponding new Wolfe line search, which make the corresponding hybrid method keep the global convergence property and the descent property.

Dai and Yuan have proposed a family of the three-term conjugate gradient method:

$$\beta_k = \frac{(1 - \lambda_k)\|g_k\|^2 + \lambda_k g_k^{\mathrm{T}}(g_k - g_{k-1})}{(1 - \mu_k - \omega_k)\|g_{k-1}\|^2 + \mu_k d_{k-1}^{\mathrm{T}}(g_k - g_{k-1}) - \omega_k d_{k-1}^{\mathrm{T}} g_{k-1}}$$

where $\lambda_k \in [0, 1]$, $\mu_k \in [0, 1]$, $\omega_k \in [0, 1 - \mu_k]$.

Let $\mu_k = 1$, $\omega_k = 0$, we have

$$\beta_k = \frac{(1 - \lambda_k)\|g_k\|^2 + \lambda_k g_k^{\mathrm{T}}(g_k - g_{k-1})}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})} \tag{22}$$

Let $\mu_k = \omega_k = 0$, then we have

$$\beta_k = \frac{(1 - \lambda_k)\|g_k\|^2 + \lambda_k g_k^{\mathrm{T}}(g_k - g_{k-1})}{\|g_{k-1}\|^2} \tag{23}$$

In our paper, β_k can be rewritten as

$$\beta_k^{(1)} = \frac{a_1\|g_k\|^2 + a_2 g_k^{\mathrm{T}}(g_k - g_{k-1})}{d_{k-1}^{\mathrm{T}}(g_k - g_{k-1})}$$

$$\beta_k^{(2)} = \frac{a_1\|g_k\|^2 + a_2 g_k^{\mathrm{T}}(g_k - g_{k-1})}{\|g_{k-1}\|^2}$$

The difference between (22), (23) and $\beta_k^{(1)}$, $\beta_k^{(2)}$ defined in the paper is the numerator $\|g_k\|^2$ and $g_k^{\mathrm{T}}(g_k - g_{k-1})$ are convex combination of the formula (22), (23), however, in the formula $\beta_k^{(1)}$, $\beta_k^{(2)}$, $a_1 + a_2 \neq 1$, which have weakened the condition.

References

Al-Baali M (1985) Descent property and global convergence of the Fletcher–Reeves method with inexact line search. IMA J Number Anal 5:121–124

Birgin EG, Martinez JM (2001) A spectral conjugate gradient method for unconstrained optimization. Appl Math Optim 43:117–128

Dai YH, Yuan Y (2000) A nonlinear conjugate gradient method with a strong global convergence property. SIAM J Optim 10:177–182

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

Hestenes MR, Stiefel EL (1952) Method of conjugate gradient for solving linear systems. J Res Natl Bur Stand 49:409–432

Liu GH, Han JY, Yin HX (1995) Global convergence of the Fletcher–Reeves algorithm with an inexact line search. Appl Math J Chin Univ Ser B 10:75–82

More JJ, Garbow BS, Hillstrom KE (1981) Testing unconstrained optimization software. ACM Trans Math Softw 7:17–41

Polak E, Ribiere G, Polak E et al (1968) Note sur la convergence de methode de directions conjuguees[J]. Rev franaise Informat recherche Opérationnelle 16(16):35–43

Polyak BT (1969) The conjugate gradient method in extreme problem. USSR Com-put Math Math Phys 9:94–112

Powell MJD (1977) Restart procedures of the conjugate gradient method. Math Progr 2:241–254

Powell MJD (1984) Nonconvex minimization calculations and the conjugate gradient method[M]//numerical analysis. Springer, Berlin, pp 122–141

Vincent TL (1983) Practical methods of optimization-volume 1: unconstrained optimization (New York: Wiley & Sons). Math Bio Sci 64:151–152

Yuan Y (1993) numerical methods for nonlinear programming. Shanghai Scientific & Technical Publishers, China

Yuan Y, Stoer J (1995) A subspace study on conjugate algorithms. ZAMM Z Angew Math Mech 75(11):69–77

Zoutendijk G (1970) Nonlinear programming computational methods[J]. Integer Nonlinear Program 143(1):37–86