A class of globally convergent conjugate gradient methods

DAI Yuhong (戴或虹) & YUAN Yaxiang (袁亚湘)

State Key Laboratory of Scientific and Engineering Computing,
Institute of Computational Mathematics and Scientific/Engineering Computing,
Academy of Mathematics and System Sciences,
Chinese Academy of Sciences, Beijing 100080, China
(email-addresses: {dyh,yyx}@lsec.cc.ac.cn)
Correspondence should be addressed to Dai Yuhong.
Received April 15, 2002.

Abstract *Conjugate gradient methods are very important ones for solving nonlinear optimization problems, especially for large scale problems. However, unlike quasi-Newton methods, conjugate gradient methods were usually analyzed individually. In this paper, we propose a class of conjugate gradient methods, which can be regarded as some kind of convex combination of the Fletcher-Reeves method and the method proposed by Dai et al. To analyze the class of methods, we introduce some unified tools that concern a general method with the scalar* β_k *having the form of* ϕ_k/ϕ_{k-1}. *Consequently, the class of conjugate gradient methods can uniformly be analyzed.*

Keywords: unconstrained optimization, conjugate gradient, line search, global convergence.

Consider the unconstrained optimization problem

$$\min f(x), \quad x \in R^n, \tag{0.1}$$

where f is smooth and its gradient g is available. Conjugate gradient methods are very important methods for solving (0.1), especially for large scale problems, which have the following form:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{0.2}$$

$$d_k = \begin{cases} -g_k, & \text{for } k = 1; \\ -g_k + \beta_k d_{k-1}, & \text{for } k \geq 2, \end{cases} \tag{0.3}$$

where $g_k = \nabla f(x_k)$, α_k is a stepsize obtained by a one-dimensional line search and β_k is a scalar.

Since Fletcher and Reeves introduced the nonlinear conjugate gradient method in 1964, many formulae have been proposed to compute the scalar β_k, see [1, 2, 3, 4, 5, 6, 7, 8, 9] *etc.* Among them, two well-known formulae for β_k are called the FR and PRP formulae (see [4, 7, 8]), and are given by

$$\beta_k^{FR} = \|g_k\|^2/\|g_{k-1}\|^2 \tag{0.4}$$

and

$$\beta_k^{PRP} = g_k^T y_{k-1}/\|g_{k-1}\|^2 \tag{0.5}$$

respectively, where $y_{k-1} = g_k - g_{k-1}$ and $\|\cdot\|$ means the Euclidean norm. The properties of nonlinear conjugate gradient methods may be quite different with the scalar β_k. A typical example is that (see [10]), the FR method may sink into a cycle of small steps thus leading to bad numerical performances, whereas the PRP method will take the steepest descent direction approximately once a small step is produced. Nonlinear conjugate gradient methods have been analyzed individually, see [11, 12, 13, 14, 15, 16, 17, 18, 19, 20] *etc.*

It is well known that some quasi-Newton methods can be expressed in a unified way. For example, the Broyden's family can be written as one parameter class which can be viewed as a combination of the BFGS and DFP methods. Consequently, properties of the methods in the Broyden's family and their convergences can be analyzed uniformly (see, [21], [22] and [23]). A larger family of quasi-Newton methods is called the Huang's family ([24]). Therefore a question of much theoretical interest is as follows. Does there exists a class of nonlinear conjugate gradient methods and its properties can be analyzed by a unified tool?

This paper will give a positive answer to this question partly through our observations on the formulae of β_k. In [1], we presented a nonlinear conjugate gradient method, which has the form (0.2)–(0.3) with

$$\beta_k^{DY} = \|g_k\|^2/d_{k-1}^T y_{k-1}. \tag{0.6}$$

Such a nonlinear conjugate gradient method was shown to be globally convergent under the Wolfe line search conditions. An algorithm based on (0.6) was tested and it performs better than the PRP method on a set of test problems([1]). By direct calculations, we can deduce an equivalent form for β_k^{DY}, namely,

$$\beta_k^{DY} = g_k^T d_k / g_{k-1}^T d_{k-1}. \tag{0.7}$$

We see that the FR formula and the above formula for β_k are special forms of

$$\beta_k = \phi_k/\phi_{k-1}, \tag{0.8}$$

In this paper, we consider a class of methods that use (0.8) to define β_k and where ϕ_k satisfies that

$$\phi_k = \lambda\|g_k\|^2 + (1-\lambda)(-g_k^T d_k), \tag{0.9}$$

with $\lambda \in [0,1]$ being a parameter. This class of conjugate gradient methods can be viewed as some kind of convex combination of the FR method and the method (0.6), because ϕ_k is a convex combination of $\|g_k\|$ and $-g_k^T d_k$.

This paper is organized as follows. Some preliminaries are given in the next section; Section 3 provides two convergence theorems for general method (0.2)–(0.3) with β_k defined by (0.8); Section 4 analyzed the class of conjugate gradient methods where ϕ_k is defined by (0.9). Some remarks are made in the last section.

1. Preliminaries

Throughout this paper, we assume that $g_k \neq 0$ for all k, for otherwise a stationary point has been found. We give the following basic assumption on the objective function.

Assumption 1.1 *(i) f is bounded below on the level set $\mathcal{L} = \{x \in \Re^n : f(x) \le f(x_1)\}$; (ii) In some neighborhood $\mathcal{N}$ of $\mathcal{L}$, f is differentiable and its gradient g is Lipschitz continuous, namely, there exists a constant $L > 0$ such that*

$$\|g(x) - g(\tilde{x})\| \le L\|x - \tilde{x}\|, \qquad \textit{for all } x,\ \tilde{x} \in \mathcal{N}. \tag{1.1}$$

Some of the results obtained in this paper depend also on the following assumption.

Assumption 1.2 *The level set $\mathcal{L} = \{x \in \Re^n : f(x) \le f(x_1)\}$ is bounded.*

If f satisfies Assumptions 1.1 and 1.2, there exists a positive constant $\overline{\gamma}$ such that

$$\|g(x)\| \leq \overline{\gamma}, \quad \text{for all } x \in \mathcal{L}. \tag{1.2}$$

The stepsize α_k in (0.2) is computed by carrying out certain line searches. The Wolfe line search [25] is to find a positive stepsize α_k such that

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{1.3}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \tag{1.4}$$

where $0 < \delta < \sigma < 1$. Under Assumption 1.1 on f, we state the following result, which was essentially obtained in [25, 26, 27].

Lemma 1.3 *Suppose that x_1 is a starting point for which Assumption 1.1 holds. Consider any iterative method (0.2), where d_k is a descent direction and α_k is computed by the Wolfe line search (1.3)–(1.4). Then*

$$\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{1.5}$$

In the convergence analyses and numerical implementations of conjugate gradient methods, the stepsize α_k is often computed by the strong Wolfe line search which requires α_k satisfying (1.3) and

$$|g(x_k + \alpha_k d_k)^T d_k| \leq -\sigma g_k^T d_k, \tag{1.6}$$

where also $0 < \delta < \sigma < 1$. For the purpose of analysis, this paper is also concerned about the line search conditions (1.3) and

$$\sigma_1 g_k^T d_k \leq g(x_k + \alpha_k d_k)^T d_k \leq -\sigma_2 g_k^T d_k, \tag{1.7}$$

where $0 < \delta < \sigma_1 < 1$ and $\sigma_2 \geq 0$. It is obvious that the Wolfe line search and the strong Wolfe line search are corresponding to (1.3) and (1.7) with $\sigma_1 = \sigma$, $\sigma_2 = \infty$ and $\sigma_1 = \sigma_2 = \sigma$ respectively.

In the latter sections, the following lemmas are also needed, the first of which is derived from [28], whereas the second is self-evident and will be used for many times.

Lemma 1.4 *Suppose that $\{a_i\}$ and $\{b_i\}$ are positive number sequences. If*

$$\sum_{k \geq 1} a_k = \infty \tag{1.8}$$

and for all $k \geq 1$,

$$b_k \leq c_1 + c_2 \sum_{i=1}^{k} a_i, \tag{1.9}$$

where c_1 and c_2 are positive constants, then we have that

$$\sum_{k \geq 1} a_k / b_k = \infty. \tag{1.10}$$

Lemma 1.5 *Consider the following* 1*-dimensional function,*

$$\rho(t) = \frac{a + bt}{c + dt}, \quad t \in R^1, \tag{1.11}$$

where a, b, c *and* $d \neq 0$ *are given real numbers. If*

$$bc - ad > 0, \tag{1.12}$$

$\rho(t)$ *is strictly monotonically increasing for* $t < -\frac{c}{d}$ *and* $t > -\frac{c}{d}$*; otherwise, if*

$$bc - ad < 0, \tag{1.13}$$

$\rho(t)$ *is strictly monotonically decreasing for* $t < -\frac{c}{d}$ *and* $t > -\frac{c}{d}$.

2. Convergence of the general method (0.8)

In this section, the general method (0.2)–(0.3) with β_k defined by (0.8) is studied. After giving a basic lemma, we establish two convergence results which depend certain conditions on ϕ_k.

For simplicity, we define

$$t_k = \frac{\|d_k\|^2}{\phi_k^2} \tag{2.1}$$

and

$$r_k = -\frac{g_k^T d_k}{\phi_k}. \tag{2.2}$$

Lemma 2.1 *For the method (0.2)–(0.3) with* β_k *defined by (0.8),*

$$t_k = -2\sum_{i=1}^{k} \frac{g_i^T d_i}{\phi_i^2} - \sum_{i=1}^{k} \frac{\|g_i\|^2}{\phi_i^2} \tag{2.3}$$

holds for all $k \geq 1$.

Proof Since $d_1 = -g_1$, (2.3) holds for $k = 1$. For $i \geq 2$, it follows from (0.3) that

$$d_i + g_i = \beta_i d_{i-1}. \tag{2.4}$$

Squaring both sides of the above equation, we get that

$$\|d_i\|^2 = -\|g_i\|^2 - 2g_i^T d_i + \beta_i^2 \|d_{i-1}\|^2. \tag{2.5}$$

Dividing (2.5) by ϕ_i^2 and applying (0.8) and (2.1),

$$t_i = t_{i-1} - \frac{2g_i^T d_i}{\phi_i^2} - \frac{\|g_i\|^2}{\phi_i^2}. \tag{2.6}$$

Summing this expression over i, we obtain

$$t_k = t_1 - 2\sum_{i=2}^{k} \frac{g_i^T d_i}{\phi_i^2} - \sum_{i=2}^{k} \frac{\|g_i\|^2}{\phi_i^2}. \tag{2.7}$$

Since $d_1 = -g_1$ and $t_1 = \|g_1\|^2/\phi_1^2$, the above relation is equivalent to (2.3). So (2.3) holds for all $k \geq 1$. □

Theorem 2.2 *Suppose that* x_1 *is a starting point for which Assumption 1.1 holds. Consider the method (0.2), (0.3) and (0.8), if for all* k d_k *is a descent direction and* α_k *is computed by the Wolfe line search (1.3)–(1.4), and if*

$$\sum_{k\geq 1} r_k^2 = \infty, \tag{2.8}$$

we have that

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{2.9}$$

Proof Since

$$-2g_i^T d_i - \|g_i\|^2 \leq \frac{(g_i^T d_i)^2}{\|g_i\|^2}, \tag{2.10}$$

it follows from (2.3) that

$$t_k \leq \sum_{i=1}^{k} \frac{(g_i^T d_i)^2}{\|g_i\|^2 \phi_i^2}, \tag{2.11}$$

or equivalently,

$$t_k \leq \sum_{i=1}^{k} \frac{r_i^2}{\|g_i\|^2}. \tag{2.12}$$

Assume that (2.9) is not true, namely,

$$\liminf_{k\to\infty} \|g_k\| \neq 0. \tag{2.13}$$

Then there exists a positive constant γ such that

$$\|g_k\| \geq \gamma, \quad \text{for all } k. \tag{2.14}$$

In this case, it follows by (2.12) that

$$t_k \leq \frac{1}{\gamma^2} \sum_{i=1}^{k} r_i^2. \tag{2.15}$$

The above relation, (2.8) and Lemma 1.4 yield

$$\sum_{i\geq 1} \frac{r_i^2}{t_i} = \infty. \tag{2.16}$$

By the definitions of t_k and r_k, we know that (2.16) contradicts (1.5). Therefore (2.9) is true. □

Theorem 2.3 *Suppose that* x_1 *is a starting point for which Assumption 1.1 holds. Consider the method (0.2), (0.3) and (0.8), if for all* k d_k *is a descent direction and* α_k *is computed by the Wolfe line search (1.3)–(1.4), and if*

$$\sum_{k\geq 1} \frac{\|g_k\|^2}{\phi_k^2} = \infty, \tag{2.17}$$

we have that $\liminf \|g_k\| = 0$.

Proof Noting that

$$t_k \geq 0 \tag{2.18}$$

for all k, we can get from (2.3) that

$$-2\sum_{i=1}^{k}\frac{g_i^T d_i}{\phi_i^2} \geq \sum_{i=1}^{k}\frac{\|g_i\|^2}{\phi_i^2}, \tag{2.19}$$

which yields that

$$4\sum_{i=1}^{k}\frac{(g_i^T d_i)^2}{\|g_i\|^2\phi_i^2} \geq -4\sum_{i=1}^{k}\frac{g_i^T d_i}{\phi_i^2} - \sum_{i=1}^{k}\frac{\|g_i\|^2}{\phi_i^2} \geq \sum_{i=1}^{k}\frac{\|g_i\|^2}{\phi_i^2}. \tag{2.20}$$

Thus if (2.17) holds, we also have that

$$\sum_{k\geq 1}\frac{(g_k^T d_k)^2}{\|g_k\|^2\phi_k^2} = \infty. \tag{2.21}$$

Because (2.11) still holds, it follows from (2.21) and Lemma 1.4 that

$$\sum_{k\geq 1}\frac{(g_k^T d_k)^2}{\|g_k\|^2\|d_k\|^2} = \infty. \tag{2.22}$$

The above relation and Lemma 1.3 clearly give (2.9). This completes our proof. □

Thus we have proved two convergence theorems for the general method (0.2)-(0.3) with β_k defined by (0.8). From the above results, we can see that the proof to the convergence of any method in the form (0.8) can be divided into two stages: the first stage is to show the descent property of the search direction and the second is to show the truth of (2.8) or (2.17).

For the method (0.6), (2.8) clearly holds since in this case $\phi_k = -g_k^T d_k$ and hence $r_k = 1$. If f satisfies Assumption 1.2, then we have (2.17) for the FR method because in this case $\phi_k = \|g_k\|^2$ and (1.2) holds. Therefore Theorems 2.2 and 2.3 are powerful tools in analyzing the convergence of any conjugate gradient method provided that β_k has the form (0.8).

It should also be noted that the sufficient descent condition, namely,

$$g_k^T d_k \leq -c\|g_k\|^2, \tag{2.23}$$

where c is a positive constant, is not invoked in Theorems 2.2 and 2.3. The sufficient descent condition (2.23) was often used or implied in the previous analyses of conjugate gradient methods (for example, see [11, 15]). This condition has been relaxed to the descent condition ($g_k^T d_k < 0$) in the convergence analyses [1] of the FR method and the convergence analyses [29] of any conjugate gradient method. Another point is that both theorems can be easily extended to any method (0.2)–(0.3) with β_k satisfying

$$|\beta_k| \leq \phi_k/\phi_{k-1}, \tag{2.24}$$

because in this case, instead of (2.3), we can show that

$$t_k \leq -2\sum_{i=1}^{k}\frac{g_i^T d_i}{\phi_i^2} - \sum_{i=1}^{k}\frac{\|g_i\|^2}{\phi_i^2}, \tag{2.25}$$

which is sufficient for us to prove Theorems 2.2 and 2.3.

3. A class of globally convergent conjugate gradient methods

In this section, we will exploit a class of conjugate gradient methods between the FR method and the method (0.6). The global convergence of the class is proved under certain line search conditions and the methods related to the class are uniformly discussed.

We consider the method (0.2) – (0.3) with ϕ_k satisfying

$$\phi_k = \lambda\|g_k\|^2 + (1-\lambda)(-g_k^T d_k), \tag{3.1}$$

where $\lambda \in [0,1]$. It is obvious that the FR method and the method (0.6) are corresponding to $\lambda = 1$ and $\lambda = 0$ respectively. (3.1) and (0.3) show that

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \beta_k g_k^T d_{k-1} \\ &= -\|g_k\|^2 + \frac{\lambda\|g_k\|^2 + (1-\lambda)(-g_k^T d_k)}{\lambda\|g_{k-1}\|^2 + (1-\lambda)(-g_{k-1}^T d_{k-1})} g_k^T d_{k-1}. \end{aligned} \tag{3.2}$$

The above relation gives that

$$g_k^T d_k = -\frac{\lambda(\|g_{k-1}\|^2 - g_k^T d_{k-1}) + (1-\lambda)(-g_{k-1}^T d_{k-1})}{\lambda\|g_{k-1}\|^2 + (1-\lambda) d_{k-1}^T y_{k-1}} \|g_k\|^2. \tag{3.3}$$

Thus by the first equality in (3.2), we deduce an equivalent form of β_k,

$$\beta_k = \frac{\|g_k\|^2}{\lambda\|g_{k-1}\|^2 + (1-\lambda) d_{k-1}^T y_{k-1}}. \tag{3.4}$$

The above form for β_k can be used for practical computations. Substituting (3.3) into (3.1), we obtain that

$$\phi_k = \frac{\lambda\|g_{k-1}\|^2 + (1-\lambda)(-g_{k-1}^T d_{k-1})}{\lambda\|g_{k-1}\|^2 + (1-\lambda) d_{k-1}^T y_{k-1}} \|g_k\|^2. \tag{3.5}$$

By this relation, we can show an important property of ϕ_k under Wolfe line searches and hence obtain the global convergence of the class of conjugate gradient methods (3.4) under some assumptions.

Theorem 3.1 *Suppose that x_1 is a starting point for which Assumptions 1.1 and 1.2 hold. Consider the method (0.2), (0.3), (0.8) and (3.1), where $\lambda \in [0,1]$ and α_k is computed by the Wolfe line search (1.3)–(1.4). If $g_k^T d_k < 0$ for all k, then*

$$\phi_k \le \frac{1}{1-\sigma}\|g_k\|^2. \tag{3.6}$$

Further, the method converges in the sense that (2.9) is true.

Proof The line search condition (1.4) implies that

$$d_{k-1}^T y_{k-1} \ge (1-\sigma)(-g_{k-1}^T d_{k-1}), \tag{3.7}$$

which, with relation (3.5), shows the truth of (3.6). It follows from (1.2) and (3.7) that

$$\sum_{k\geq 1} \frac{\|g_k\|^2}{\phi_k^2} = \infty. \tag{3.8}$$

Thus (2.9) follows from Theorem 2.3. □

Let us now define

$$\overline{r}_k = -\frac{g_k^T d_k}{\|g_k\|^2} \tag{3.9}$$

and

$$l_k = \frac{g_{k+1}^T d_k}{g_k^T d_k}. \tag{3.10}$$

Then by (3.3), we can write

$$\overline{r}_k = \frac{\lambda + (1-\lambda+\lambda l_{k-1})\overline{r}_{k-1}}{\lambda + (1-\lambda)(1-l_{k-1})\overline{r}_{k-1}}. \tag{3.11}$$

By the above relation, we can show that, if the line search conditions are (1.3) and (1.7) where the scalars σ_1 and σ_2 satisfy certain condition, then for any $\lambda \in (0,1]$, the method (0.2), (0.3), (0.8) and (3.1) ensures the descent property of each search direction and converges globally. The assumption on the objective function used here is slightly weaker those than in Theorem 3.1.

Theorem 3.2 *Suppose that x_1 is a starting point for which Assumption 1.1 holds. Consider the method (0.2), (0.3), (0.8) and (3.1), where $\lambda \in (0,1]$ and α_k satisfies the line search conditions (1.3) and (1.7). If the scalars σ_1 and σ_2 in (1.7) is such that*

$$\sigma_1 + \sigma_2 \leq \lambda^{-1}, \tag{3.12}$$

then we have for all $k \geq 1$,

$$\overline{r}_k > 0. \tag{3.13}$$

Further, the method converges in the sense that (2.9) is true.

Proof The right hand side of (3.11) is a function of λ, l_{k-1} and $\overline{r}_{k-1}$, which is denoted as $\psi(\lambda, l_{k-1}, \overline{r}_{k-1})$. First, we show that

$$0 < \overline{r}_k < (1-\sigma_1)^{-1} \tag{3.14}$$

for all $k \geq 1$. Since $d_1 = -g_1$ and hence $\overline{r}_1 = 1$, (3.14) clearly holds for $k = 1$. Suppose that (3.14) is true for some $k-1$. It follows from (1.7) that

$$-\sigma_2 \leq l_{k-1} \leq \sigma_1. \tag{3.15}$$

Then by Lemma 1.5, we get that

$$\overline{r}_k \leq \psi(\lambda, \sigma_1, \overline{r}_{k-1}) < \psi(\lambda, \sigma_1, (1-\sigma_1)^{-1}) = (1-\sigma_1)^{-1}. \tag{3.16}$$

On the other hand, by Lemma 1.5 and relation (3.12), we also have that

$$\overline{r}_k \geq \psi(\lambda, -\sigma_2, \overline{r}_{k-1}) > \psi(\lambda, -\sigma_2, (1-\sigma_1)^{-1}) \geq 0. \tag{3.17}$$

Thus (3.14) is true for k. By induction, (3.14) holds for all $k \geq 1$.

To show the truth of (2.9), by Theorem 2.2, it suffices to prove that

$$\max\{r_{k-1}, r_k\} \geq c_1 \tag{3.18}$$

for all $k \geq 2$ and some constant $c_1 > 0$. In fact, if

$$\overline{r}_{k-1} \leq 1, \tag{3.19}$$

we can get by Lemma 1.5 that

$$\overline{r}_k \geq \psi(\lambda, -\sigma_2, 1) \triangleq c_2. \tag{3.20}$$

Since $c_2 \in (0,1)$, we then obtain

$$\max\{\overline{r}_{k-1}, \overline{r}_k\} \geq c_2 \tag{3.21}$$

for all $k \geq 2$. By the definition (2.2) of r_k and relation (3.1), we have that

$$r_k = \frac{\overline{r}_k}{\lambda + (1-\lambda)\overline{r}_k}, \tag{3.22}$$

which, with (3.21) and Lemma 1.5, implies that (3.18) holds with $c_1 = c_2$. This completes our proof. □

Thus we have established two convergence results for the class of conjugate gradient methods (3.4). Letting $\lambda = 1$ in Theorem 3.2, we again obtain the convergence result of the FR method in [12]. For the case when $\lambda = 0$, the method is proved to generate a descent search direction at every iteration and converge globally under the Wolfe line search conditions (1.3)–(1.4) (see [14]). Such a result can be regarded in certain sense as the limit of the results in Theorem 3.2 when $\lambda \to 0$, since (3.12) implies that σ_2 may tend to infinity when λ tends to zero.

In the following, we study methods related to the class of conjugate gradient methods (3.4). To combine the nice global-convergence properties of the FR method and the good numerical performances of the PRP method, [30] discussed the methods related to the FR method and extended the result in [11] to any method (0.2) and (0.3) with β_k satisfying

$$0 \leq \beta_k \leq \beta_k^{FR}. \tag{3.23}$$

[15] further extended the result to the case that

$$|\beta_k| \leq \beta_k^{FR}. \tag{3.24}$$

For the nonlinear conjugate method (0.6), [1] proved that the method (0.2)–(0.3) with β_k satisfying

$$\beta_k \in \left[\frac{\sigma - 1}{1 + \sigma}\bar{\beta}_k, \bar{\beta}_k\right], \tag{3.25}$$

where $\bar{\beta}_k$ stands for the formula (0.6), and with α_k chosen by the Wolfe line search gives the convergence relation (2.9). If the line search conditions are (1.3) and (1.6) with $\sigma \leq 1/2$, these results can be seen as special cases of the following general result.

Theorem 3.3 *Suppose that x_1 is a starting point for which Assumption 1.1 holds. Consider the method (0.2) and (0.3), where*

$$\beta_k = \frac{\tau_k \|g_k\|^2}{\lambda \|g_{k-1}\|^2 + (1-\lambda) d_{k-1}^T y_{k-1}}, \tag{3.26}$$

and where α_k is computed by the strong Wolfe line search (1.3) and (1.6) with $\sigma \le 1/2$. For any $\lambda \in [0,1]$, if

$$\tau_k \in \left[\frac{\sigma - 1}{1 + (1-2\lambda)\sigma}, 1 \right], \tag{3.27}$$

then the method produces a descent direction at every iteration and converges globally in the sense that (2.9) is true.

Proof Denote

$$\bar{\beta}_k = \frac{\lambda \|g_k\|^2 + (1-\lambda)(-g_k^T d_k)}{\lambda \|g_{k-1}\|^2 + (1-\lambda)(-g_{k-1}^T d_{k-1})} \tag{3.28}$$

and

$$\xi_k = \frac{\beta_k}{\bar{\beta}_k}. \tag{3.29}$$

Direct calculations show that

$$\overline{r}_k = \frac{\lambda + [\tau_k l_{k-1} + (1-\lambda)(1-l_{k-1})]\overline{r}_{k-1}}{\lambda + (1-\lambda)(1-l_{k-1})\overline{r}_{k-1}} \tag{3.30}$$

and

$$\xi_k = \frac{[\lambda + (1-\lambda)\overline{r}_{k-1}]\tau_k}{\lambda + (1-\lambda)(1 - l_{k-1} + \tau_k l_{k-1})\overline{r}_{k-1}}, \tag{3.31}$$

where $\overline{r}_k$ and l_k are defined by (3.9) and (3.10). Now, the right hand side of (3.30) is a function of λ, τ_k, l_{k-1} and $\overline{r}_{k-1}$, which can be denoted as $\psi(\lambda, \tau_k, l_{k-1}, \overline{r}_{k-1})$. We first show that

$$0 < \overline{r}_k < (1-\sigma)^{-1} \tag{3.32}$$

holds for all $k \ge 1$. Since $\overline{r}_k = 1$, (3.32) holds for $k = 1$. Suppose that (3.32) is true for some $k-1$. It follows from (1.7) that

$$|l_{k-1}| \le \sigma. \tag{3.33}$$

This relation and Lemma 1.5 give that

$$\begin{aligned} \overline{r}_k &\le \max\{\psi(\lambda, 1, l_{k-1}, \overline{r}_{k-1}), \psi(\lambda, \frac{\sigma-1}{1+(1-2\lambda)\sigma}, l_{k-1}, \overline{r}_{k-1})\} \\ &\le \max\{\psi(\lambda, 1, \sigma, \overline{r}_{k-1}), \psi(\lambda, \frac{\sigma-1}{1+(1-2\lambda)\sigma}, -\sigma, \overline{r}_{k-1})\} \\ &< \max\{\psi(\lambda, 1, \sigma, (1-\sigma)^{-1}), \psi(\lambda, \frac{\sigma-1}{1+(1-2\lambda)\sigma}, -\sigma, (1-\sigma)^{-1})\} \\ &= (1-\sigma)^{-1}, \end{aligned} \tag{3.34}$$

where $\sigma \le 1/2$ is also used in the equality. For the opposite direction, we can prove that

$$\overline{r}_k > \min\{\psi(\lambda, 1, -\sigma, (1-\sigma)^{-1}), \psi(\lambda, \frac{\sigma-1}{1+(1-2\lambda)\sigma}, \sigma, (1-\sigma)^{-1})\} \ge 0. \tag{3.35}$$

Thus (3.32) is true for k. Therefore by induction, (3.32) holds for all $k \geq 1$.

Now we prove that

$$\xi_k \in [-1, 1] \tag{3.36}$$

for all $k \geq 2$. Denoting D_k to be the denominator of ξ_k in (3.31), direct calculations show that

$$(1 - \xi_k) D_k = (1 - \tau_k)[\lambda + (1 - \lambda)(1 - l_{k-1})\overline{r}_{k-1}] \tag{3.37}$$

and

$$(\xi_k + 1) D_k = [\lambda + (1 - \lambda)(1 + l_{k-1})\overline{r}_{k-1}]\tau_k + [\lambda + (1 - \lambda)(1 - l_{k-1})\overline{r}_{k-1}]. \tag{3.38}$$

Applying (3.27), (3.32) and (3.33), we can show that $D_k > 0$ and the right hand terms in relations (3.37) and (3.38) are nonnegative. So (3.36) holds. Besides it, similarly to the proof of Theorem 3.2, one can verify that (3.18) is also true for some constant $c_1 > 0$. By (3.18), (3.36) and the related discussion in Section 3, we know that (2.9) must hold. □

4. Some remarks

We have studied the convergence properties of the general method (0.8) and provided two sufficient conditions which ensure the global convergence of the method. The results are powerful tools in analyzing the convergence of any conjugate gradient method in the form (0.8) and hence enable us to establish convergence results of the class of conjugate gradient methods (3.4).

¿From Theorems 2.2, 2.3 and 3.1, we can see that, the descent property of the search direction plays an important role in establishing convergence results of the method in the form (0.8). At the same time, we can also see that the sufficient descent condition (2.23) is not necessary in the convergence analysis of the method in the form (0.8).

It can be seen from Theorem 3.2 that the properties of the class of conjugate gradient methods (3.4) seem to more resemble those of the FR method with an exception of the method (0.6). One evident is that, for the method (0.8) and (3.1) where $\lambda \in (0, 1]$, if the line search conditions are (1.3) and (1.7) with σ_1 and σ_2 satisfying (3.12), then due to (3.21), we know that the sufficient descent condition (2.23) holds for at least one of any neighboring two iterations. However, such a property does not hold any more for the method (0.6) using the Wolfe line search. For the method (0.6), it can be shown that (2.23) is true for most of the iterations, see [31].

¿From the view of theory, it would be interesting to investigate whether Theorems 3.2 and 3.3 can be extended to the case that $\lambda > 1$ or $\lambda < 0$. As described in [1], an algorithm based on the method (0.6) has been found which performs much better than the PRP method. Therefore from the view of computation, one may ask whether a more efficient algorithm can be exploited according to the results of this paper. These questions still remain under investigations.

Acknowledgements

This paper is written based on the report Dai and Yuan (1998) [1]. This work was partially supported by the National Natural Science Foundation of China (Grant Nos. 19525101, 19731010, 19801033 and 10171104), and also by an Innovation Fund of the Academy of Mathematics and System Sciences of CAS.

[1]Dai Y H, Yuan Y. A class of globally convergent conjugate gradient methods, Research report ICM-98-030, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1998

References

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

[2] Daniel, J. W., The conjugate gradient method for linear and nonlinear operator equations, SIAM J. Numer. Anal., 1967, 4: 10–26.

[3] Fletcher, R., Practical methods of optimization vol. 1: unconstrained optimization, New York: John Wiley and Sons, 1987.

[4] Fletcher, R., Reeves, C., Function minimization by conjugate gradients, Comput. J., 1964, 7: 149–154.

[5] Hestenes, M. R., Stiefel, E. L., Methods of conjugate gradients for solving linear systems, J. Res. Nat. Bur. Standards, 1952, 49(5): 409–436.

[6] Liu, Y., Storey, C., Efficient Generalized Conjugate Gradient Algorithms, Part 1: Theory, Journal of Optimization Theory and Applications, 1991, 69: 129–137.

[7] Polak, E., Ribière, G., Note sur la convergence de directions conjugées, Rev. française Informat Recherche Opérationelle, 1969, 16: 35–43.

[8] Polyak, B. T., The conjugate gradient method in extremem problems, USSR Comp. Math. and Math. Phys., 1969, 9: 94–112.

[9] Shanno, D. F., Conjugate gradient methods with inexact searches, Math. Oper. Res., 1978, 3: 244–256.

[10] Powell, M. J. D., Restart procedures of the conjugate gradient method, Math. Program. 1977, 2: 241–254.

[11] Al-Baali, M., Descent property and global convergence of the Fletcher-Reeves method with inexact line search, IMA J. Numer. Anal., 1985, 5: 121–124.

[12] Dai, Y. H., Yuan, Y., Convergence properties of the Fletcher-Reeves method, IMA J. Numer. Anal., 1996, 16(2): 155–164.

[13] Dai, Y. H., Yuan, Y., Convergence properties of the conjugate descent method, Advances in Mathematics (China), 1996, 25(6): 552–562.

[14] Dai, Y. H., Yuan, Y., Some properties of a new conjugate gradient method, In: Y. Yuan ed., Advances in Nonlinear Programming. Boston: Kluwer, 1998, 251–262.

[15] Gilbert, J. C., Nocedal, J., Global convergence properties of conjugate gradient methods for optimization, SIAM J. Optimization, 1992, 2(1): 21–42.

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

[17] Liu, G. H., Han, J. Y., Yin, H. X., Global convergence of the Fletcher-Reeves algorithm with an inexact line search, Appl. Math. J. Chinese Univ. (Series B), 1995, 10: 75–82.

[18] Powell, M. J. D., Nonconvex minimization calculations and the conjugate gradient method, In: Lecture Notes in Mathematics (Vol. 1066), Berlin: Springer-Verlag, 1984. 122–141.

[19] Qi, H. D., Han, J. Y., Liu, G. H., A modified Hestenes-Stiefel conjugate gradient algorithm, Chinese Annals of Mathematics 1996, 17A (3): 277–284. (in Chinese)

[20] Touati-Ahmed, D., Storey, C., Efficient hybrid conjugate gradient techniques, J. Optimization Theory Appl., 1990, 64: 379–397.

[21] Broyden, C. G., The convergence of a class of double-rank minimization algorithms 1. general considerations, J. Inst. Math. Appl., 1970, 6: 76–90.

[22] Byrd, R., Nocedal, J., Yuan, Y., Global convergence of a class of variable metric algorithms, SIAM J. Numer. Anal., 1987, 4: 1171–1190.

[23] Peng, J. M., Yuan, Y., Properties of Ψ-function and its application, Chinese Numerical Mathematics, 1994, 16: 102–107. (in Chinese)

[24] Huang, H. Y., Unified approach to quadratically convergent algorithms for function minimization, J. Optimization Theory and Appl., 1970, 5: 405–423.

[25] Wolfe, P., Convergence conditions for ascent methods, SIAM Review, 1969, 11: 226–235.

[26] Wolfe, P., Convergence conditions for ascent methods. II: Some corrections, SIAM Review, 1969, 13: 185–188.

[27] Zoutendijk, G., Nonlinear Programming, Computational Methods, In: Abadie J, ed. Integer and Nonlinear Programming. Amsterdam: North-Holland, 1970, 37–86.

[28] Pu, D., Yu, W., On the convergence properties of the DFP algorithms, Annals of Operations Research, 1990, 24: 175–184.

[29] Dai, Y. H., Han, J. Y., Liu, G. H., Sun, D. F., Yin, H. X., Yuan, Y., Convergence properties of nonlinear conjugate gradient methods, SIAM Journal on Optimization, 1999, 10(2): 345–358.

[30] Hu, Y. F., Storey, C., Global convergence result for conjugate gradient methods, JOTA, 1991, 71(2): 399–405.

[31] Dai, Y. H., New properties of a nonlinear conjugate gradient method, Numerische Mathematik, 2001, 89: 83–98.