

A new family of globally convergent conjugate gradient methods

B. Sellami¹ · Y. Chaïb¹

Published online: 8 February 2016
© Springer Science+Business Media New York 2016

Abstract Conjugate gradient methods are an important class of methods for unconstrained optimization, especially for large-scale problems. Recently, they have been much studied. In this paper, a new family of conjugate gradient method is proposed for unconstrained optimization. This method includes the already existing two practical nonlinear conjugate gradient methods, which produces a descent search direction at every iteration and converges globally provided that the line search satisfies the Wolfe conditions. The numerical experiments are done to test the efficiency of the new method, which implies the new method is promising. In addition the methods related to this family are uniformly discussed.

Keywords Unconstrained optimization · Conjugate gradient method · Line search · Global convergence

Mathematics Subject Classification 65K05 · 90C25 · 90C26 · 90C27 · 90C30

1 Introduction

Consider the unconstrained optimization problem

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

where f is a smooth function and its gradient is available. Conjugate gradient methods are a class of important methods for solving (1), especially for large scale problems, which have the following form:

$$x_{k+1} = x_k + \alpha_k d_k \quad (2)$$

✉ B. Sellami
basellami@gmail.com

Y. Chaïb
chaibyacine@yahoo.com

¹ Mohamed Chrif Messaadia University, Souk-Ahras, Algeria

where x_k is the current iterate, α_k is a positive scalar and called the step length which is determined by some line search, and d_k is the search direction generated by the rule

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

where $g_k = \nabla f(x_k)$ is the gradient of f at x_k , and β_k is a scalar. The strong Wolfe conditions, namely,

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k \tag{4}$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \leq -\sigma g_k^T d_k, \tag{5}$$

where $0 < \delta < \sigma < 1$. The scalar β_k is chosen so that the method (2), (3) reduces to the linear conjugate gradient method in the case when f is convex quadratic and exact line search ($g(x_k + \alpha_k d_k)^T d_k = 0$) is used. For general functions, however, different formula for scalar β_k result in distinct nonlinear conjugate gradient methods, (see Dai and Yuan 1999; Fletcher and Reeves 1964; Polyak 1969; Fletcher 1987; Shanno 1978). Since Fletcher and Reeves (FR) introduced the linear conjugate gradient method in 1964, with

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \tag{6}$$

where $\|\cdot\|$ means the Euclidean norm. For non-quadratic objective functions, the global convergence of (FR) method was proved when the exact line search or strong Wolfe line search (Al-Baali 1985; Dai and Yuan 1996) was used. However, if the condition (5) is satisfied for $\sigma < 1$, the above method of (FR) with the strong Wolfe line search can ensure a descent search direction and converge globally provided only for the case when f is quadratic (Dai et al. 2000; Dai and Yuan 1999). The conjugate descent (CD) method of Fletcher (1987), where

$$\beta_k^{CD} = \frac{\|g_k\|^2}{-g_{k-1}^T d_{k-1}}, \tag{7}$$

ensures a descent direction for general functions if the line search satisfies the strong Wolfe conditions (4), (5) with $\sigma < 1$. But the global convergence of the method is proved (see Dai et al. 2000) only for the case when the line search satisfies (4) and

$$\sigma g_k^T d_k \leq g(x_k + \alpha_k d_k)^T d_k \leq 0. \tag{8}$$

For any positive constant σ_2 , an example is constructed in Dai et al. (2000) showing that conjugate descent method with α_k satisfying (4) and

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

need not converge. Recently, Dai and Yuan (1999) proposed a nonlinear conjugate gradient method, which has the form (2), (3) with

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}, \tag{10}$$

where $y_{k-1} = g_k - g_{k-1}$. A remarkable property of the DY method is that it provides a descent search direction at every iteration and converges globally provided that the step size satisfies the Wolfe conditions, namely, (4) and

$$\sigma g_k^T d_k \leq g(x_k + \alpha_k d_k)^T d_k. \tag{11}$$

By direct calculations, we can deduce an equivalent form for β_k^{DY} , namely

$$\beta_k^{DY} = \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}. \tag{12}$$

In Dai and Yuan (2003), proposed a class of globally convergent conjugate gradient methods, in which

$$\beta_k = \frac{\|g_k\|^2}{\lambda \|g_{k-1}\|^2 + (1 - \lambda) (d_{k-1}^T y_{k-1})}, \tag{13}$$

where $\lambda \in [0, 1]$ is a parameter, and proved that the family of methods using line searches that satisfy (4) and (9) converges globally if the parameters σ_1, σ_2 , and λ are such that

$$\sigma_1 + \sigma_2 \leq \lambda^{-1}.$$

In addition, Sellami et al. (2015) proposed a new two-parameter family of conjugate gradient methods, in which

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

where $\lambda_k \in [0, 1]$ and $\mu_k \in [0, 1 - \lambda_k]$ are parameters, and proved that the new two-parameter family can ensure a descent search direction at every iteration and converges globally under line search condition (4) and (9) where the scalars σ_1 and σ_2 satisfy the condition

$$\sigma_1 + \sigma_2 \leq \frac{1 + \mu_k \sigma_1}{1 - \lambda_k}.$$

Observing that the formula (7) and (10) share same numerators and two denominators. In this paper we can use combinations of these numerators and denominators to obtain the following new family of conjugate gradient methods

$$\beta_k^* = \frac{(1 - \lambda) \|g_k\|^2 + \lambda (-g_k^T d_k)}{(1 - \lambda) (-g_{k-1}^T d_{k-1}) + \lambda (-g_{k-1}^T d_{k-1})} \tag{14}$$

Thus by the above equality in (14), we deduce an equivalent form of β_k^* ,

$$\beta_k^* = \frac{(1 - \lambda) \|g_k\|^2 + \lambda (-g_k^T d_k)}{-g_{k-1}^T d_{k-1}} \tag{15}$$

with $\lambda \in [0, 1]$ being a parameter. We see that the above formula for β_k^* is special forms of

$$\beta_k^* = \frac{\phi_k}{\phi'_{k-1}}, \tag{16}$$

where ϕ_k satisfies that

$$\phi_k = (1 - \lambda) \|g_k\|^2 + \lambda (-g_k^T d_k), \tag{17}$$

and

$$\phi'_{k-1} = (1 - \lambda) (-g_{k-1}^T d_{k-1}) + \lambda (-g_{k-1}^T d_{k-1}) = -g_{k-1}^T d_{k-1} \tag{18}$$

It is clear that the formula (15) is a generalization of the two previous methods are defined by (7) and (10).

The rest of this paper is organized as follows. Some preliminaries are given in the next section. Section 3 provides two convergence theorems for the general method (2), (3) with β_k^* defined by (16). Section 4 includes the main convergence properties of the new family of conjugate gradient methods with Wolfe line search, and we study methods related to the new nonlinear conjugate gradient method (15). The preliminary numerical results are contained in Sect. 5. Conclusions and discussions are made in the last section.

2 Preliminaries

For convenience, 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 2.1 (i) f is bounded below on the level set $\mathcal{L} = \{x \in R^n; f(x) \leq f(x_1)\}$;
 (ii) In some neighborhood N of \mathcal{L} , f is differentiable and its gradient g is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(\tilde{x})\| \leq L \|x - \tilde{x}\|, \quad \text{for all } x, \tilde{x} \in N \tag{19}$$

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

Assumption 2.2 The level set $\mathcal{L} = \{x \in R^n; f(x) \leq f(x_1)\}$ is bounded.

If f satisfies Assumption 2.1 and 2.2, there exists a positive constant γ such that

$$\|g(x)\| \leq \gamma, \quad \text{for all } x \in \mathcal{L}. \tag{20}$$

The conclusion of the following lemma, often called the Zoutendijk condition, is used to prove the global convergence of nonlinear conjugate gradient methods. It was originally given in Zoutendijk (1970).

Lemma 2.3 Suppose Assumption 2.1 holds. Let $\{x_k\}$ be generated by (2) and d_k satisfy $g_k^T d_k < 0$. If α_k is determined by the Wolfe line search (4), (11), then we have

$$\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{21}$$

In the latter section, we need the following lemmas, the first of which is derived from Hu and Storey (1991), Pu and Yu (1990), whereas the second is self-evident and will be used for many times.

Lemma 2.4 Suppose that $\{a_i\}$ and $\{b_i\}$ are positive number sequences. If

$$\sum_{k \geq 1} a_k = \infty, \tag{22}$$

and there exist two constants c_1 and c_2 such that for all $k \geq 1$,

$$b_k \leq c_1 + c_2 \sum_{i=1}^k a_i, \tag{23}$$

then we have that

$$\sum_{k \geq 1} \frac{a_k}{b_k} = \infty. \tag{24}$$

Lemma 2.5 Consider the following 1-dimensional function,

$$\rho(t) = \frac{a + bt}{c + dt}, \quad t \in \mathbb{R}^1, \tag{25}$$

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

$$bc - ad > 0, \tag{26}$$

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

$$bc - ad < 0, \tag{27}$$

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

3 Algorithm and convergence properties

Now we can present a new descent conjugate gradient method, namely NDCG method, as follows:

Algorithm 3.1 Step 0: Given $x_1 \in \mathbb{R}$, set $d_1 = -g_1, k = 1$. If $g_1 = 0$, then stop.

Step 1: Find a $\alpha_k > 0$ satisfying the Wolfe conditions (4) and (11).

Step 2: Let $x_{k+1} = x_k + \alpha_k d_k$ and $g_{k+1} = g(x_{k+1})$. If $g_{k+1} = 0$, then stop.

Step 3: Compute β_{k+1}^* by the formula (15) and generate d_{k+1} by (3).

Step 4: Set $k := k + 1$, go to Step 1.

In order to establish the global convergence result for the Algorithm 3.1, we will impose the following basic lemma.

For simplicity, we define

$$r_k = -\frac{g_k^T d_k}{\phi_k}, \tag{28}$$

and

$$t_k = \frac{\|d_k\|^2}{\phi_k^2}. \tag{29}$$

Lemma 3.1 For the method (2), (3) with β_k^* defined by (16),

$$t_k = 2 \sum_{i=1}^k \frac{r_i}{\phi_i} - \sum_{i=1}^k \frac{\|g_i\|^2}{\phi_i^2}, \tag{30}$$

holds for all $k \geq 1$.

Proof Since $d_1 = -g_1$, (30) holds for $k = 1$. For $i \geq 2$, it follows from (3) that

$$d_i + g_i = \beta_i^* d_{i-1}. \tag{31}$$

Squaring both sides of the above equation, we get that

$$\|d_i\|^2 = -\|g_i\|^2 - 2g_i^T d_i + \beta_i^{*2} \|d_{i-1}\|^2. \tag{32}$$

Dividing (32) by ϕ_i^2 and applying (16) and (29),

$$t_i = \frac{\|d_{i-1}\|^2}{\phi_{i-1}^2} + 2\frac{r_i}{\phi_i} - \frac{\|g_i\|^2}{\phi_i^2}. \tag{33}$$

Using (18), (17) and since, $d_1 = -g_1$ we get that

$$\frac{\|d_1\|^2}{\phi_1^2} = \frac{\|g_1\|^2}{\|g_1\|^4} = \frac{\|g_1\|^2}{\phi_1^2}. \tag{34}$$

Summing the above expression (33) over i , we obtain

$$t_k = t_1 + 2 \sum_{i=2}^k \frac{r_i}{\phi_i} - \sum_{i=2}^k \frac{\|g_i\|^2}{\phi_i^2}. \tag{35}$$

Since $d_1 = -g_1$ and $t_1 = \frac{\|g_1\|^2}{\phi_1^2}$, the above relation is equivalent to (30). So (30) holds for $k \geq 1$. □

Theorem 3.2 *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the method (2), (3) and (16), if for all k , d_k satisfy $g_k^T d_k < 0$ and α_k is determined by the Wolfe line search (4), (11), and if*

$$\sum_{k \geq 1} r_k^2 = \infty, \tag{36}$$

we have that

$$\lim_{k \rightarrow \infty} \inf \|g_k\| = 0. \tag{37}$$

Proof Equation (3) can be re-written as

$$g_i^T d_i + \|g_i\|^2 = \beta_i^* g_i^T d_{i-1}. \tag{38}$$

Squaring both sides of the above equation, we get that

$$-2g_i^T d_i - \|g_i\|^2 \leq \frac{(g_i^T d_i)^2}{\|g_i\|^2}, \tag{39}$$

dividing (39) by ϕ_i^2 and applying (30)

$$t_k \leq \sum_{i=1}^k \frac{r_i^2}{\|g_i\|^2}. \tag{40}$$

We proceed by contradiction. Assume that

$$\lim_{k \rightarrow \infty} \inf \|g_k\| \neq 0. \tag{41}$$

Then there exists a positive constant γ such that

$$\|g_k\| \geq \gamma, \text{ for all } k. \tag{42}$$

We can see from (40) that,

$$t_k \leq \frac{1}{\gamma^2} \sum_{i=1}^k r_i^2. \tag{43}$$

The above relation, (36) and Lemma 2.4, yield

$$\sum_{i \geq 1} \frac{r_i^2}{t_i} = \infty. \tag{44}$$

Thus, by the definition (28) and (29), we know that (44) contradicts (21). This concludes the proof. \square

Theorem 3.3 *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the method (2), (3) and (16), if for all k , d_k satisfy $g_k^T d_k < 0$ and α_k is determined by the Wolfe line search (4), (11), and if*

$$\sum_{k \geq 1} \frac{\|g_k\|^2}{\phi_k^2} = \infty, \tag{45}$$

we have that

$$\lim_{k \rightarrow \infty} \inf \|g_k\| = 0. \tag{46}$$

Proof Noting that

$$t_k \geq 0 \text{ for all } k, \tag{47}$$

Squaring the right side of equation (39), we get that

$$\left(-2g_i^T d_i - \|g_i\|^2\right)^2 \geq 0.$$

Hence, we have

$$4\left(g_i^T d_i\right)^2 + \|g_i\|^4 + 4\left(g_i^T d_i\right)\|g_i\|^2 \geq 0. \tag{48}$$

Summing this expression over i and dividing (48) by $(\phi_i^2 \|g_i\|^2)$, we obtain

$$4 \sum_{i=1}^k \frac{(g_i^T d_i)^2}{\phi_i^2 \|g_i\|^2} \geq -4 \sum_{i=1}^k \frac{(g_i^T d_i)}{\phi_i^2} - \sum_{i=1}^k \frac{\|g_i\|^2}{\phi_i^2}. \tag{49}$$

On the other hand, we can get from (30) and (47)

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

Direct calculation show that,

$$-4 \sum_{i=1}^k \frac{g_i^T d_i}{\phi_i^2} - \sum_{i=1}^k \frac{\|g_i\|^2}{\phi_i^2} \geq \sum_{i=1}^k \frac{\|g_i\|^2}{\phi_i^2}. \tag{50}$$

The above relation (49) and (50) imply that

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

Thus if (45) holds, we also have that

$$\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|g_k\|^2 \phi_k^2} = \infty. \tag{51}$$

Because (40) still holds, it follows from (51), the definition of r_k and Lemma 2.4, that

$$\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|g_k\|^2 \|d_k\|^2} = \infty. \tag{52}$$

The above relation and Lemma 2.3 clearly give (37). This completes our proof.

Thus we have proved two convergence theorems for the general method (2), (3) with β_k^* defined by (16).

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

$$g_k^T d_k \leq -c \|g_k\|^2, \tag{53}$$

where c is a positive constant, is not invoked in Theorems 3.2 and 3.3. The sufficient descent condition (53) was often used or implied in the previous analysis of conjugate gradient methods (see Al-Baali 1985; Gilbert and Nocedal 1992). This condition has been relaxed to the descent condition ($g_k^T d_k < 0$) in the convergence analysis (Dai and Yuan 1999) of the FR method and the convergence analysis (Dai et al. 2000) of any conjugate gradient method. □

4 Global convergence of new conjugate gradient method

In this section, we establish some global convergence of the new family of nonlinear conjugate gradient methods under certain line searches conditions and the methods related to this family are uniformly discussed.

We consider the method (2), (3) with ϕ_k satisfying

$$\phi_k = (1 - \lambda) \|g_k\|^2 + \lambda (-g_k^T d_k), \tag{54}$$

where $\lambda \in [0, 1]$. (54) and (3) show that

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \beta_k^* g_k^T d_{k-1} \\ &= -\|g_k\|^2 + \frac{(1 - \lambda) \|g_k\|^2 + \lambda (-g_k^T d_k)}{-g_{k-1}^T d_{k-1}} g_k^T d_{k-1}. \end{aligned} \tag{55}$$

The above relation imply that

$$g_k^T d_k = -\frac{(1 - \lambda) (-g_k^T d_{k-1}) - g_{k-1}^T d_{k-1}}{\lambda g_k^T d_{k-1} - g_{k-1}^T d_{k-1}} \|g_k\|^2. \tag{56}$$

Thus by (55), we deduce an equivalent form of β_k^* ,

$$\beta_k^* = \frac{\|g_k\|^2}{\lambda g_k^T d_{k-1} - g_{k-1}^T d_{k-1}}. \tag{57}$$

Substituting (56) into (54), we obtain that

$$\phi_k = \frac{-g_{k-1}^T d_{k-1}}{\lambda g_k^T d_{k-1} - g_{k-1}^T d_{k-1}} \|g_k\|^2. \tag{58}$$

By this relation, we can show an important property of ϕ_k under Wolfe line searches and hence obtain the global convergence of the new family of nonlinear conjugate gradient methods (57) under some assumptions.

Theorem 4.1 *Suppose that x_1 is a starting point for which Assumption 2.1 and 2.2 hold. Consider the method (2), (3), (16) and (54), if $g_k^T d_k < 0$ for all k and α_k is computed by the Wolfe line search (4), (11), then*

$$\frac{\phi_k}{\|g_k\|^2} \leq (1 - \lambda\sigma)^{-1}. \tag{59}$$

Further, the method converges in the sense that

$$\lim_{k \rightarrow \infty} \inf \|g_k\| = 0. \tag{60}$$

Proof Since (11), we have that

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k. \tag{61}$$

By direct calculations show that

$$\lambda g_k^T d_{k-1} - g_{k-1}^T d_{k-1} \geq (1 - \lambda\sigma) (-g_{k-1}^T d_{k-1}). \tag{62}$$

Dividing (58) by $\|g_k\|^2$ and applying (62) implies the truth of (59). Therefore, by (20) and (62) that

$$\sum_{k \geq 1} \frac{\|g_k\|^2}{\phi_k^2} \geq \frac{(1 - \lambda\sigma)^2}{\|g_k\|^2} \geq \frac{(1 - \lambda\sigma)^2}{\gamma^2} = \infty. \tag{63}$$

Thus (37) follows from Theorem 3.3.

In the following, we can show that, for any $\lambda \in (0, 1]$, the method (2), (3), (16) and (54) ensures the descent property of each search direction and converges globally under line search condition (4) and (9) where the scalar σ_2 satisfy certain condition. For this purpose, we define

$$\bar{r}_k = -\frac{g_k^T d_k}{\|g_k\|^2}, \tag{64}$$

and

$$l_k = \frac{g_{k+1}^T d_k}{g_k^T d_k}, \tag{65}$$

it is obvious that d_k is a descent direction if and only if $\bar{r}_k > 0$. For the above relation, (56) and (65), we can write

$$\bar{r}_k = \frac{1 + (1 - \lambda)l_{k-1}}{1 - \lambda l_{k-1}}. \tag{66}$$

□

Theorem 4.2 *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the method (2), (3), (16) and (54), where $\lambda \in [0, 1)$ and α_k satisfies the line search conditions (4) and (9). If the scalar σ_2 in (9) is such that*

$$\sigma_2 \leq (1 - \lambda)^{-1}, \tag{67}$$

then we have for all $k \geq 1$

$$0 < \bar{r}_k < (1 - \sigma_1)^{-1}. \tag{68}$$

Further, the method converges in the sense that (37) is true.

Proof The right hand side of (66) is a function of λ, l_{k-1} and \bar{r}_{k-1} , which is denoted as

$h(\lambda, l_{k-1}, \bar{r}_{k-1})$. We prove (68) by induction. Noting that $d_1 = -g_1$ and hence $\bar{r}_1 = 1$, we see that (68) is true for $k = 1$. We now suppose that (68) holds for $k - 1$, namely,

$$0 < \bar{r}_{k-1} < (1 - \sigma_1)^{-1}. \tag{69}$$

It follows from (9)

$$-\sigma_2 \leq l_{k-1} \leq \sigma_1. \tag{70}$$

Then by Lemma 2.5, the fact that $\lambda \in [0, 1)$, we get that

$$\begin{aligned} \bar{r}_k &\leq h(\lambda, \sigma_1, \bar{r}_{k-1}) < h(\lambda, \sigma_1, (1 - \sigma_1)^{-1}) \\ &= 1 + \frac{\sigma_1}{1 - \lambda\sigma_1} \\ &\leq 1 + \frac{\sigma_1}{1 - \sigma_1} \\ &= (1 - \sigma_1)^{-1}. \end{aligned} \tag{71}$$

On the other hand, by Lemma 2.5 and relation (67), we also have that

$$\bar{r}_k \geq h(\lambda, -\sigma_2, \bar{r}_{k-1}) > h(\lambda, -\sigma_2, (1 - \sigma_1)^{-1}) \geq 0. \tag{72}$$

Thus (68) is true for k , by induction, (68) holds for $k \geq 1$.

To show the truth of (37), by Theorem 3.2, it suffices to prove that

$$\max\{r_{k-1}, r_k\} \geq c_1, \tag{73}$$

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

$$\bar{r}_{k-1} \leq 1, \tag{74}$$

by Lemma 2.5, the fact that $\lambda \in [0, 1)$, we can get that

$$\bar{r}_k \geq h(\lambda, -\sigma_2, 1) \triangleq c_2. \tag{75}$$

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

$$\max\{\bar{r}_{k-1}, \bar{r}_k\} \geq c_2, \tag{76}$$

for all $k \geq 2$. By the definition (28) of r_k and relation (54), we have that

$$r_k = \frac{\bar{r}_k}{1 + \lambda(\bar{r}_k - 1)}. \tag{77}$$

Which, with (76) and lemma (23), implies that (73) holds with $c_1 = c_2$. This completes our proof. □

Thus we have some general convergence results are established for the new family of nonlinear conjugate gradient methods (57). It is easy to see from (57) that the new family of conjugate gradient methods includes the two nonlinear conjugate gradient methods mentioned above. For the case when $\lambda = 1$, the method is proved to generate a descent search direction at every iteration and converge globally of the DY method under the Wolfe line search conditions (4), (11) (see Dai and Yuan 1999). If $\lambda = 0$, then the method ensures a descent direction for general functions and is proved to global convergence under strong Wolfe line search (4), (5) of the method CD (see Dai et al. 2000).

In addition, the methods related to the FR method and the DY method in Hu and Storey (1991), Dai and Yuan (1999) can also be regarded as special cases of the new family methods (57). For example, to combine the nice global convergence properties of the FR method and the good numerical performances of the PRP method, namely

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}. \tag{78}$$

Hu and Storey (1991) extended the result in Al-Baali (1985) to any method (2) and (3) with β_k satisfying

$$\beta_k \in [0, \beta_k^{FR}]. \tag{79}$$

Gilbert and Nocedal (1992) further extended the result to the case that

$$\beta_k \in [-\beta_k^{FR}, \beta_k^{FR}]. \tag{80}$$

Dai and Yuan (2001) studied the hybrid conjugate gradient algorithms and proposed the following hybrid methods

$$\beta_k = \max \left\{ 0, \min \left\{ \beta_k^{HS}, \beta_k^{DY} \right\} \right\}. \tag{81}$$

where β_k^{HS} is the choice of Hestenes and Stiefel (1952) and β_k^{DY} appears in Dai and Yuan (1999). Furthermore, Dai and Yuan (1999) proved that the method (2) and (3) with β_k satisfying

$$\beta_k \in \left[\frac{\sigma - 1}{1 + \sigma} \overline{\beta}_k, \overline{\beta}_k \right], \tag{82}$$

where $\overline{\beta}_k$ stands for the formula (10), and with α_k chosen by the Wolfe line search give the convergence relation (37). For methods related to the method (57). We have the following result, where s_k is given by

$$s_k = \frac{\beta_k}{\beta_k^*}, \tag{83}$$

where β_k^* stands for the formula (15). We prove that any method (20), (21) with the strong Wolfe line search produces a descent search direction at every iteration and converges globally if the scalar β_k is such that

$$-c \leq s_k \leq (1 - \sigma)^{-1}, \tag{84}$$

where $c = (1 + \sigma)/(1 - \sigma) > 0$.

Theorem 4.3 *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the method (2) and (3), where*

$$\beta_k = \frac{\tau_k \|g_k\|^2}{\lambda g_k^T d_{k-1} - g_{k-1}^T d_{k-1}}, \tag{85}$$

and where α_k is computed by the strong Wolfe line search (4) and (9) with $\sigma \leq \frac{1}{2}$. For any $\lambda \in [0, 1]$, if

$$\tau_k \in \left[\frac{1 + \lambda\sigma}{\sigma - 1}, \frac{1 - \lambda\sigma}{1 - \sigma} \right], \tag{86}$$

and β_k is such that

$$s_k \in [-c, (1 - \sigma)^{-1}], \tag{87}$$

then if $g_k \neq 0$ for all $k \geq 1$, we have that

$$0 < \bar{r}_k < (1 - \sigma)^{-1} \text{ for all } k \geq 1. \tag{88}$$

Further, the method converges in the sense that (37) is true.

Proof From relation (15), (85) and by direct calculations we can show that

$$\bar{r}_k = \frac{1 - (\lambda - \tau_k)l_{k-1}}{1 - \lambda l_{k-1}}, \tag{89}$$

and

$$s_k = \frac{\tau_k}{1 - \lambda(1 - \tau_k)l_{k-1}}, \tag{90}$$

where \bar{r}_k and l_k are defined by (64) and (65). Now the right hand side of (89) is a function of λ, τ_k, l_{k-1} and \bar{r}_{k-1} , which can be denoted as $h(\lambda, \tau_k, l_{k-1}, \bar{r}_{k-1})$. We prove (88) by induction. Noting that $d_1 = -g_1$ and hence $\bar{r}_1 = 1$, we see that (4.34) is true for $k = 1$. We now suppose that (88) holds for $k - 1$, namely,

$$0 < \bar{r}_{k-1} < (1 - \sigma)^{-1}. \tag{91}$$

It follows from (5)

$$|l_{k-1}| \leq \sigma. \tag{92}$$

Then by Lemma 2.5, and the fact that $\lambda \in [0, 1]$, we get that

$$\begin{aligned} \bar{r}_k &\leq \max \left\{ h \left(\lambda, \frac{1 - \lambda\sigma}{1 - \sigma}, l_{k-1}, \bar{r}_{k-1} \right), h \left(\lambda, \frac{1 + \lambda\sigma}{\sigma - 1}, l_{k-1}, \bar{r}_{k-1} \right) \right\} \\ &\leq \max \left\{ h \left(\lambda, \frac{1 - \lambda\sigma}{1 - \sigma}, \sigma, \bar{r}_{k-1} \right), h \left(\lambda, \frac{1 + \lambda\sigma}{\sigma - 1}, -\sigma, \bar{r}_{k-1} \right) \right\} \\ &< \max \left\{ h \left(\lambda, \frac{1 - \lambda\sigma}{1 - \sigma}, \sigma, (1 - \sigma)^{-1} \right), h \left(\lambda, \frac{1 + \lambda\sigma}{\sigma - 1}, -\sigma, (1 - \sigma)^{-1} \right) \right\} \\ &= 1 + \frac{\sigma}{1 - \sigma} = (1 - \sigma)^{-1}, \end{aligned} \tag{93}$$

where $\sigma \leq \frac{1}{2}$ is also used in the equality. For the opposite direction, we can prove that

$$\bar{r}_k > \min \left\{ h \left(\lambda, \frac{1 - \lambda\sigma}{1 - \sigma}, -\sigma, (1 - \sigma)^{-1} \right), h \left(\lambda, \frac{1 + \lambda\sigma}{\sigma - 1}, \sigma, (1 - \sigma)^{-1} \right) \right\} \geq 0. \tag{94}$$

Thus (88) is true for k , by induction, (88) holds for $k \geq 1$.

We now prove (37) by contradiction and assuming that

$$\|g(x)\| \geq \gamma, \text{ for some } \gamma > 0, \tag{95}$$

for all $k \geq 1$, since $d_k + g_k = \beta_k d_{k-1}$, we have that

$$\|d_k\|^2 = \beta_k^2 \|d_{k-1}\|^2 - 2g_k^T d_k - \|g_k\|^2. \tag{96}$$

Dividing both sides of (96) by $(g_k^T d_k)^2$ and using (64) and (83), we obtain

$$\begin{aligned} \frac{\|d_k\|^2}{(g_k^T d_k)^2} &= \frac{\beta_k^2 \|d_{k-1}\|^2}{(g_k^T d_k)^2} + \frac{2}{\bar{r}_k \|g_k\|^2} - \frac{1}{\bar{r}_k^2 \|g_k\|^2} \\ &= \frac{(s_k \beta_k^*)^2 \|d_{k-1}\|^2}{(g_k^T d_k)^2} + \frac{1}{\|g_k\|^2} \left[1 - \left(1 - \frac{1}{\bar{r}_k} \right)^2 \right]. \end{aligned} \tag{97}$$

In addition, by the definition (64) of \bar{r}_k , the relation (3) and (83), we get

$$\bar{r}_k \|g_k\|^2 = -g_k^T d_k = \|g_k\|^2 - s_k \beta_k^* g_k^T d_{k-1}, \tag{98}$$

the above relation and the definition (65) imply that

$$s_k \beta_k^* = \frac{(1 - \bar{r}_k)}{l_{k-1} (g_{k-1}^T d_{k-1})} \|g_k\|^2. \tag{99}$$

Relation (97) and (99), we obtain

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} = \frac{(1 - \bar{r}_k)^2 \|d_{k-1}\|^2}{\bar{r}_k^2 l_{k-1}^2 (g_{k-1}^T d_{k-1})^2} + \frac{1}{\|g_k\|^2} \left[1 - \left(1 - \frac{1}{\bar{r}_k} \right)^2 \right]. \tag{100}$$

Denote

$$m_k = \frac{1 - \bar{r}_k}{\bar{r}_k l_{k-1}}, \tag{101}$$

where $l_{k-1} \neq 0$. Now we prove that

$$|m_k| \leq 1, \text{ for all } k \geq 2. \tag{102}$$

the right hand side of (101) is a function of l_{k-1} and \bar{r}_k , which can be denoted as $h(l_{k-1}, \bar{r}_k)$. We can get by (88), (92) and Lemma 2.5 that

$$\begin{aligned} m_k &\leq \max \{ h(\sigma, \bar{r}_k), h(-\sigma, \bar{r}_k) \} \\ &< \max \{ h(\sigma, (1 - \sigma)^{-1}), h(-\sigma, (1 - \sigma)^{-1}) \} = 1. \end{aligned} \tag{103}$$

Thus we have that

$$\begin{aligned}
 m_k &\geq \min \{h(-\sigma, \bar{r}_k), h(\sigma, \bar{r}_k)\} \\
 &> \min \{h(-\sigma, (1-\sigma)^{-1}), h(\sigma, (1-\sigma)^{-1})\} = -1.
 \end{aligned}
 \tag{104}$$

Therefore (101) holds for all $k \geq 2$.

By (102) and (100), we obtain

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{1}{\|g_k\|^2}.
 \tag{105}$$

Because $\|d_1\|^2 / (g_1^T d_1)^2 = 1 / \|g_1\|^2$, (105) shows that

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \leq \sum_{i=1}^k \frac{1}{\|g_i\|^2},
 \tag{106}$$

for all k . Then we get from this and (95) that

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \frac{\gamma^2}{k},
 \tag{107}$$

which implies that

$$\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = +\infty.
 \tag{108}$$

This contradicts the Zoutendijk condition (21). Therefore (37) holds. □

5 Numerical results

In this section, we will test the following four conjugate gradient algorithms:

- PRP^{SW}: the PRP method with the strong Wolfe conditions, where $\delta = 10^{-2}$ and $\sigma = 0, 1$.
- PRP⁺: the PRP method with nonnegative values of $\beta_k = \max\{0, \beta_k^{PRP}\}$ and the strong Wolfe conditions, where $\delta = 10^{-2}$ and $\sigma = 0, 1$.
- NDCG^{SW}: Algorithm 3.1 with the Wolfe conditions (4) and (9), where the scalar σ_2 satisfy the condition (67), in addition, $\delta = 10^{-2}$, $\sigma_1 = \sigma_2 = \sigma = 0, 1$, $\lambda = 0, 5$.
- NDCG^W: Algorithm 3.1 with the standard Wolfe conditions, where $\delta = 10^{-2}$, $\sigma = 0, 1$, $\lambda = 0, 5$.

In this paper, all codes were written in Matlab and run on PC with 3.0 GHz CPU processor and 1GB RAM memory and Linux operation system. During our experiments, the strategy for the initial step length is to assume that the first-order change in the function at iterate x_k will be the same as that obtained at the previous step (Hu and Storey 1991). In other words, we choose the initial guess α_0 satisfying:

$$\alpha_0 = \alpha_{k-1} \frac{\Psi_{k-1}}{\Psi_k} \text{ for all } k > 1,$$

where $\Psi_k = g_k^T d_k$, when $k = 1$, we choose $\alpha_0 = \frac{1}{\|g(x_1)\|}$. In the case when an uphill search direction does occur, we restart the algorithm by setting $d_k = -g_k$, but this case never occurs for NDCG^{SW} and NDCG^W . We stop the iteration if the inequality $\|g(x_k)\| < 10^{-6}$ is satisfied. The iteration is also stopped if the number of iteration exceed 10,000, but we find that this never occurs for our tested problems. The test problems we used are described in Hillstrome et al. (1981). Each problem was tested with various values of n changing from $n = 500$ to 1000.

Table 1 list numerical results. The meaning of each column is as follows:

“N”	The number of the test problem
“Problem”	The name of the test problem
“n”	The dimension of the test problem
“NI”	The number of iterations
“NF”	The total number of function evaluations
“NG”	The total number of gradient evaluations
“CPUtime(s)”	The total CPU time in seconds which should be taken to compute all of these problems

We can see from the above table that, the average performances of the NDCG^{SW} is better than that of the PRP method. Also see that, the NDCG^W outperforms other three algorithms for solving these problems, especially for problems 3, 4, 7, 9, 11 and 12. Table 1, shows the performance of these methods relative to CPU time. To solve all the 26 problems, the CPU time (in seconds) required by the PRP^{SW} , PRP_+^{SW} , NDCG^{SW} and NDCG^W are 3.5688e+2, 3.5441e+2, 3.3435e+2 and 3.3112e+2, respectively. These preliminary results obtained are encouraging.

6 Conclusions and discussions

In this paper, we have proposed a new family of nonlinear conjugate gradient methods, and studied the global convergence of these methods. The new family not only includes the two already known simple and practical conjugate gradient methods, but has other family of conjugate gradient methods as subfamily. First, we can see that, the descent property of the search direction plays an important role in establishing some general convergence results of the method in the form (16) with weak Wolfe line search (4) and (11) even in the absence of the sufficient descent condition (3.27), namely, Theorems 3.2, 3.3, 4.1. Next, from Theorem 4.2, we proved that the new family can ensure a descent search direction at every iteration and converges globally under line search condition (4) and (9) where the scalar σ_2 satisfy the condition (67). From Theorem (56), we have carefully studied methods related to the method (57). Denote s_k to be the size of β_k with respect to β_k^* . If τ_k and s_k belongs to some interval, namely, (86) and (87) respectively, the corresponding methods are shown to produce a descent search direction at every iteration and converge globally provided that the line search satisfies the strong Wolfe conditions (4) and (9) with $\sigma \leq \frac{1}{2}$. In summary, our computational results show that this new descent nonlinear conjugate gradient method, namely NDCG^W method not only converges globally, but also outperforms the original PRP method in average. The results, we hope, can stimulate more study on the theory and implementations on the conjugate gradient methods with the Wolfe line search. For future research, we should investigate to

Table 1 Test results on $PRP^{SW} / PRP_+^{SW} / NDCG^{SW} / NDCG^W$ methods

N	Problem	n	PRP^{SW}	PRP_+^{SW}	$NDCG^{SW}$	$NDCG^W$
			NI/NF/NG	NI/NF/NG	NI/NF/NG	NI/NF/NG
1	Extended rosenbrock	500	34/127/17	45/165/113	46/109/72	50/123/88
		1000	29/90/42	29/108/38	57/202/126	49/141/81
2	Extended powell singular	500	326/582/478	89/191/154	99/212/177	94/193/145
		1000	165/323/260	57/122/112	100/238/197	96/217/170
3	Penalty 1	500	68/165/122	79/227/159	45/247/154	49/240/160
		1000	73/169/135	94/297/189	48/158/123	66/236/160
4	Penalty 2	500	838/1102/986	897/1515/1142	224/146/156	441/310/252
		1000	115/217/166	109/289/170	71/205/153	88/244/171
5	Variably dimensioned	500	47/50/28	47/50/28	47/50/28	47/50/28
		1000	23/70/60	23/70/60	23/70/60	23/70/60
6	Trigonometric	500	51/140/103	49/132/103	55/149/105	55/109/70
		1000	57/102/68	57/102/68	61/110/79	63/106/71
7	Discrete boundary value	500	1347/2122/2115	1347/2122/2115	139/263/240	124/191/161
		1000	140/189/185	140/189/185	37/68/56	34/47/39
8	Discrete integral equation	500	13/17/10	13/17/10	13/17/10	13/17/10
		1000	13/17/10	13/17/10	13/17/10	13/17/10
9	Chebyquad	500	360/411/267	291/118/99	158/78/56	200/105/79
		1000	289/321/169	126/120/103	91/80/21	103/90/67
10	Broyden tridiagonal	500	84/90/87	103/151/125	195/214/201	167/141/163
		1000	105/224/102	373/225/263	1008/1074/1051	2014/2030/2070
11	Broyden banded	500	360/113/156	264/106/100	261/93/99	167/69/76
		1000	507/206/171	225/138/131	139/71/31	76/112/68
12	Helical valley	500	8750/2890/3050	7350/2572/2192	44/33/38	56/45/41
		1000	7673/5099/4381	6984/4346/2678	47/34/29	72/49/54
13	Powel badly scaled	500	74/92/41	74/92/41	74/92/41	74/92/41
		1000	74/92/41	74/92/41	74/92/41	74/92/41
	CPU time(s)		3.5688e+2	3.5441e+2	3.3435e+2	3.3112e+2

find the practical performance of the method (57). Furthermore, we can investigate whether Theorems 4.2 and 4.3 can be extended to the case that $\lambda > 1$ or $\lambda < 0$.

Acknowledgments We would like to thank to Professor Paul Armand (University of Limoges, France), who has always been generous with his time and advice.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflict of interest.

References

Al-Baali, M. (1985). Descent property and global convergence of the fletcherreeves method with inexact line search. *IMA Journal of Numerical Analysis*, 5(1), 121–124.

- Al-Baali, M. (1985). New property and global convergence of the fletcher-reeves method with inexact line searches. *IMA Journal of Numerical Analysis*, 5(1), 122–124.
- Dai, Y. H., & Yuan, Y. (1996). Convergence properties of the fletcher-reeves method. *IMA Journal of Numerical Analysis*, 16(2), 155–164.
- Dai, Y. H., & Yuan, Y. (2001). An efficient hybrid conjugate gradient method for unconstrained optimization. *Annals of Operations Research*, 103(1–4), 33–47.
- Dai, Y.-H., & Yuan, Y. (1999). A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization*, 10(1), 177–182.
- Dai, Y., Han, J., Liu, G., Sun, D., Yin, H., & Yuan, Y.-X. (2000). Convergence properties of nonlinear conjugate gradient methods. *SIAM Journal on Optimization*, 10(2), 345–358.
- Dai, Y., & Yuan, Y. (2003). A class of globally convergent conjugate gradient methods. *Science in China Series A: Mathematics*, 46(2), 251–261.
- Fletcher, R., & Reeves, C. M. (1964). Function minimization by conjugate gradients. *The computer Journal*, 7(2), 149–154.
- Fletcher, R. (1987). *Practical methods of optimization*. Hoboken: Wiley.
- Gilbert, J. C., & Nocedal, J. (1992). Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal on Optimization*, 2(1), 21–42.
- Hestenes, M. R., & Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Standards*, 49, 409–436.
- Hu, Y. F., & Storey, C. (1991). Global convergence result for conjugate gradient methods. *Journal of Optimization Theory and Applications*, 71(2), 399–405.
- Hillstrome, K. E., More, J. J., & Garbow, B. S. (1981). Testing unconstrained optimization software. *Mathematical Software*, 7, 17–41.
- Polyak, B. T. (1969). The conjugate gradient method in extremal problems. *USSR Computational Mathematics and Mathematical Physics*, 9(4), 94–112.
- Pu, D., & Yu, W. (1990). On the convergence property of the dfp algorithm. *Annals of Operations Research*, 24(1), 175–184.
- Sellami, B., Laskri, Y., & Benzine, R. (2015). A new two-parameter family of nonlinear conjugate gradient methods. *Optimization*, 64(4), 993–1009.
- Shanno, D. F. (1978). Conjugate gradient methods with inexact searches. *Mathematics of Operations Research*, 3(3), 244–256.
- Zoutendijk, G. (1970). Nonlinear programming, computational methods. *Integer and Nonlinear Programming*, 143(1), 37–86.