RAIRO-Oper. Res. 50 (2016) 1013–1026
DOI: 10.1051/ro/2015064

RAIRO Operations Research
www.rairo-ro.org

NEW CONJUGATE GRADIENT METHOD FOR UNCONSTRAINED OPTIMIZATION

Badreddine Sellami[1] and Yacine Chaib[1]

Abstract. In this paper, a new conjugate gradient method is proposed for large-scale unconstrained optimization. This method includes the already existing three 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 confirms the promising potentials of the new method.

Mathematics Subject Classification. 65K05, 90C25, 90C26, 90C27, 90C30.

Received April 17, 2015. Accepted December 21, 2015.

1. Introduction

Consider the unconstrained optimization problem

$$\min \; f(x), \qquad x \in R^n, \tag{1.1}$$

where f is a smooth function and its gradient is available. Conjugate gradient methods are very important methods for solving (1.1), especially for large scale problems, which have the following form

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

where x_k is the current iterate point, 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

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

where $g_k = \nabla f(x_k)$ is the gradient of f at x_k, and β_k is a scalar. The standard Wolfe conditions [21, 22] are given by

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k \tag{1.4}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \tag{1.5}$$

Keywords. Unconstrained optimization, conjugate gradient method, line search, global convergence.

[1] Department of mathematics and informatics, Mohamed Cherif Messaadia University, Souk-Ahras, Algeria.
`bsellami@univ-soukahras.dz; chaibyacine@yahoo.com`

where $0 < \delta < \sigma < 1$. Also, the strong Wolfe conditions consist of (1.4) and

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

the scalar β_k is chosen so that the method (1.2)–(1.3) 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 [6, 10, 13, 14, 17, 20]) . The best-known formulas for β_k are the following Polak–Ribière–Polyak (PRP)[16, 17] and Hestenes–Stiefel (HS) [13] formulas, which are given by

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} \tag{1.7}$$

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \tag{1.8}$$

respectively, where $\|.\|$ means the Euclidean norm and $y_{k-1} = g_k - g_{k-1}$. The PRP and HS methods with the exact line search are not globally convergent; see the counter example of Powell [18]. Recently, Dai and Yuan (DY) [6] proposed a nonlinear conjugate gradient method, which has the form (1.2) and (1.3) with

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}\cdot \tag{1.9}$$

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 (1.4) and (1.5). By direct calculation, we can deduce an equivalent form for β_k^{DY}, namely

$$\beta_k^{DY} = \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}\cdot \tag{1.10}$$

In [8], Dai and Yuan proposed a family of globally convergent conjugate methods, in which

$$\beta_k = \frac{\|g_k\|^2}{\lambda \|g_{k-1}\|^2 + (1-\lambda)(d_{k-1}^T y_{k-1})}, \tag{1.11}$$

where $\lambda \in [0, 1]$ is a parameter, and proved that the family of methods using line searches that satisfy (1.4) and

$$\sigma_1 g_k^T d_k \leq g(x_k + \alpha_k d_k)^T d_k \leq -\sigma_2 g_k^T d_k, \tag{1.12}$$

converges globally if the parameters σ_1, σ_2, and λ are such that

$$\sigma_1 + \sigma_2 \leq \lambda^{-1}. \tag{1.13}$$

In addition, Sellami *et al.* [19] proposed a new two-parameter family of conjugate gradient methods, in which

$$\beta_k = \frac{(1-\lambda_k)\|g_k\|^2 + \lambda_k(-g_k^T d_k)}{(1-\lambda_k-\mu_k)\|g_{k-1}\|^2 + (\lambda_k+\mu_k)(-g_{k-1}^T d_{k-1})}, \tag{1.14}$$

where $\lambda_k \in [0, 1]$ and $\mu_k \in [0, 1-\lambda_k]$ are parameters, and proved that the two-parameter family can ensure a descent search direction at every iteration and converges globally under line search condition (1.4) and (1.12) where the scalars σ_1 and σ_2 satisfy the condition

$$\sigma_1 + \sigma_2 \leq \frac{1+\mu_k\sigma_1}{1-\lambda_k}.\cdot \tag{1.15}$$

Further in [7], Dai and Yuan proposed a three-parameter family of conjugate gradient methods whose β_k is defined as

$$\beta_k = \frac{(1-\lambda_k)\|g_k\|^2 + \lambda_k g_k^T y_{k-1}}{(1-\mu_k-\omega_k)\|g_{k-1}\|^2 + \mu_k d_{k-1}^T y_{k-1} - \omega_k d_{k-1}^T g_{k-1}}, \tag{1.16}$$

where $\lambda_k \in [0,1]$, $\mu_k \in [0,1]$ and $\omega_k \in [0, 1-\mu_k]$ are parameters. With Powell's restart criterion, namely,

$$g_k^T g_{k-1} \le \xi \|g_k\|^2, \tag{1.17}$$

where $\xi > 0$ is some positive constant, which produces a descent search direction at every iteration and converges globally provided that the line search satisfies Wolfe conditions (1.4) and (1.5). Moreover, a well-established survey of development of different versions of nonlinear conjugate gradient methods, with special attention to global convergence properties, is presented by Hager and Zhang [12]. This family of algorithms includes a lot of variants, well known in the literature, with important convergence properties and numerical efficiency. Motivated by the good numerical performances of PRP method and the nice global-convergence properties of DY and HS methods, we can combine the previous methods into one unified method to ensure a descent search direction at every iteration and converges globally under the Wolfe line search conditions this would allows us to obtain the best numerical results that can outperform the classical methods. This paper gives a new conjugate gradient method for large-scale unconstrained optimization. We observe that the formulas (1.7) and (1.8) share the same numerators meanwhile (1.8) and (1.9) share the same denominators. We can use combinations of these numerators and denominators to obtain the following new formula which is given by

$$\beta_k^* = \frac{\lambda_k \|g_k\|^2 + (1-\lambda_k) g_k^T y_{k-1}}{(1-\lambda_k-\mu_k)\|g_{k-1}\|^2 + (\lambda_k+\mu_k)(y_{k-1}^T d_{k-1})}, \tag{1.18}$$

where $\lambda_k \in [0,1]$ and $\mu_k \in [0, 1-\lambda_k]$ are parameters.

We see that the above formula for β_k^* is special form of

$$\beta_k^* = \frac{\phi_k}{\phi'_{k-1}}, \tag{1.19}$$

where ϕ_k satisfies that

$$\phi_k = \lambda_k \|g_k\|^2 + (1-\lambda_k) g_k^T y_{k-1}, \tag{1.20}$$

and

$$\phi'_{k-1} = (1-\lambda_k-\mu_k)\|g_{k-1}\|^2 + (\lambda_k+\mu_k)(y_{k-1}^T d_{k-1}). \tag{1.21}$$

It is clear that the formula (1.21) is a generalization of the three previous methods.

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 (1.2)–(1.3) with β_k^* defined by (1.19). Section 4 includes the main convergence properties of the new method with Wolfe line search. The preliminary numerical results are described in Section 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 $\pounds = \{x \in R^n; f(x) \le f(x_1)\}$;
(ii) In some neighborhood $\mathcal{N}$ of $\pounds$, f is differentiable and its gradient g is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(\tilde{x})\| \le L\|x - \tilde{x}\|, \qquad \text{for all } x, \tilde{x} \in \mathcal{N}. \tag{2.1}$$

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

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

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

$$\|g(x)\| \leq \gamma, \qquad \text{for all } x \in \pounds. \tag{2.2}$$

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 [23].

Lemma 2.3. *Suppose Assumption* 2.1 *holds. Let* $\{x_k\}$*be generated by* (1.2) *and* d_k *satisfy* $g_k^T d_k < 0$. *If* α_k *is determined by the Wolfe line search* (1.4) *and* (1.5), *then we have*

$$\sum_{k\geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{2.3}$$

In the latter section, we need the following lemmas, the first of which is derived from [2], 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*

$$\sum_{k\geq 1} a_k = \infty, \tag{2.4}$$

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

$$b_k \leq c_1 + c_2 \sum_{i=1}^{k} a_i, \tag{2.5}$$

then we have that

$$\sum_{k\geq 1} \frac{a_k}{b_k} = \infty. \tag{2.6}$$

Lemma 2.5. *Consider the following 1-dimensional function,*

$$\rho(t) = \frac{a+bt}{c+dt}, \quad t \in R^1, \tag{2.7}$$

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

$$bc - ad > 0, \tag{2.8}$$

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

$$bc - ad < 0, \tag{2.9}$$

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

3. Algorithm

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

Algorithm 3.1.

Step 0: Given $x_1 \in R^n$, set $d_1 = -g_1$, $k = 1$. If $g_1 = 0$, then stop.
Step 1: Find $\alpha_k > 0$ satisfying the Wolfe conditions (1.4) and (1.5).
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 (1.21) and generate d_{k+1} by (1.3).
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

$$r_k = -\frac{g_k^T d_k}{\phi_k}, \tag{3.1}$$

and

$$t_k = \frac{\|d_k\|^2}{\phi_k^2}. \tag{3.2}$$

Lemma 3.2. *For the method* (1.2) *and* (1.3) *with* β^*_k *defined by* (1.19),

$$t_k = 2\sum_{i=1}^{k} \frac{r_i}{\phi_i} - \sum_{i=1}^{k} \frac{\|g_i\|^2}{\phi_i^2}, \tag{3.3}$$

holds for all $k \geq 1$.

Proof. Since $d_1 = -g_1$, (3.3) holds for $k = 1$. For $i \geq 2$, it follows from (1.3) that

$$d_i + g_i = \beta^*_i d_{i-1}. \tag{3.4}$$

Squaring both sides of the above equation, we get that

$$\|d_i\|^2 = -\|g_i\|^2 - 2g_i^T d_i + \beta_i^{*2} \|d_{i-1}\|^2. \tag{3.5}$$

Dividing (3.5) by ϕ_i^2 and applying (1.19) and (3.2),

$$t_i = \frac{\|d_{i-1}\|^2}{{\phi'_{i-1}}^2} + 2\frac{r_i}{\phi_i} - \frac{\|g_i\|^2}{\phi_i^2}. \tag{3.6}$$

Using (1.20)–(1.21) and since, $d_1 = -g_1$ we get that

$$\frac{\|d_1\|^2}{{\phi'_1}^2} = \frac{\|g_1\|^2}{\|g_1\|^4} = \frac{\|g_1\|^2}{\phi_1^2}. \tag{3.7}$$

Summing the above expression (3.6) over i, we obtain

$$t_k = t_1 + 2\sum_{i=2}^{k} \frac{r_i}{\phi_i} - \sum_{i=2}^{k} \frac{\|g_i\|^2}{\phi_i^2}. \tag{3.8}$$

Since $d_1 = -g_1$ and $t_1 = \dfrac{\|g_1\|^2}{\phi_1^2}$, the above relation is equivalent to (3.3). So (3.3) holds for $k \geq 1$. □

Theorem 3.3. *Suppose that x_1 is a starting point for which Assumption* 2.1 *holds. Consider the method* (1.2)–(1.3) *and* (1.19)*, if for all k, d_k satisfy $g_k^T d_k < 0$ and α_k is determined by the Wolfe line search* (1.4) *and* (1.5)*, and if*

$$\sum_{k\geq 1} r_k^2 = \infty, \tag{3.9}$$

then we have

$$\lim_{k\longrightarrow\infty} \inf \|g_k\| = 0. \tag{3.10}$$

Proof. (1.3) can be re-written as

$$g_i^T d_i + \|g_i\|^2 = \beta_i^* g_i^T d_{i-1}. \tag{3.11}$$

Squaring both sides of the above equation, we get that

$$-2g_i^T d_i - \|g_i\|^2 \leq \frac{(g_i^T d_i)^2}{\|g_i\|^2}, \tag{3.12}$$

dividing (3.12) by ϕ_i^2 and applying (3.3)

$$t_k \leq \sum_{i=1}^{k} \frac{r_i^2}{\|g_i\|^2}. \tag{3.13}$$

We proceed by contradiction. Assume that

$$\lim_{k\longrightarrow\infty} \inf \|g_k\| \neq 0. \tag{3.14}$$

Then there exists a positive constant γ such that

$$\|g_k\|^2 \geq \gamma, \qquad \text{for all } k. \tag{3.15}$$

We can see from (3.13) that,

$$t_k \leq \frac{1}{\gamma^2} \sum_{i=1}^{k} r_i^2. \tag{3.16}$$

The above relation, (3.9) and Lemma 2.4, yield

$$\sum_{i\geq 1} \frac{r_i^2}{t_i} = \infty. \tag{3.17}$$

Thus, by the definition (3.1) and (3.2), we know that (3.17) contradicts (2.3). This concludes the proof. □

Theorem 3.4. *Suppose that x_1 is a starting point for which Assumption* 2.1 *holds. Consider the method* (1.2)–(1.3) *and* (1.19)*, if for all k, d_k satisfy $g_k^T d_k < 0$ and α_k is determined by the Wolfe line search* (1.4) *and* (1.5)*, and if*

$$\sum_{k\geq 1} \frac{\|g_k\|^2}{\phi_k^2} = \infty, \tag{3.18}$$

then we have

$$\lim_{k\longrightarrow\infty} \inf \|g_k\| = 0. \tag{3.19}$$

Proof. Noting that

$$t_k \geq 0 \qquad \text{for all } k, \tag{3.20}$$

we can get from (3.3)

$$2\sum_{i=1}^{k} \frac{r_i}{\phi_i} \geq \sum_{i=1}^{k} \frac{\|g_i\|^2}{\phi_i^2}. \tag{3.21}$$

Direct calculation show that,

$$4\sum_{i=1}^{k} \frac{r_i^2}{\|g_i\|^2} \geq 4\sum_{i=1}^{k} \frac{r_i}{\phi_i} - 2\sum_{i=1}^{k} \frac{\|g_i\|^2}{\phi_i^2} \geq 0. \tag{3.22}$$

Or equivalently,

$$4\sum_{i=1}^{k} \frac{r_i^2}{\|g_i\|^2} \geq 4\sum_{i=1}^{k} \frac{r_i}{\phi_i} - \sum_{i=1}^{k} \frac{\|g_i\|^2}{\phi_i^2} \geq \sum_{i=1}^{k} \frac{\|g_i\|^2}{\phi_i^2}. \tag{3.23}$$

Thus if (3.18) holds, we also have that

$$\sum_{k\geq 1} \frac{(g_k^T d_k)^2}{\|g_k\|^2 \phi_k^2} = \infty. \tag{3.24}$$

Because (3.13) still holds, it follows from (3.24), the definition of r_k and Lemma 2.4, that

$$\sum_{k\geq 1} \frac{(g_k^T d_k)^2}{\|g_k\|^2 \|d_k\|^2} = \infty. \tag{3.25}$$

The above relation and Lemma 2.3 clearly give (3.10). This completes our proof. □

Thus we have proved two convergence theorems for the general method (1.2) and (1.3) with β_k^* defined by (1.19).

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

$$g_k^T d_k \leq -c\|g_k\|^2, \tag{3.26}$$

where c is a positive constant, is not invoked in Theorems 3.2 and 3.3. The sufficient descent condition (3.26) was often used or implied in the previous analysis of conjugate gradient methods (see [1,2,11]). This condition has been relaxed to the descent condition ($g_k^T d_k < 0$) in the convergence analysis [6] of the FR method and the convergence analysis [9] of any conjugate gradient method.

4. Global convergence of new conjugate gradient method

In this section, we establish some global convergence of the new method (1.19) under certain line searches conditions.

We consider the method (1.2) and (1.3) with ϕ_k satisfying

$$\phi_k = \lambda_k \|g_k\|^2 + (1-\lambda_k) g_k^T y_{k-1}, \tag{4.1}$$

where $\lambda_k \in [0,1]$. (4.1) and (1.3) show that

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \beta_k^* g_k^T d_{k-1} \\ &= -\|g_k\|^2 + \frac{\lambda_k \|g_k\|^2 + (1-\lambda_k) g_k^T y_{k-1}}{(1-\lambda_k-\mu_k)\|g_{k-1}\|^2 + (\lambda_k+\mu_k) y_{k-1}^T d_{k-1}} g_k^T d_{k-1}. \end{aligned} \tag{4.2}$$

The above relation imply that

$$g_k^T d_k = -\frac{(1-\lambda_k-\mu_k)\left\|g_{k-1}\right\|^2 - \lambda_k g_k^T d_{k-1} - (1-2\lambda_k-\mu_k) y_{k-1}^T d_{k-1}}{(1-\lambda_k-\mu_k)\left\|g_{k-1}\right\|^2 + (\mu_k+\lambda_k) y_{k-1}^T d_{k-1}} \left\|g_k\right\|^2. \tag{4.3}$$

Thus by (4.2), we deduce an equivalent form of β_k^*,

$$\beta_k^* = \frac{\lambda_k g_{k-1}^T d_{k-1} + y_{k-1}^T d_{k-1}}{(1-\lambda_k-\mu_k)\left\|g_{k-1}\right\|^2 + \mu_k(-d_{k-1}^T g_{k-1}) + \lambda_k(y_{k-1}^T d_{k-1})} \frac{\left\|g_k\right\|^2}{g_k^T d_{k-1}}. \tag{4.4}$$

Substituting (4.3) into (4.1), we obtain that

$$\phi_k = \frac{\lambda_k g_{k-1}^T d_{k-1} + y_{k-1}^T d_{k-1}}{g_k^T d_{k-1}} \left\|g_k\right\|^2. \tag{4.5}$$

By this relation, we can show an important property of ϕ_k under Wolfe line searches and hence obtain the global convergence of the new method (4.4) 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* (1.2)–(1.3)–(1.19) *and* (4.1), *if $g_k^T d_k < 0$ for all $k$and α_k is computed by the Wolfe line search* (1.4) *and* (1.5), *then we have*

$$\frac{\phi_k}{\left\|g_k\right\|^2} \leq (1-\lambda_k-\sigma)^{-1}. \tag{4.6}$$

Further, the method converges in the sense that

$$\lim_{k\longrightarrow\infty} \inf \left\|g_k\right\| = 0. \tag{4.7}$$

Proof. Since (1.5), we have that

$$g(x_k+\alpha_k d_k)^T d_k \geq \sigma g_k^T d_k. \tag{4.8}$$

By direct calculation, it shows that

$$d_{k-1}^T y_{k-1} \geq (1-\sigma)(-g_{k-1}^T d_{k-1}). \tag{4.9}$$

Dividing (4.5) by $\left\|g_k\right\|^2$ and applying (4.9) implies the truth of (4.6). Therefore, by (2.2) and (4.9) we have that

$$\sum_{k\geq 1} \frac{\left\|g_k\right\|^2}{\phi_k^2} = \infty. \tag{4.10}$$

Thus (3.10) follows from Theorem 3.3. □

In the following, we can show that, for any $\lambda_k \in (0,1]$, the method (1.2)–(1.3)–(1.19) and (4.1) ensures the descent property of each search direction and converges globally under line search condition (1.4) and (1.12) where the scalars σ_1 and σ_2 satisfy certain condition. For this purpose, we define

$$\overline{r}_k = -\frac{g_k^T d_k}{\left\|g_k\right\|^2}, \tag{4.11}$$

and

$$l_k = \frac{g_{k+1}^T d_k}{g_k^T d_k}, \tag{4.12}$$

it is obvious that d_k is a descent direction if and only if $\overline{r}_k > 0$. For The above relation, (4.3) and (4.12), we can write

$$\overline{r}_k = \frac{(1-\lambda_k-\mu_k) - [\mu_k(1-l_{k-1}) + \lambda_k(2-l_{k-1}) + l_{k-1} - 1]\,\overline{r}_{k-1}}{(1-\lambda_k-\mu_k) + [(\mu_k+\lambda_k)(1-l_{k-1})]\,\overline{r}_{k-1}}. \tag{4.13}$$

The following theorem indicates that, if α_k is computed by the Wolfe line search (1.4) and (1.12), then the search direction d_k satisfies the descent property.

Theorem 4.2. *Suppose that x_1 is a starting point for which Assumption* 2.1 *holds. Consider the method* (1.2)–(1.3)–(1.19) *and* (4.1)*, where $\lambda_k \in [0,1)$, $\mu_k \in [0, 1-\lambda_k]$ and α_k satisfies the line search conditions* (1.4) *and* (1.12) *with $\frac{1}{2} \le \sigma_1 \le 1$ and $\sigma_2 > 0$, if*

$$\sigma_1 - \sigma_2 \le \frac{2(1-\mu_k) - 3\lambda_k}{1-\lambda_k-\mu_k}, \tag{4.14}$$

then we have for all $k \ge 1$

$$0 < \bar{r}_k < (1-\sigma_1)^{-1}. \tag{4.15}$$

Proof. The right hand side of (4.13) is a function of λ_k, μ_k, l_{k-1} and $\overline{r}_{k-1}$, which is denoted as $h(\lambda_k, \mu_k, l_{k-1}, \bar{r}_{k-1})$. We prove (4.15) by induction. Noting that $d_1 = -g_1$ and hence $\bar{r}_1 = 1$, we see that (4.15) is true for $k = 1$. We now suppose that (4.15) holds for $k-1$, namely,

$$0 < \bar{r}_{k-1} < (1-\sigma_1)^{-1}. \tag{4.16}$$

It follows from (1.12)

$$-\sigma_2 \le l_{k-1} \le \sigma_1. \tag{4.17}$$

Then by Lemma 2.5, the fact that $\lambda_k \in (0,1]$ and the fact that $\mu_k \ge 0$, we get that

$$\begin{aligned} \bar{r}_k &\le h(\lambda_k, \mu_k, \sigma_1, \bar{r}_{k-1}) < h(\lambda_k, \mu_k, \sigma_1, (1-\sigma_1)^{-1}) \\ &= 1 + \frac{1-\lambda_k-\sigma_1}{(1-2\lambda_k-2\mu_k)(1-\sigma_1)} \\ &\le 1 + \frac{1-\sigma_1}{(1-2\lambda_k-2\mu_k)(1-\sigma_1)} \\ &\le 1 + \frac{\sigma_1}{1-\sigma_1} = (1-\sigma_1)^{-1}. \end{aligned} \tag{4.18}$$

On the other hand, by Lemma 2.5 and relation (4.14), we also have that

$$\begin{aligned} \bar{r}_k &\ge h(\lambda_k, \mu_k, -\sigma_2, \bar{r}_{k-1}) > h(\lambda_k, \mu_k, -\sigma_2, (1-\sigma_1)^{-1}) \\ &= 1 + \frac{1+\sigma_2-\lambda_k}{(1-\sigma_1)(1-\lambda_k-\mu_k) - (\mu_k+\lambda_k)(1+\sigma_2)} \ge 0. \end{aligned} \tag{4.19}$$

Thus (4.15) is true for k, by induction, (4.15) holds for $k \ge 1$, and hence the descent property

$$g_k^T d_k < 0, \forall k$$

holds, as long as $g_k \neq 0$. □

By Theorem 4.2, we can immediately give the following convergence result for new conjugate gradient method (1.19).

Theorem 4.3. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the method (1.2)–(1.3)–(1.19) and (4.1), where $\lambda_k \in [0,1)$, $\mu_k \in [0, 1-\lambda_k]$. If the step length α_k is computed by the Wolfe line search (1.4) and (1.12) with $\frac{1}{2} \leq \sigma_1 \leq 1$ and $\sigma_2 > 0$, and if the scalars σ_1 and σ_2 satisfy the condition (4.14), then the method converges in the sense that,*

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

Proof. To show the truth of (3.10), by Theorem 3.2, it suffices to prove that

$$\max\{\bar{r}_{k-1}, \bar{r}_k\} \geq c_1, \tag{4.20}$$

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

$$\bar{r}_{k-1} \leq 1, \tag{4.21}$$

by Lemma 2.5, the fact that $\lambda_k \in (0,1]$ and the fact that $\mu_k \geq 0$, we can get that

$$\bar{r}_k \geq h(\lambda_k, \mu_k, -\sigma_2, 1) \stackrel{\Delta}{=} c_2. \tag{4.22}$$

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

$$\max\{\bar{r}_{k-1}, \bar{r}_k\} \geq c_2, \tag{4.23}$$

for all $k \geq 2$. By the definition (3.1) of r_k and relation (4.1), we have that

$$r_k = \frac{\bar{r}_k}{1 + (1-\lambda_k)\eta_k}, \tag{4.24}$$

where $\eta_k = -\frac{g_k^T g_{k-1}}{\|g_k\|^2}$. Which, with (4.23) and Lemma 2.5, implies that (4.20) holds with $c_1 = c_2$. This completes our proof. □

Thus we have some general convergence results achieved for the new method (4.4). It is easy to see from (4.4) that the new method include the three nonlinear conjugate gradient methods mentioned above. Letting $\lambda_k \equiv 0$ and $\mu_k \equiv 0$, from Theorem 4.2, we again obtain the convergence result of the PRP method in [16,17]. For the case when $\lambda_k \equiv 1$ and $\mu_k \equiv 0$ the method is proved to generate a descent search direction at every iteration and converges globally provided that the stepsize satisfies the Wolfe conditions (1.4) and (1.5) (see [5]). If $\lambda_k \equiv 0$ and $\mu_k \equiv 1$, then the method ensures a descent direction for general functions and is proved to global convergence under strong Wolfe line search (1.4) and (1.6) of the method HS (see [13]).

5. Numerical experiments

In this section, we report some numerical results obtained with the new proposed conjugated gradient method. The code is written in Fortran and compiler settings on the PC machine (AMD, 1.61 GHZ, 960 M memory) with Windows operation system. There are a number of 68 large-scale unconstrained optimization test problems in generalized or extended from CUTE [3,4] collection. For each test function we have taken seven numerical experiments with the number of variables increasing as $n = 1000, 2000, 3000, 4000, 5000, 8000, 10\,000$.

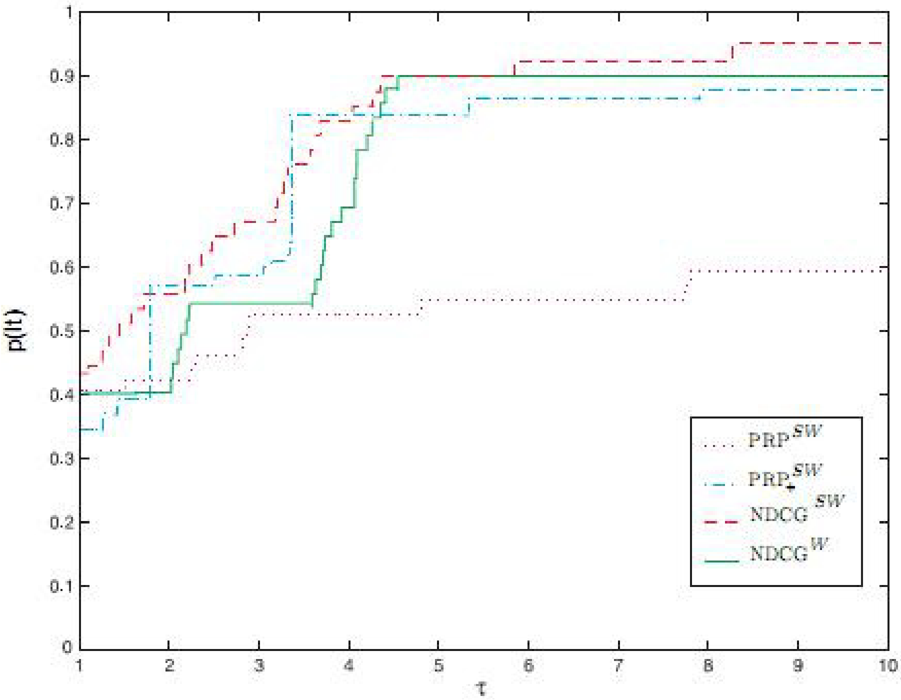

FIGURE 1. Performance files based on iterations.

We adopt the performance profiles by Delan and Moré [15] to compare the performance between 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$_+^{SW}$: the PRP method with nonnegative values of $\beta_k = \max\left\{0, \beta_k^{PRP}\right\}$, proposed by Powell [18] and analysed by Gilbert and Nocedal [11], under the strong Wolfe conditions, where $\delta = 10^{-2}$ and $\sigma = 0.1$.
- NDCG^{SW}: Algorithm 3.1 with the Wolfe conditions (1.4) and (1.12), where the scalars σ_1 and σ_2 satisfy the condition (4.14), in addition, $\delta = 10^{-2}$, $\sigma_1 = \sigma_2 = \sigma = 0.1$, $\lambda_k = \lambda = 0.5$ and $\mu_k = \mu = 0.4$.
- NDCG^{W}: Algorithm 3.1 with the standard Wolfe conditions, where $\delta = 10^{-2}$, $\sigma = 0.1$, $\lambda_k = \lambda = 0.5$ and $\mu_k = \mu = 0.5$

During our experiments, the strategy for the initial step length is to assume that the first-order change in the function at iteration x_k will be the same as that obtained at the previous step [17]. In other words, we choose the initial guess α_0 satisfying

$$\alpha_0 = \alpha_{k-1}\frac{\Psi_{k-1}}{\Psi_k} \qquad\qquad \forall k > 1,$$

where $\Psi_k = g_k^T d_k$, when $k = 1$, we choose $\alpha_0 = \dfrac{1}{\|g(x_1)\|}$. 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 $\|g_k\|_\infty \leq 10^{-5}$, where $\|.\|_\infty$ is the maximum absolute component of a vector. Figures 1–3 give performance profiles of the four methods for the number of iterations, function and gradient evaluations, and the CPU time, respectively.

From the above three figures, we can see that all the methods are efficient. The new method NDCG performs better than the PRP^{SW} and PRP$_+^{SW}$ methods, for the given test problems. These obtained preliminary results are indeed encouraging.

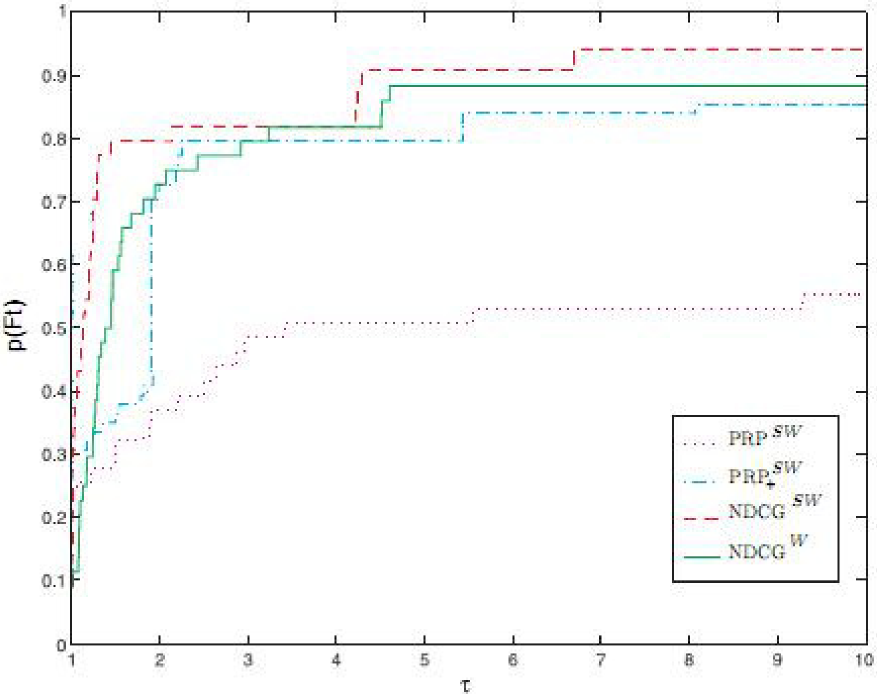

FIGURE 2. Performance files based on function and gradient evaluations.

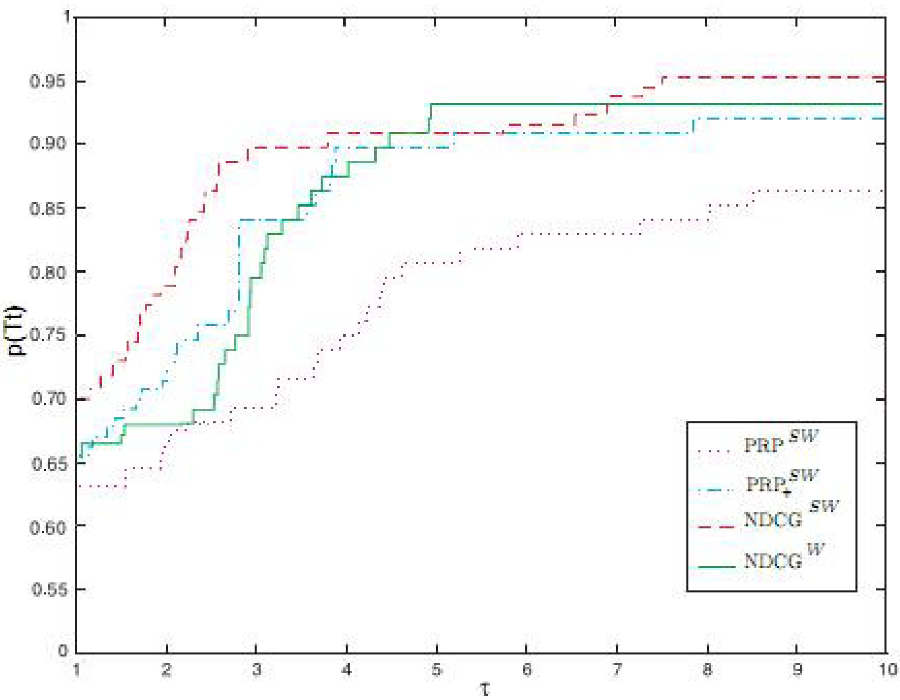

FIGURE 3. Performance files based on CPU time.

In the performance profile plot, the top curve corresponds to the method that solved the most problems in time that is within a factor τ of the best time. The percentage of the test problems for which a method is reported as the fastest is given on the left axis of the plot. The right side of the plot gives the percentage of the test problems that were successfully solved by each of the methods. In essence, the right side is a measure of the algorithm's robustness.

In Figure 1, we compare the performance based on number of iterations. Since the top curve in Figure 1 corresponds to NDCG^{SW}, this algorithm is clearly the fastest for this set of 68 test problems. In particular, NDCG^{SW} is fastest for about 43% (29 out of 68) of the test problems, and it ultimately solves 95% of the test problems. Notice that relative to the number of iterations, NDCG^{W} and PRP^{SW} have almost identical performance (fastest for 27 problems) achieving better than PRP$_{+}^{SW}$ (fastest for 24 problems). Also, it is interesting to observe in Figure 1 that the NDCG^{SW} and NDCG^{W} codes are the top performer, relative to the iteration metric, for values of $\tau \geq 4.5$.

In Figure 2, we compare performance based on the number of function and gradient evaluations. For our CUTEr test set, we found that, on average, the CPU time to evaluate the derivative of f was about 2 times the CPU time to evaluate f itself. Figure 2 gives the performance profiles based on the metric NF + 2NG, where NF is the number of function evaluations and NG is the number of gradient evaluations. Relative to this metric, NDCG^{SW} achieves the top performance (fastest for 32%), followed by PRP$_{+}^{SW}$ (fastest for 30%), then PRP^{SW} (fastest for 25%), and then NDCG^{W} (fastest for 11%). Also, it is interesting to observe in Figure 2 that the NDCG^{SW} and NDCG^{W} codes are the top performers, relative to the number of function and gradient evaluations, for values of $\tau \geq 5$.

In Figure 3, we use CPU time to compare the performance of the conjugate gradient codes NDCG^{SW}, NDCG^{W}, PRP^{SW}, and PRP^{SW}. Figure 3 indicates that, relative to the CPU time metric, NDCG^{SW} is fastest (fastest for 70%), then PRP$_{+}^{SW}$ and NDCG^{W} almost identical performance (fastest for 65%), and then PRP^{SW} (fastest for 63%). Hence, NDCG^{SW} and NDCG^{W} codes are the top performers, relative to the CPU time metric, for values of $\tau \geq 7$.

In conclusion, Figures 1–3 suggest that our proposed method NDCG exhibits the best overall performance since it illustrates the highest probability of being the optimal solver, followed by the PRP^{SW} and PRP$_{+}^{SW}$ conjugate gradient methods relative to all performance metrics.

6. Conclusions and discussions

In this paper, we have proposed a new conjugate gradient method, and studied the global convergence of these method. 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 (1.19) with weak Wolfe line search (1.4) and (1.5), namely, Theorems 3.2, 3.3 and 4.1. Next, from Theorems 4.2 and 4.3, we proved that the new method (1.19) can ensure a descent search direction at every iteration and converges globally provided that the stepsize satisfies the Wolfe conditions (1.4) and (1.12) where the scalars σ_1 and σ_2 satisfy the condition (4.14). In summary, our computational results show that this new descent nonlinear conjugate gradient method, namely NDCG method not only converges globally, but also outperforms the original PRP method. 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 find the practical performance of the method (4.4).

Appendix A

The following table lists the names of the 68 test problems.

TABLE A.1. Test problems.

n	Problem	n	Problem	n	Problem	n	Problem
1	ARWHEAD	2	BD1	3	BDEXP	4	BEALE
5	BIGGSB1	6	BROWNAL	7	BROYDN7D	8	COSINE
9	CRAGGLVY	10	DENSCHNB	11	DENSCHNF	12	DIXMAANA
13	DIXMAANB	14	DIXMAANG	15	DIXMAANE	16	DIXMAANF
17	DIXMAANG	18	DIXMAANI	19	DIXMAANJ	20	DIXMAANK
21	DIXMAANL	22	DIXON3DQ	23	DQDRTIC	24	DQRTIC
25	EDENSCH	26	EG2	27	ENGVAL1	28	FLETCHCR
29	FREUROTH	30	GHUMPS	31	GROSEN	32	GPSC1
33	HIEBERT	34	HIMMELBLAU	35	LIARWHD	36	MARATOS
37	NONCVXU2	38	NONDIA	39	NONDQUAR	40	PENALTY1
41	PENALTY	42	POWELL	43	POWELLBS	44	POWELLSG
45	POWER	46	PPQ2	47	PSC1	48	QF1
49	QF2	50	QP1	51	QP2	52	QUARTC
53	RAYDAN1	54	RAYDAN2	55	ROSEN	56	SINQUAD
57	SQ1	58	SQ2	59	SROSENBR	60	TRIDIA
61	WHITEHOLST	62	WOODS	63	TRIDIAG1	64	TRIDIAG2
65	EXTRIGON	66	GTRIDIAGI	67	DIAG2	68	CLIFF

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

REFERENCES

[1] M. Al-Baali, Descent property and global convergence of the fletcherreeves method with inexact line search. *IMA J. Numer. Anal.* **5** (1985) 121–124.
[2] M. Al-Baali, New property and global convergence of the fletcher-reeves method with inexact line searches. *IMA J. Numer. Anal.* **5** (1985) 122–124.
[3] N. Andrei, An unconstrained optimization test functions collection. *Adv. Model. Optim* **10** (2008) 147–161.
[4] I. Bongartz, A.R. Conn, N. Gould and P.L. Toint, Cute: Constrained and unconstrained testing environment. *ACM Trans. Math. Software (TOMS)* **21** (1995) 123–160.
[5] Y. Dai and Y. Yuan, Some properties of a new conjugate gradient method. In *Advances in Nonlinear Programming*. Springer (1998) 251–262.
[6] Y. Dai and Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property. *SIAM J. Optim.* **10** (1999) 177–182.
[7] Y. Dai and Y. Yuan, A three-parameter family of nonlinear conjugate gradient methods. *Math. Comput.* **70** (2001) 1155–1167.
[8] Y. Dai and Y. Yuan, A class of globally convergent conjugate gradient methods. *Sci. China Ser. A: Math.* **46** (2003) 251–261.
[9] Y. Dai, J. Han, G. Liu, D. Sun, H. Yin and Y. Yuan, Convergence properties of nonlinear conjugate gradient methods. *SIAM J. Optim.* **10** (2000) 345–358.
[10] R. Fletcher and C.M., Reeves, Function minimization by conjugate gradients. *Comput. J.* **7** (1964) 149–154.
[11] J.C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization. *SIAM J. Optim.* **2** (1992) 21–42.
[12] W.W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods. *Pacific J. Optim.* **2** (2006) 35–58.
[13] M. Hestenes, Methods of conjugate gradients for solving linear systems. *Res. J. Natl. Bur. Stan.* **49** (1952) 409–436.
[14] Y. Hu and C. Storey, Global convergence result for conjugate gradient methods. *J. Optim. Theory Appl.* **71** (1991) 399–405.
[15] J. Moré and E.D. Dolan, Benchmarking optimization software with performance files. *Math. Program.* **91** (2002) 201–2013.
[16] E. Polak and G. Ribiere, Note sur la convergence de méthodes de directions conjuguées. *Revue française d'informatique et de recherche opérationnelle, série rouge* **3** (1969) 35–43.
[17] B.T. Polyak, The conjugate gradient method in extremal problems. *USSR Comput. Math. Math. Phys.* **9** (1969) 94–112.
[18] M. Powell, Nonconvex minimization calculations and the conjugate gradient method. *Numer. Anal.* (1984) 122–141.
[19] B. Sellami, Y. Laskri and R. Benzine, A new two-parameter family of nonlinear conjugate gradient methods. *Optimization* **64** (2015) 993–1009.
[20] D.F. Shanno, Conjugate gradient methods with inexact searches. *Math. Oper. Res.* **3** (1978) 244–256.
[21] P. Wolfe, Convergence conditions for ascent methods. *SIAM Rev.* **11** (1969) 226–235.
[22] P. Wolfe, Convergence conditions for ascent methods. ii: Some corrections. *SIAM Rev.* **13** (1971) 185–188.
[23] G. Zoutendijk, Nonlinear programming, computational methods. In *Integer and Nonlinear Programming*, edited by J. Abadie. North-Holland, Amsterdam (1970) 37–86.