RAIRO-Oper. Res. 51 (2017) 1101–1117
DOI: 10.1051/ro/2017028

RAIRO Operations Research
www.rairo-ro.org

GLOBALLY CONVERGENCE OF NONLINEAR CONJUGATE GRADIENT METHOD FOR UNCONSTRAINED OPTIMIZATION

B. Sellami[1], M. Belloufi[1] and Y. Chaib[1]

Abstract. The conjugate gradient method is a useful and powerful approach for solving large-scale minimization problems. In this paper, a new nonlinear conjugate gradient method is proposed for large-scale unconstrained optimization. This method include the already existing two practical nonlinear conjugate gradient methods, to combine the nice global convergence properties of Fletcher-Reeves method (abbreviated FR) and the good numerical performances of the Polak–Ribiére–Polyak method (abbreviated PRP), which produces a descent search direction at every iteration and converges globally provided that the line search satisfies the Wolfe conditions. Our numerical results show that of the new method is very efficient for the given test problems. In addition we will study the methods related to the new nonlinear conjugate gradient method.

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

Received January 9, 2017. Accepted April 5, 2017.

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 a class of 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, 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.

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; mbelloufi;chaibyacine@yahoo.com

The strong Wolfe conditions, namely,

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k g_k^T d_k \quad (1.4)$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \le -\sigma g_k^T d_k, \quad (1.5)$$

where $0 < \delta < \sigma < 1$. 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 [7, 10, 12, 15, 18]). The best-known formulas for k are the following Fletcher-Reeves (FR) and Polak-Ribière-Polyak (PRP):

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \quad (1.6)$$

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \quad (1.7)$$

where $\|.\|$ means the Euclidean norm and $y_{k-1} = g_k - g_{k-1}$. For non-quadratic objective functions, the global convergence of (FR) method was proved when the exact line search or strong Wolfe line search [2, 9] was used. However, if the condition (1.5) 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 [9], see the counter example of Powell [16].

Recently, Dai and Yuan (DY) [7] proposed a nonlinear conjugate gradient method, which has the form (1.2), (1.3) with

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} \cdot \quad (1.8)$$

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 (see [6]), namely, (1.4) and

$$\sigma g_k^T d_k \le g(x_k + \alpha_k d_k)^T d_k \cdot . \quad (1.9)$$

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

$$\beta_k = \frac{\|g_k\|^2}{\lambda \|g_{k-1}\|^2 + (1-\lambda)(d_{k-1}^T y_{k-1})}, \quad (1.10)$$

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

$$\sigma_1 g_k^T d_k \le g(x_k + \alpha_k d_k)^T d_k \le -\sigma_2 g_k^T d_k, \quad (1.11)$$

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

$$\sigma_1 + \sigma_2 \le \lambda^{-1}, \quad (1.12)$$

where $0 < \delta < \sigma_1 < 1$ and $\sigma_2 > 0$. In addition, Sellami *et al.* [17] proposed a new family of conjugate gradient methods, in which

$$\beta_k = \frac{(1-\lambda_k)\|g_k\|^2 + \lambda_k(-g_k^T d_k)}{(1-\lambda_k-\mu_k)\|g_{k-1}\|^2 + (\lambda_k+\mu_k)(-g_{k-1}^T d_{k-1})}, \quad (1.13)$$

where $\lambda_k \in [0, 1]$ and $\mu_k \in [0, 1-\lambda]$ are parameters, and proved that the new family can ensure a descent search direction at every iteration and converges globally under line search condition (1.4) and (1.11) where the scalars σ_1 and σ_2 satisfy the condition

$$\sigma_1 + \sigma_2 \le \frac{1 + \mu_k \sigma_1}{1 - \lambda_k} \cdot \quad (1.14)$$

Observing that the formula (1.6) and (1.7) share same denominators and two numerators, we can use combinations of these numerators and denominators to obtain the following new nonlinear conjugate gradient method

$$\beta_k^* = \frac{\lambda \|g_k\|^2 + (1-\lambda)(g_k^T y_{k-1})}{\lambda \|g_{k-1}\|^2 + (1-\lambda) \|g_{k-1}\|^2}. \tag{1.15}$$

Thus by the above equality in (1.14), we deduce an equivalent form of β_k^*,

$$\beta_k^* = \frac{\lambda \|g_k\|^2 + (1-\lambda)(g_k^T y_{k-1})}{\|g_{k-1}\|^2}, \tag{1.16}$$

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

$$\beta_k^* = \frac{\phi_k}{\phi'_{k-1}}, \tag{1.17}$$

where ϕ_k satisfies that

$$\phi_k = \lambda \|g_k\|^2 + (1-\lambda)(g_k^T y_{k-1}), \tag{1.18}$$

and

$$\phi'_{k-1} = \lambda \|g_{k-1}\|^2 + (1-\lambda) \|g_{k-1}\|^2 = \|g_{k-1}\|^2 . \tag{1.19}$$

It is clear that the formula (1.17) is a generalization of the two previous methods are defined by (1.6) and (1.7). 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 (1.17). Section 4 includes the main convergence properties of the new nonlinear conjugate gradient method with Wolfe line search, and we study methods related to the new nonlinear conjugate gradient method (1.17). The preliminary numerical results are contained in Section 5. Conclusions and discussions are made in the last section.

2. Preliminaries

For convenience, we assume that $g_k \neq 0$ for $k \geq 1$. We also assume that $g_0 = 0$ for $k = 0$, which gives us $y_0 = g_1 - g_0 = g_1$, 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 $\mathcal{N}$ of $\mathcal{L}$, f is differentiable and its gradient g is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

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

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

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

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

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

The conclusion of the following lemma, often called the Zoutendijk condition, is used to prove the global convergence of nonlinear conjugate gradient methods.

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), (1.9), *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 exist two constants c_1 *and* c_2 *such that for all* $k \geq 1$,

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

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 and convergence analysis

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 a $\alpha_k > 0$ satisfying the Wolfe conditions (1.4) and (1.9).
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 (1.17) 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 3.1. *For the method* (1.2), (1.3) *with* β_k^* *defined by* (1.17),

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

holds for all $k \geq 1$.

Proof. For $k = 1$, we have that $d_1 = -g_1$, then by the definition (3.2), we get that

$$t_1 = \frac{\|d_1\|^2}{\phi_1^2}.$$

Or equivalently,

$$t_1 = 2\frac{\|d_1\|^2}{\phi_1^2} - \frac{\|d_1\|^2}{\phi_1^2}.$$

By the definition (3.1) of r_k, the above relation is equivalent to (3.3) for $d_1 = -g_1$, so (3.3) holds. 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 (1.17) 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 (1.18), (1.19) 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 3.2. *Suppose that* x_1 *is a starting point for which Assumption* 2.1 *holds. Consider the method* (1.2), (1.3) *and* (1.17), *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* (1.9), *if*

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

we have that

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

Proof. (1.3) can be re-written as

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

Square $(g_i^T d_i + \|g_i\|^2)$ in order to obtain,

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

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. Assuming 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 3.3. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the method (1.2), (1.3) and (1.17), 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 (1.9), if*

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

we have that

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

Proof. Noting that

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

Squaring the left side of equation (3.12), we get that

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

Hence, we have

$$4(g_i^T d_i)^2 + \|g_i\|^4 + 4(g_i^T d_i)\|g_i\|^2 \geq 0. \tag{3.21}$$

Summing this expression over i and dividing (3.21) by $(\phi_i^2 \|g_i\|^2)$, we obtain

$$4\sum_{i=1}^{k}\frac{(g_i^T d_i)^2}{\phi_i^2\|g_i\|^2} \geq -4\sum_{i=1}^{k}\frac{(g_i^T d_i)}{\phi_i^2} - \sum_{i=1}^{k}\frac{\|g_i\|^2}{\phi_i^2}. \tag{3.22}$$

On the other hand, we can get from (3.3), (3.20) and the definition of r_k

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

Direct calculation show that,

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

The above relation (3.22) and (3.23) 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 (3.18) holds, we also have that

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

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), (1.3) with β_k^* defined by (1.17). 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, 11]). This condition has been relaxed to the descent condition ($g_k^T d_k < 0$) in the convergence analysis [7] of the FR method and the convergence analysis [5] of any conjugate gradient method.

4. Global convergence

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

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

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

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

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

By the definition of y_k, we obtain that

$$\left[g_k^T d_k + \|g_k\|^2\right] \|g_{k-1}\|^2 = \left[\lambda \|g_k\|^2 + (1-\lambda)(g_k^T g_k - g_k^T g_{k-1})\right] g_k^T d_{k-1}.$$

Which implies that

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

The above relation imply that

$$g_k^T d_k = -\frac{\|g_{k-1}\|^2 - g_k^T d_{k-1} + (1-\lambda) g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} \|g_k\|^2. \tag{4.3}$$

Thus by using (4.2) and (4.3) we obtain

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

The above relation imply that

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

The above form for β_k^* and relation (1.17), we obtain that

$$\phi_k = \frac{g_k^T d_{k-1} + (1-\lambda)(-g_{k-1}^T d_{k-1})}{g_k^T d_{k-1}} \|g_k\|^2. \tag{4.5}$$

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 nonlinear conjugate gradient method (4.4) under some assumptions.

Theorem 4.1. *Suppose that x_1 is a starting point for which Assumptions* 2.1 *and* 2.2 *hold. Consider the method* (1.2), (1.3), (1.17) *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), (1.9), *then*

$$\frac{\phi_k}{\|g_k\|^2} \leq (\sigma + \lambda - 1)\sigma^{-1}. \tag{4.6}$$

Further, the method converges in the sense that

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

Proof. Since (1.9), 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 calculations show that

$$1 + \frac{(1-\lambda)(-g_{k-1}^T d_{k-1})}{g_k^T d_{k-1}} \leq 1 + \frac{(1-\lambda)(-g_{k-1}^T d_{k-1})}{\sigma g_{k-1}^T d_{k-1}}. \tag{4.9}$$

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

$$\sum_{k \geq 1} \frac{\|g_k\|^2}{\phi_k^2} \geq \frac{\sigma^2}{(\sigma + \lambda - 1)^2 \|g_k\|^2} \geq \frac{\sigma^2}{(\sigma + \lambda - 1)^2 \gamma^2} = \infty. \tag{4.10}$$

Thus (3.10) follows from Theorem 3.3. □

In the following, we can show that, for any $\lambda \in (0, 1]$, the method (1.2), (1.3), (1.17) and (4.1) ensures the descent property of each search direction and converges globally under line search condition (1.4) and (1.11) where the scalars σ_1 and σ_2 satisfy certain condition. For this purpose, we define

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

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

$$\overline{r}_k = 1 + (l_{k-1} + \lambda - 1)\overline{r}_{k-1}. \tag{4.13}$$

Theorem 4.2. *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the method (1.2), (1.3), (1.17) and (4.1), where $\lambda \in (0, 1]$ and α_k satisfies the line search conditions (1.4) and (1.11). If $g_k^T g_{k-1} = 0$ for $k \geq 1$, and if the scalars σ_1 and σ_2 in (1.11) with $\sigma_1, \sigma_2 > 0$ and $\sigma_1 < 1$, is such that*

$$\sigma_1 + \sigma_2 \leq \lambda, \tag{4.14}$$

then we have for all $k \geq 1$

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

Further, the method converges in the sense that (3.10) is true.

Proof. The right hand side of (4.13) is a function of λ, l_{k-1} and $\overline{r}_{k-1}$, which is denoted as $h(\lambda, l_{k-1}, \overline{r}_{k-1})$. 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 (1.11)

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

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

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

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

$$\bar{r}_k \geq h(\lambda, -\sigma_2, \bar{r}_{k-1}) > h(\lambda, -\sigma_2, (1 - \sigma_1)^{-1}) = [-(\sigma_1 + \sigma_2) + \lambda](1 - \sigma_1)^{-1} \geq 0. \tag{4.19}$$

Thus (4.15) is true for k, by induction, (4.15) holds for $k \geq 1$.

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

$$\max\{r_{k-1}, r_k\} \geq c_1, \tag{4.20}$$

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

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

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

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

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

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

Where $\eta_k = -\dfrac{g_k^T g_{k-1}}{\|g_k\|^2}$.

Since $g_k^T g_{k-1} = 0$. Hence, we have

$$r_k = \bar{r}_k \tag{4.24}$$

which, with (4.23) and (4.24), implies that (4.20) holds with $c_1 = c_2$.

Now, if

$$\overline{r}_{k-1} > 1,$$

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

$$\overline{r}_k \geq h(\lambda, -\sigma_2, 1) \stackrel{\Delta}{=} c_2.$$

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

$$\max\{\overline{r}_{k-1}, \overline{r}_k\} > 1 > c_2.$$

We complete the same steps the proof in the previous case (the case $\overline{r}_{k-1} \leq 1$), we get the same result in which (4.20) holds with $c_1 = c_2$. This completes our proof. □

Thus we have some general convergence results are established for the new nonlinear conjugate gradient method (4.4). It is easy to see from (4.4) that the new nonlinear conjugate gradient method include the two nonlinear conjugate gradient methods mentioned above. For the case when $\lambda \equiv 1$, in Theorem 4.2, we again obtain the convergence result of the FR method in [9]. Letting $\lambda \equiv 0$, in Theorem 4.2, we again obtain the convergence result of the PRP method in [15].

In addition, the methods related to the FR method and the DY method in [7, 13] can also be regarded as special cases of the new method (4.4). For example, to combine the nice global convergence properties of the FR method and the good numerical performance of the PRP method.

Hu and Storey [13] extended the result in [1] to any method (1.2) and (1.3) with β_k satisfying

$$\beta_k \in \left[0, \beta_k^{FR}\right]. \tag{4.25}$$

Gilbert and Nocedal [11] further extended the result to the case that

$$\beta_k \in \left[-\beta_k^{FR}, \beta_k^{FR}\right]. \tag{4.26}$$

Dai and Yuan [7] proved that the method (1.2) and (1.3) with β_k satisfying

$$\beta_k \in \left[\frac{\sigma - 1}{1 + \sigma}\overline{\beta_k}, \overline{\beta_k}\right], \tag{4.27}$$

where $\overline{\beta_k}$ stands for the formula (1.8), and with α_k chosen by the Wolfe line search give the convergence relation (3.10), if the line search conditions are (1.4) and (1.5) with $\sigma \leq \frac{1}{2}$. For methods related to the method (4.4). We have the following result, where s_k is given by

$$s_k = \frac{\beta_k}{\beta_k^*}, \tag{4.28}$$

where β_k^* stands for the formula (1.17). We prove that any method (2.2), (2.3) 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

$$-c \leq s_k \leq (1-\sigma)^{-1}, \tag{4.29}$$

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 (1.2) and (1.3), where*

$$\beta_k = \tau_k \frac{g_k^T d_{k-1} + (1-\lambda)(-g_{k-1}^T d_{k-1})}{g_k^T d_{k-1}} \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \tag{4.30}$$

and where α_k is computed by the strong Wolfe line search (1.4) and (1.5) with $\sigma \leq \frac{1}{2}$. For any $\lambda \in [0,1]$, if

$$\tau_k \in \left[\frac{\sigma}{\lambda - (1+\sigma)}, 1\right], \tag{4.31}$$

and β_k is such that

$$s_k \in \left[-c, (1-\sigma)^{-1}\right], \tag{4.32}$$

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

$$0 < \bar{r}_k < (1-\sigma)^{-1} \qquad \text{for all } k \geq 1. \tag{4.33}$$

Further, the method converges in the sense that (3.10) is true.

Proof. From relation (4.30), (4.28) and by direct calculations we can show that

$$\overline{r}_k = 1 + \left[(l_{k-1} + \lambda - 1)\overline{r}_{k-1}\right]\tau_k, \tag{4.34}$$

and

$$s_k = \frac{(\lambda + l_{k-1} - 1)\tau_k}{(1 + l_{k-1})}, \tag{4.35}$$

where $\overline{r}_k$ and l_k are defined by (4.11) and (4.12). Now the right hand side of (4.34) is a function of λ, τ_k, l_{k-1} and $\overline{r}_{k-1}$, which can be denoted as $h(\lambda, \tau_k, l_{k-1}, \overline{r}_{k-1})$. We prove (4.33) by induction. Noting that $d_1 = -g_1$ and hence $\bar{r}_1 = 1$, we see that (4.33) is true for $k = 1$. We now suppose that (4.33) holds for $k-1$, namely,

$$0 < \bar{r}_{k-1} < (1-\sigma)^{-1}. \tag{4.36}$$

It follows from (1.5)

$$|l_{k-1}| \leq \sigma. \tag{4.37}$$

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

$$\bar{r}_k \le \max\left\{h(\lambda, 1, l_{k-1}, \bar{r}_{k-1}), h(\lambda, , \frac{\sigma}{\lambda - (1+\sigma)}, l_{k-1}, \bar{r}_{k-1})\right\} \tag{4.38}$$

$$\le \max\left\{h(\lambda, , 1, \sigma, \bar{r}_{k-1}), h(\lambda, , \frac{\sigma}{\lambda - (1+\sigma)}, -\sigma, \bar{r}_{k-1})\right\}$$

$$< \max\left\{h(\lambda, , 1, \sigma, (1-\sigma)^{-1}), h(\lambda, , \frac{\sigma}{\lambda - (1+\sigma)}, -\sigma, (1-\sigma)^{-1})\right\}$$

$$= 1 + \frac{\sigma}{1-\sigma} = (1-\sigma)^{-1}, \tag{4.39}$$

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

$$\bar{r}_k > \min\left\{h(\lambda, , 1, -\sigma, (1-\sigma)^{-1}), h(\lambda, , \frac{\sigma}{\lambda - (1+\sigma)}, \sigma, (1-\sigma)^{-1})\right\} \ge 0. \tag{4.40}$$

Thus (4.33) is true for k, by induction, (4.33) holds for $k \ge 1$.

We now prove (3.10) by contradiction and assuming that

$$\|g(x)\| \ge \gamma, \qquad \text{for some } \gamma > 0 \text{ and all } k \ge 1. \tag{4.41}$$

Since $d_k + g_k = \beta_k d_{k-1}$, we have that

$$\|d_k\|^2 = \beta_k^2 \|d_{k-1}\|^2 - 2 g_k^T d_k - \|g_k\|^2. \tag{4.42}$$

Dividing both sides of (4.42) by $\left(g_k^T d_k\right)^2$ and using (4.11) and (4.29), we obtain

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

$$= \frac{(s_k \beta_k^*)^2 \|d_{k-1}\|^2}{\left(g_k^T d_k\right)^2} + \frac{1}{\|g_k\|^2}\left[1 - \left(1 - \frac{1}{\bar{r}_k}\right)^2\right]. \tag{4.43}$$

In addition, by the definition (4.11) of $\bar{r}_k$, the relations (1.3) and (4.29), we get

$$\bar{r}_k \|g_k\|^2 = -g_k^T d_k = \|g_k\|^2 - s_k \beta_k^* g_k^T d_{k-1}, \tag{4.44}$$

the above relation and the definition (4.12) imply that

$$s_k \beta_k^* = \frac{(1 - \bar{r}_k)}{l_{k-1}(g_{k-1}^T d_{k-1})} \|g_k\|^2. \tag{4.45}$$

Relation (4.43) and (4.45), we obtain

$$\frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} = \frac{(1-\bar{r}_k)^2 \|d_{k-1}\|^2}{\bar{r}_k^2 l_{k-1}^2 (g_{k-1}^T d_{k-1})^2} + \frac{1}{\|g_k\|^2}\left[1 - \left(1 - \frac{1}{\bar{r}_k}\right)^2\right]. \tag{4.46}$$

Denote

$$m_k = \frac{1 - \bar{r}_k}{\bar{r}_k l_{k-1}}, \tag{4.47}$$

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

$$|m_k| \leq 1, \qquad \text{for all } k \geq 2. \tag{4.48}$$

the right hand side of (4.47) is a function of l_{k-1} and $\overline{r}_k$, which can be denoted as $h(l_{k-1}, \bar{r}_k)$. We can get by (4.33), (4.37) and Lemma 2.5 that

$$\begin{aligned} m_k &\leq \max\left\{h(\sigma, \bar{r}_k), h(-\sigma, \bar{r}_k)\right\} \\ &< \max\left\{h(\sigma, (1-\sigma)^{-1}), h(-\sigma, (1-\sigma)^{-1})\right\} = 1. \end{aligned} \tag{4.49}$$

Thus we have that

$$\begin{aligned} m_k &\geq \min\left\{h(-\sigma, \bar{r}_k), h(\sigma, \bar{r}_k)\right\} \\ &> \min\left\{h(-\sigma, (1-\sigma)^{-1}), h(\sigma, (1-\sigma)^{-1})\right\} = -1. \end{aligned} \tag{4.50}$$

Therefore (4.48) holds for all $k \geq 2$.

By (4.48) and (4.46), we obtain

$$\frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} \leq \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{1}{\|g_k\|^2}. \tag{4.51}$$

Because $\|d_1\|^2 \diagup \left(g_1^T d_1\right)^2 = 1 \diagup \|g_1\|^2$, (4.51) shows that

$$\frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} \leq \sum_{i=1}^{k} \frac{1}{\|g_i\|^2}, \tag{4.52}$$

for all k. Then we get from this and (4.41) that

$$\frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} \geq \frac{\gamma}{k}, \tag{4.53}$$

which implies that

$$\sum_{k\geq 1} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} = +\infty. \tag{4.54}$$

This contradicts the Zoutendijk condition (2.3). Therefore (3.10) holds. □

5. Numerical results

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, 960M memory) with Windows operation system. There are a number of 68 unconstrained test problems in generalized or extended from CUTE [4] and [3] collection with dimensions ranging from 2 to 8000. We adopt the performance profiles by Delan and Moré [14] to compare the performance between the following four conjugate gradient algorithms:

PRP^{SW}: The PRP method with the strong Wolfe conditions, where $\delta = 10^{-4}$ and $\sigma = 0.1$.

PRP_{+}^{SW}: The PRP method with nonnegative values of $\beta_k = \max\left\{0, \beta_k^{PRP}\right\}$ and the strong Wolfe conditions, where $\delta = 10^{-4}$ and $\sigma = 0.1$.

NDCG^{SW}: Algorithm 3.1 with the Wolfe conditions (1.4) and (1.11), where the scalars σ_1 and σ_2 satisfy the condition (4.14), in addition, $\delta = 10^{-4}$, $\sigma_1 = \sigma_2 = \sigma = 0.1$, $\lambda = 0.5$.

NDCG^{W}: Algorithm 3.1 with the standard Wolfe conditions, where $\delta = 10^{-4}$, $\sigma = 0.1$, $\lambda = 0.5$.

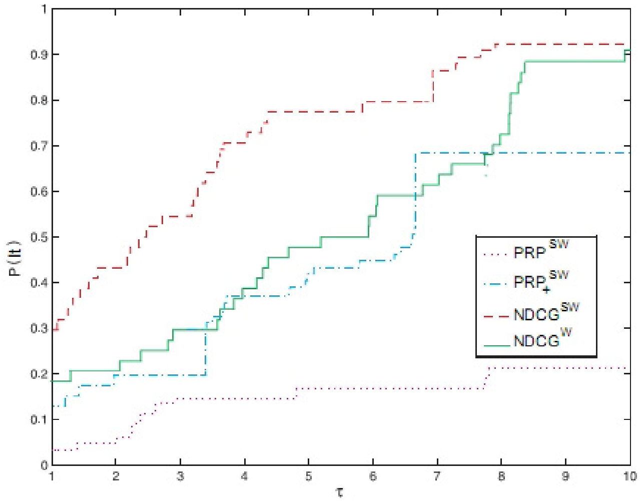

FIGURE 1. Performance files based on Iterations.

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 [13]. 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 preliminary results obtained are encouraging.

6. Conclusions and discussions

In this paper, we have proposed a new nonlinear conjugate gradient method, and studied the global convergence of this method. The new method include the two already known simple and practical conjugate gradient methods. 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 (1.17) with weak Wolfe line search (1.4) and (1.9) 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 method can ensure a descent search direction at every iteration and converges globally under line search conditions (1.4) and (1.11) where the scalars σ_1 and σ_2 satisfy the condition (4.14). From Theorem (4.3), we have carefully studied methods related to the method (4.4). 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, (4.31) and (4.32) 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 (1.4) and (1.5) with $\sigma \leq \dfrac{1}{2}$. In summary, our computational results show that this new descent nonlinear conjugate gradient method,

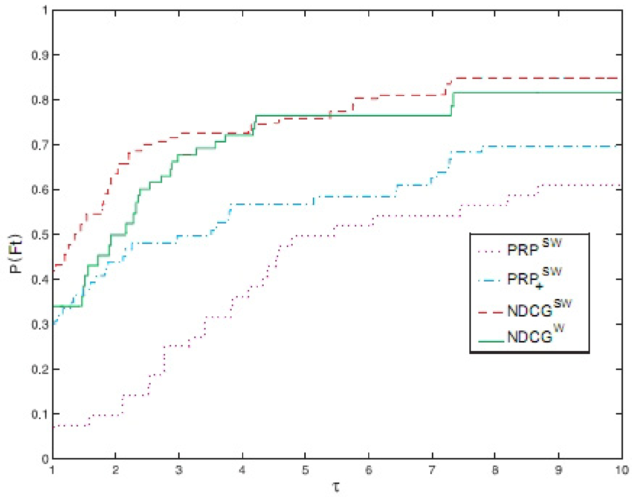

FIGURE 2. Performance files based on function and gradient evaluations.

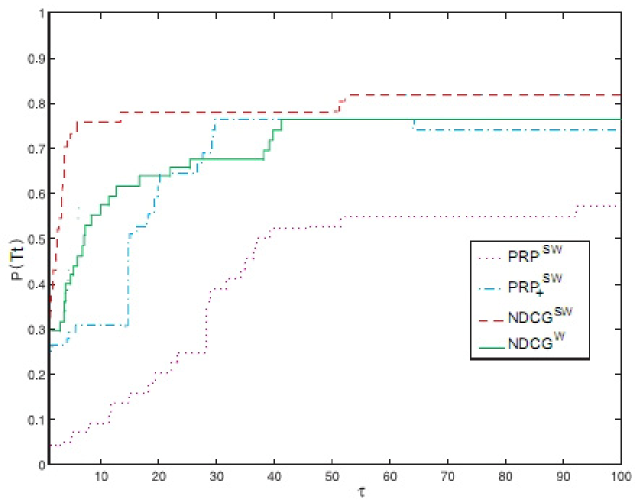

FIGURE 3. Performance files based on CPU time.

namely NDCG method not only converges globally, but also performs better than 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).

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

Appendix

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

Problem	n	Problem	n	Problem	n	Problem	n
ARWHEAD	100	BD1	100	BDEXP	100	BEALE	12
	500		500		500		52
	1000		1000		1000		102
BIGGSB1	6	BROWNAL	10	BROYDN7D	10	COSINE	96
	50		50		50		456
	102		100		100		906
CRAGGLVY	500	DENSCHNB	2	DENSCHNF	2	DIXMAANA	15
	1000		3		3		90
	5000		7		7		300
DIXMAANB	15	DIXMAANC	15	DIXMAANE	15	DIXMAANF	15
	90		90		90		90
	300		300		300		300
DIXMAANG	15	DIXMAANI	15	DIXMAANJ	15	DIXMAANK	15
	90		90		90		90
	300		300		300		300
DIXMAANL	15	DIXON3DQ	500	DQDRTIC	10	DQRTIC	10
	90		1000		50		50
	300		5000		100		100
EDENSCH	36	EG2	10	ENGVAL1	2	FLETCHCR	10
	1000		25		50		50
	2000		50		100		100
FREUROTH	2	GHUMPS	14	GROSEN	60	GPSC1	924
	10		506		180		3444
	50		650		612		7564
HIEBERT	2	HIMMELBLAU	2	LIARWHD	500	MARATOS	2
	5		24		1000		6
	10		45		5000		8
NONCVXU2	10	NONDIA	500	NONDQUAR	100	PENALTY1	4
	20		1000		500		10
	30		5000		1000		50
PENALTY	4	POWELL	60	POWELLBS	2	POWELLSG	500
	10		80		4		1000
	50		100		8		5000
POWER	10	PPQ2	30	PSC1	4	QF1	30
	20		50		25		50
	30		75		50		75
QF2	2	QP1	2	QP2	4	QUARTC	100
	4		4		25		500
	6		6		100		1000
RAYDAN1	20	RAYDAN2	20	ROSEN	2	SINQUAD	5
	50		50		4		50
	100		100		6		100
SQ1	28	SQ2	28	SROSENBR	500	TRIDIA	30
	100		100		1000		50
	499		499		5000		100
WHITEHOLST	12	WOODS	4	TRIDIAG1	10	TRIDIAG2	10
	31		100		20		20
	68		1000		30		30
EXTRIGON	2	GTRIDIAGI	14	DIAG2	4	CLIFF	2
	5		506		6		4
	10		650		8		6

References

[1] M. Al-Baali, Descent property and global convergence of the fletcher reeves 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. Softw. (TOMS)* **21** (1995) 123–160.

[5] 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.

[6] Y. Dai and Y. Yuan, Some properties of a new conjugate gradient method, in Advances in Nonlinear Programming. Springer (1998) 251–262.

[7] Y. Dai and Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property. *SIAM J. Optim.* **10** (1999) 177–182.

[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 and Y.-X. YUAN, Convergence properties of the fletcher-reeves method. *IMA J. Numer. Anal.* **16** (1996) 155–164.

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

[11] J.C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization. *SIAM J. Optim.* **2** (1992) 21–42.

[12] M. Hestenes, Methods of conjugate gradients for solving linear systems. *Res. Nation. Bureau Standards* **49** (1952) 409–436.

[13] Y. Hu and C. Storey, Global convergence result for conjugate gradient methods. *J. Optim. Theory Appl.* **71** (1991) 399–405.

[14] J. Moré and E.D. Dolan, Benchmarking optimization software with performance files. *Math. Program* **91** (2002) 201–2013.

[15] B.T. Polyak, The conjugate gradient method in extremal problems. *USSR Comput. Math. Math. Phys.* **9** (1969) 94–112.

[16] M. Powell, Nonconvex minimization calculations and the conjugate gradient method. *Numer. Anal.* (1984) 122–141.

[17] B. Sellami, Y. Laskri and R. Benzine, A new two-parameter family of nonlinear conjugate gradient methods. *Optimization* **64** (2015) 993–1009.

[18] D.F. Shanno, Conjugate gradient methods with inexact searches. *Math. Oper. Res.* **3** (1978) 244–256.