

**A THREE-PARAMETER FAMILY OF  
NONLINEAR CONJUGATE GRADIENT  
METHODS**

by

**Y. H. Dai and Y. Yuan**

Report No. **ICM-98-050**

September 1998

# A THREE-PARAMETER FAMILY OF NONLINEAR CONJUGATE GRADIENT METHODS\*

Y. H. Dai and Y. Yuan

*State Key Laboratory of Scientific and Engineering Computing,  
Institute of Computational Mathematics and Scientific/Engineering Computing,  
Chinese Academy of Sciences, P. O. Box 2719, Beijing 100080, China  
Email-addresses: dyh@lsec.cc.ac.cn, yyx@lsec.cc.ac.cn*

## Abstract

In this paper, we propose a three-parameter family of conjugate gradient methods for unconstrained optimization. The three-parameter family of methods not only includes the already-existing six practical nonlinear conjugate gradient methods, but subsumes some other families of nonlinear conjugate gradient methods as its subfamilies. With Powell's restart criterion, the three-parameter family of methods with the strong Wolfe line search is shown to ensure the descent property of each search direction. Some general convergence results are also established for the three-parameter family of methods. This paper can also be regarded as a brief review on nonlinear conjugate gradient methods.

**Key words:** unconstrained optimization, conjugate gradient methods, line search, global convergence.

**AMS classification:** 65k, 90c.

## 1. Introduction

Consider the unconstrained optimization problem

$$\min_{x \in R^n} f(x), \quad (1.1)$$

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

$$x_{k+1} = x_k + \alpha_k d_k, \quad (1.2)$$

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

---

\*Research partly supported by Chinese NSF grants 19525101, 19731010 and 19801033.

where  $g_k = \nabla f(x_k)$ ,  $\alpha_k$  is a stepsize obtained by a one-dimensional line search and  $\beta_k$  is a scalar. The strong Wolfe conditions, namely,

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

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

where  $0 < \delta < \sigma < 1$ , are often imposed on the line search (in this case, we call the line search as the strong Wolfe line search). The scalar  $\beta_k$  is so chosen 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 generally functions, however, different formulae for scalar  $\beta_k$  result in distinct nonlinear conjugate gradient methods. Several famous formulae for  $\beta_k$  are the Fletcher-Reeves (FR), Polak-Ribière-Polyak (PRP), and Hestenes-Stiefel (HS) formulae (see [15; 27, 28; 18]), which are given by

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

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

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

respectively, where  $\|\cdot\|$  means the Euclidean norm and  $y_{k-1} = g_k - g_{k-1}$ . The convergence properties of the FR, PRP, and HS methods have been studied in many references, for example [1, 5, 7, 16, 17, 21, 25, 28, 31]. However, if the condition imposed on  $\sigma$  in (1.5) is only that  $\sigma < 1$ , neither of the above three famous nonlinear conjugate gradient methods with the strong Wolfe line search can ensure a descent search direction even if  $f$  is quadratic (see [3, 7]).

The conjugate descent (CD) method by Fletcher [14], where

$$\beta_k^{CD} = \frac{\|g_k\|^2}{-d_{k-1}^T g_{k-1}}, \quad (1.9)$$

ensures a descent direction for generally functions if the line search satisfies the strong Wolfe conditions (1.4)–(1.5) with  $\sigma < 1$ . But the global convergence of the CD method is proved (see [8]) only for the case when the line search satisfies (1.4) and

$$\sigma g_k^T d_k \leq g(x_k + \alpha_k d_k)^T d_k \leq 0. \quad (1.10)$$

For any positive constant  $\bar{\sigma}$ , an example is constructed in [8] showing that the conjugate descent method with  $\alpha_k$  satisfying (1.4) and

$$\sigma g_k^T d_k \leq g(x_k + \alpha_k d_k)^T d_k \leq -\bar{\sigma} g_k^T d_k \quad (1.11)$$

needs not converge.

Recently, Dai and Yuan [6] proposed a new conjugate gradient method, in which

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}. \quad (1.12)$$

A remarkable property of the DY method is that it provides a descent search direction at every iteration and converges globally provided that the stepsize satisfies the Wolfe conditions, namely, (1.4) and

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k. \quad (1.13)$$

Some other properties of the DY method were exposed in the references [4, 9, 11, 12].

In [11], Dai and Yuan proposed a family of globally convergent conjugate methods, in which

$$\beta_k = \frac{\|g_k\|^2}{\lambda \|g_{k-1}\|^2 + (1 - \lambda) d_{k-1}^T y_{k-1}}, \quad (1.14)$$

where  $\lambda \in [0, 1]$  is a parameter. [12] further studied the case when  $\lambda \in (-\infty, +\infty)$ , and proved that the family of methods using line searches that satisfy (1.4) and (1.11) converges globally if the parameters  $\sigma$ ,  $\bar{\sigma}$ , and  $\lambda$  are such that

$$\sigma - 1 \leq (\sigma + \bar{\sigma})\lambda \leq 1. \quad (1.15)$$

Another marginal but interesting note of the DY method is that, the formula (1.12) for  $\beta_k$  has the same nominator as (1.6) and the same denominator as (1.8) (see [6]). In [23], Nazareth regarded the FR, PRP, HS, and DY formulae as the four leading contenders for the scalar  $\beta_k$  and proposed a two-parameter family of conjugate gradient methods:

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

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

In this paper, we will propose a three-parameter family of conjugate gradient methods, which includes the five nonlinear conjugate gradient methods mentioned above and the one in [22]. The three-parameter family of methods also has several other families of conjugate gradient methods and some hybrid methods as its special cases (see the next section). In Section 3, we study the descent property of the three-parameter family of methods. We prove that, if Powell [24]'s restart criterion is used, the three-parameter family of methods with the strong Wolfe line search produces a descent search direction at every iteration unless the current point is a stationary point. In Section 4, some general convergence results are established for the three-parameter family of methods. Brief discussions are given in the last section.

## 2. A three-parameter family of conjugate gradient methods

In [22], Liu and Storey presented the following formula for the scalar  $\beta_k$ :

$$\beta_k^{LS} = \frac{g_k^T y_{k-1}}{-d_{k-1}^T g_{k-1}}. \quad (2.1)$$

A useful property of formulae (1.8) and (2.1) in computations is observed in [22], that is, the next direction  $d_{k+1}$  in (1.2) is independent of the length of  $d_k$  when  $\beta_k$  takes the form of (1.8) or (2.1).

Many authors have presented other choices for the scalar  $\beta_k$ , for example Buckley and Lenir [2], Daniel [13], Gilbert and Nocedal [16], Qi *et al* [28], Shanno [29], and Touati-Ahmed and

Storey [30]. Observing that the formulae (1.6)–(1.9), (1.12) and (2.1) share two nominators and three denominators, we can use the combinations of these nominators and denominators to obtain the following three-parameter family:

$$\beta_k = \frac{(1 - \lambda_k)\|g_k\|^2 + \lambda_k g_k^T y_{k-1}}{(1 - \mu_k - \omega_k)\|g_{k-1}\|^2 + \mu_k d_{k-1}^T y_{k-1} - \omega_k d_{k-1}^T g_{k-1}}, \quad (2.2)$$

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

$$g_k^T y_{k-1} = \|g_k\|^2 - g_k^T g_{k-1} \quad (2.3)$$

and

$$g_{k-1}^T d_{k-1} = -\|g_{k-1}\|^2 + \beta_{k-1} g_{k-1}^T d_{k-2}, \quad (2.4)$$

we can rewrite formula (2.2) as

$$\beta_k = \frac{\|g_k\|^2 - \lambda_k g_k^T g_{k-1}}{\|g_{k-1}\|^2 + \mu_k g_k^T d_{k-1} - \omega_k \beta_{k-1} g_{k-1}^T d_{k-2}}. \quad (2.5)$$

If the objective function is convex quadratic and the stepsize is the exact one-dimensional minimizer, the above formula for  $\beta_k$  clearly reduces to the FR formula (1.6), since in this case we have that

$$g_k^T d_{k-1} = 0 \quad (2.6)$$

and

$$g_k^T g_{k-1} = 0. \quad (2.7)$$

However, for general functions, relations (2.6) and (2.7) need not hold. Therefore the methods (1.2), (1.3) and (2.5) with different values of  $\lambda_k$ ,  $\mu_k$  and  $\omega_k$  form a three-parameter family of nonlinear conjugate gradient methods.

It is easy to see from (2.2) that the three-parameter family of methods includes the already-existing six simple and practical nonlinear conjugate gradient methods as its extreme case. If  $\omega_k \equiv 0$ , then the family reduces to the two-parameter family of conjugate gradient methods in [23]. Further, if  $\omega_k \equiv 0$  and  $\lambda_k \equiv \lambda$ , then the family reduces to the one-parameter of conjugate gradient methods in [11]. Therefore the three-parameter family has the one-parameter family in [11] and the two-parameter family in [23] as its subfamilies.

In addition, the hybrid methods in [6, 16, 19] can also be regarded as special cases of the three-parameter family. For example, to combine the nice global convergence properties of the FR method and the good numerical performances of the PRP method, Hu and Storey [19] proposed a hybrid method, where

$$\beta_k = \max\{0, \min\{\beta_k^{PRP}, \beta_k^{FR}\}\}. \quad (2.8)$$

One can easily see that, formula (2.8) is corresponding to (2.5) with

$$\begin{aligned} \lambda_k &= \begin{cases} \frac{\|g_k\|^2}{g_k^T g_{k-1}}, & \text{if } g_k^T g_{k-1} \geq \|g_k\|^2; \\ 1, & \text{if } g_k^T g_{k-1} \in (0, \|g_k\|^2); \\ 0, & \text{if } g_k^T g_{k-1} \leq 0, \end{cases} \\ \mu_k &\equiv 0, \\ \omega_k &\equiv 0. \end{aligned} \quad (2.9)$$

### 3. Powell's restart criterion and descent property

As mentioned in the first section, if we only require that  $\sigma < 1$ , any of the FR, PRP, and HS methods with the strong Wolfe line search may produce ascent search directions even if the objective function is quadratic. Thus, special attentions must be given to the problem how to keep the descent property of conjugate gradient methods. In this section, we will prove that, if Powell [24]'s restart criterion is applied, the three-parameter family of methods with the strong Wolfe line search can guarantee the descent property of each search direction.

When dealing with Beale's three-term conjugate gradient method, Powell [24] suggested a restart with  $d_k = -g_k$  if the following condition is satisfied:

$$|g_k^T g_{k-1}| \leq \xi \|g_k\|^2, \quad (3.1)$$

where  $\xi > 0$  is some positive constant. As Powell [24] observed, such a restart criterion can avoid that Beale's recurrence converges to a non-stationary point (a strict convergence result was given in [10] for Beale's method with Powell's restart criterion), and improve the numerical behaviors of Beale's method. In fact, for standard conjugate gradient methods, if the function is convex quadratic and the line search is exact, then relation (2.7) implies that no restarts would take place and finite termination could occur. Thus, the quantity  $\frac{|g_k^T g_{k-1}|}{\|g_k\|^2}$  would indicate strong local nonquadratic behavior and hence would be indicative of a need for restarting. In the implementations of conjugate gradient methods, Powell's restart criterion has been used by many authors, for example Buckley and Lenir [2] and Khoda *et al* [20].

In the following, to show the importance of Powell's restart criterion in keeping the descent property of conjugate gradient methods, we first take the HS method as an illustrative example. For this purpose, we define

$$r_k = -\frac{g_k^T d_k}{\|g_k\|^2}. \quad (3.2)$$

It is obvious that  $d_k$  is a descent direction if and only if  $r_k > 0$ . For the HS method (1.2), (1.3) and (1.8), direct calculations yield

$$r_k = \frac{-g_{k-1}^T d_{k-1}}{d_{k-1}^T y_{k-1}} \left[ 1 - \frac{g_k^T g_{k-1}}{\|g_k\|^2} \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \right]. \quad (3.3)$$

Suppose that  $d_{k-1}$  is a descent direction and the  $(k-1)$ -th line search satisfies the strong Wolfe conditions (1.4)–(1.5). Then we have that  $g_{k-1}^T d_{k-1} < 0$  and  $d_{k-1}^T y_{k-1} > 0$ . Furthermore, it follows from (1.5) that

$$\left| \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \right| \leq \sigma. \quad (3.4)$$

Therefore by (3.3), if Powell's restart criterion (3.2) is used, the HS method can ensure the descent property of the next direction  $d_k$  provided that the parameters  $\xi$  and  $\sigma$  are such that the following condition holds:

$$\xi \sigma < 1. \quad (3.5)$$

For the three-parameter family of conjugate gradient methods, we can prove the following general theorem.

**Theorem 3.1** Consider any method in the form (1.2), (1.3) and (2.2) with  $\lambda_k \in [0, 1]$ ,  $\mu_k \in [0, 1]$  and  $\omega_k \in [0, 1 - \mu_k]$ , with the search condition (1.5), and with Powell's restart criterion (3.1). If the parameters  $\xi$  and  $\sigma$  satisfy

$$(1 + \xi)\sigma \leq \frac{1}{2}, \quad (3.6)$$

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

$$0 < r_k < \frac{1}{1 - (1 + \xi)\sigma}. \quad (3.7)$$

**Proof** Without loss of generality, assume that (3.1) holds for all  $k$ . We prove (3.7) by induction. Noting that  $d_1 = -g_1$  and hence  $r_1 = 1$ , (3.7) is true for  $k = 1$ . We now suppose that (3.7) holds for  $k - 1$ , namely,

$$0 < r_{k-1} < \frac{1}{1 - (1 + \xi)\sigma}. \quad (3.8)$$

By (1.3) and (2.5), direct calculations show that

$$r_k = 1 - \left[ 1 - \lambda_k \frac{g_k^T g_{k-1}}{\|g_k\|^2} \right] b_k, \quad (3.9)$$

where

$$b_k = \frac{g_k^T d_{k-1}}{(1 - \mu_k - \omega_k)\|g_{k-1}\|^2 + \mu_k g_k^T d_{k-1} - (\mu_k + \omega_k)g_{k-1}^T d_{k-1}}. \quad (3.10)$$

Using (1.5), (3.8) and the fact that  $\mu_k, \omega_k \geq 0$  in (3.9), we get that

$$\begin{aligned} b_k &\leq \frac{-\sigma g_{k-1}^T d_{k-1}}{(1 - \mu_k - \omega_k)\|g_{k-1}\|^2 - \sigma \mu_k g_{k-1}^T d_{k-1} - (\mu_k + \omega_k)g_{k-1}^T d_{k-1}} \\ &= \frac{\sigma r_{k-1}}{(1 - \mu_k - \omega_k) + [(1 + \sigma)\mu_k + \omega_k]r_{k-1}} \\ &< \frac{\sigma[1 - (1 + \xi)\sigma]^{-1}}{(1 - \mu_k - \omega_k) + [(1 + \sigma)\mu_k + \omega_k][1 - (1 + \xi)\sigma]^{-1}} \\ &\leq \frac{\sigma}{[1 - (1 + \xi)\sigma] + (2 + \xi)\sigma\mu_k + (1 + \xi)\sigma\omega_k} \\ &\leq \frac{\sigma}{1 - (1 + \xi)\sigma}. \end{aligned} \quad (3.11)$$

Similarly, we can prove that

$$b_k > -\frac{\sigma}{1 - (1 + \xi)\sigma}. \quad (3.12)$$

Thus from (3.9), (3.11), (3.12), (3.1), (3.6) and the fact that  $\lambda_k \in [0, 1]$ , we obtain

$$0 \leq 1 - \frac{(1 + \xi)\sigma}{1 - (1 + \xi)\sigma} < r_k < 1 + \frac{(1 + \xi)\sigma}{1 - (1 + \xi)\sigma} = \frac{1}{1 - (1 + \xi)\sigma}. \quad (3.13)$$

Therefore (3.7) is also true for  $k$ . By induction, (3.7) holds for all  $k \geq 1$ .  $\square$

In real computations, Powell [24] suggested the value of  $\xi$  in (3.1) could be  $\xi = 0.2$ . Here we should point out that condition (3.6) allows relatively large values of  $\xi$  and hence is flexible, because the parameter  $\sigma$  in (1.5) is generally set to a small value, normally  $\sigma = 0.1$ .

## 4. Convergence properties

In this section, we study the global convergence properties of the three-parameter family of nonlinear conjugate gradient methods. For convenience, we assume that  $g_k \neq 0$  for all  $k$ , for otherwise a stationary point has been found. We also assume that  $\beta_k \neq 0$  for all  $k$ . This is because the direction in (1.3) reduces to  $-g_k$  if  $\beta_k = 0$ . Then either we can take  $x_{\bar{k}}$  as the new starting point where  $\bar{k}$  is the largest index for which  $\beta_{\bar{k}} = 0$ , or the convergence relation  $\liminf \|g_k\| = 0$  holds.

The following assumption is imposed on the objective function throughout this section.

**Assumption 4.1** (i) *The level set  $\mathcal{L} = \{x \in \mathbb{R}^n : f(x) \leq f(x_1)\}$  is bounded; (ii) *In some neighborhood  $\mathcal{N}$  of  $\mathcal{L}$ ,  $f$  is differentiable and its gradient  $g$  is Lipschitz continuous, namely, there exists a positive constant  $L$  such that**

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \text{for all } x, y \in \mathcal{N}. \quad (4.1)$$

The above assumption implies that there exists a positive constant  $\bar{\gamma}$  such that

$$\|g(x)\| \leq \bar{\gamma}, \quad \text{for all } x \in \mathcal{L}. \quad (4.2)$$

To give the first convergence result for the three-parameter family of methods, we need the following lemma, which can be proved similarly to Theorem 3.3 in [11].

**Lemma 4.2** *Suppose that  $x_1$  is a starting point for which Assumption 4.1 holds. Consider the method (1.2)–(1.3), where  $d_k$  is a descent direction and  $\alpha_k$  satisfies the Wolfe conditions (1.4) and (1.13). If there exists a positive sequence  $\{\phi_k\}$  such that*

$$|\beta_k| \leq \frac{\phi_k}{\phi_{k-1}} \quad (4.3)$$

and

$$\sum_{k \geq 1} \frac{\|g_k\|^2}{\phi_k^2} = +\infty, \quad (4.4)$$

then the method converges in the sense that

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \quad (4.5)$$

By Lemma 4.2, we can prove the following general result for the three-parameter family of nonlinear conjugate gradient methods.

**Theorem 4.3** *Suppose that  $x_1$  is a starting point for which Assumption 4.1 holds. Consider any method in the form (1.2), (1.3) and (2.2) with  $\lambda_k \in [0, 1]$ ,  $\mu_k \in [0, 1]$  and  $\omega_k \in [0, 1 - \mu_k]$ , with the strong Wolfe line search (1.4)–(1.5), and with Powell's restart criterion (3.1). Denote  $l_k = |\beta_k / \beta_k^{FR}|$ . If (3.6) holds, and if the parameters  $\lambda_k$ ,  $\mu_k$  and  $\omega_k$  are such that*

$$\prod_{j=2}^k l_j \leq \tau \sqrt{k} \quad (4.6)$$

holds for some constant  $\tau > 0$  and all  $k \geq 2$ , then the method converges in the sense that (4.5) holds.

**Proof** Since the parameters  $\xi$  and  $\sigma$  satisfies (3.6), we have by Theorem 3.1 that (3.7) holds for all  $k$ , which implies that each  $d_k$  is a descent direction. Define

$$\phi_k = \left( \prod_{j=2}^k l_j \right) \|g_k\|^2. \quad (4.7)$$

Then by (1.6) and the definition of  $l_k$ , we can write

$$|\beta_k| = \frac{\phi_k}{\phi_{k-1}}. \quad (4.8)$$

It follows from (4.2), (4.6) and (4.7) that

$$\frac{\|g_k\|^2}{\phi_k^2} \geq \frac{1}{\tau^2 \bar{\gamma}^2 k}, \quad (4.9)$$

which implies the truth of (4.4). Therefore by Lemma 4.2, (4.5) holds.  $\square$

Now we discuss some special choices that satisfy condition (4.6). As mentioned in the last paragraph of Section 2, the hybrid methods in [16] and [19] can be regarded as special cases of the three-parameter family of conjugate gradient methods. Using the above theorem, we can again deduce the global convergence of thy hybrid methods. For example, for the hybrid method (2.8), we have that  $0 \leq l_k \leq 1$ , which indicates that relation (4.6) holds with  $\tau = 1$ . Hence (4.5) holds.

Assume that  $\prod_{j=2}^{k-1} l_j \leq \tau \sqrt{k-1}$  for some  $k \geq 3$ . Then at the  $k$ -th iteration, (4.6) must hold if

$$l_k \leq \sqrt{1 + \frac{1}{k-1}}. \quad (4.10)$$

Note that if  $\lambda_k = \mu_k = \omega_k = 0$ ,  $\beta_k$  reduces to  $\beta_k^{FR}$  and hence  $l_k = 1$ . Thus even for large  $k$ , we can see from relation (4.10) that there exists some interval in  $[0, 1]$  for each of the parameters  $\lambda_k$ ,  $\mu_k$  and  $\omega_k$  such that (4.6) hold. As pointed by the referee, one problem with condition (4.6) is that it depends on the sequences of points produced by the methods. Though there always exists choices of  $\lambda_k$ ,  $\mu_k$  and  $\omega_k$  that satisfy condition (4.6) (as stated above), this restriction may reduce the admissible intervals for  $\lambda_k \in [0, 1]$ ,  $\mu_k \in [0, 1]$ , and  $\omega_k \in [0, 1 - \mu_k]$ . Generally, for some  $k$  the value of  $l_k$  may be less than 1 and as a result, this will allow relatively large values of the consequent  $l_k$ . For example, if the step  $\|x_k - x_{k-1}\|$  is very small and  $\lambda_k$  is close to 1 at some iteration far way from the solution, then  $\beta_k$  and hence  $l_k$  may be much smaller than 1. Another point that should be pointed out here is that, we can also enlarge the admissible intervals for  $\lambda_k$ ,  $\mu_k$  and  $\omega_k$  by setting  $\tau$  in (4.6) to a large value. Inequality (4.6) suggests that one possibility is to choose  $\lambda_k$ ,  $\mu_k$  and  $\omega_k$  such that the absolute value of right hand side of (2.2) is as small as possible.

By Lemma 4.2, we can also prove the following convergence result.

**Theorem 4.4** *Suppose that  $x_1$  is a starting point for which Assumption 4.1 holds. Consider any method in the form (1.2), (1.3) and (2.2) with  $\lambda_k \in [0, 1]$ ,  $\mu_k \in [0, 1]$  and  $\omega_k \in [0, 1 - \mu_k]$ , with the strong Wolfe line search (1.4)–(1.5), and with Powell's restart criterion (3.1). If (3.6) holds, and if the parameters  $\lambda_k$ ,  $\mu_k$  and  $\omega_k$  are such that*

$$0 \leq \lambda_k g_k^T g_{k-1} \leq \|g_k\|^2 \quad (4.11)$$

and

$$\mu_k g_k^T d_{k-1} - \omega_k \beta_{k-1} g_{k-1}^T d_{k-2} \geq -\lambda_{k-1} g_{k-1}^T g_{k-2} \quad (4.12)$$

hold for all  $k \geq 2$ , then the method converges in the sense that (4.5) holds.

**Proof** Since the parameters  $\xi$  and  $\sigma$  satisfies (3.6), we have by Theorem 3.1 that (3.7) holds for all  $k$ , which implies that each  $d_k$  is a descent direction. From (4.11), (4.12) with  $k$  replaced by  $k+1$ , and the equivalent formula (2.5) of  $\beta_k$ , we see that relation (4.3) holds with

$$\phi_k = (1 - \mu_{k+1} - \omega_{k+1}) \|g_k\|^2 + \mu_{k+1} d_k^T y_k - \omega_{k+1} d_k^T g_k. \quad (4.13)$$

Using (1.5) and (3.7) in (4.13), we can prove

$$\begin{aligned} \phi_k &\leq [(1 - \mu_{k+1} - \omega_{k+1}) + (1 + \sigma)\mu_{k+1}r_k + \omega_{k+1}r_k] \|g_k\|^2 \\ &\leq [1 + (2 + \sigma)(1 - (1 + \xi)\sigma)^{-1}] \|g_k\|^2 \\ &= \frac{3 - \xi\sigma}{1 - (1 + \xi)\sigma} \|g_k\|^2 \end{aligned} \quad (4.14)$$

for any  $\mu_{k+1} \in [0, 1]$  and  $\omega_{k+1} \in [0, 1 - \mu_{k+1}]$ . The above relation and (4.2) imply that

$$\sum_{k \geq 1} \frac{\|g_k\|^2}{\phi_k^2} = +\infty. \quad (4.15)$$

Thus by Lemma 4.2, (4.5) holds.  $\square$

If  $\xi \leq 1$ , the second inequality in (4.11) clearly holds since by (3.1) and the fact that  $\lambda_k \in [0, 1]$ , we have that

$$\lambda_k g_k^T g_{k-1} \leq |g_k^T g_{k-1}| \leq \xi \|g_k\|^2 \leq \|g_k\|^2. \quad (4.16)$$

It is easy to see that the FR method ( $\lambda_k = \mu_k = \omega_k = 0$ ) satisfies the two conditions (4.11) and (4.12). However, for other methods it is not clear whether these conditions hold, as these conditions depend on the sequences of the points generated by the methods. For example, in the extreme case  $g_k^T g_{k-1} < 0$ ,  $g_k^T d_{k-1} < 0$  and  $\beta_{k-1} g_{k-1}^T d_{k-2} > 0$ , we need to choose  $\lambda_k = \mu_{k+1} = \omega_{k+1} = 0$ . If  $g_k^T g_{k-1} > 0$ , there always exist admissible intervals for  $\lambda_k$ ,  $\mu_{k+1}$  and  $\omega_{k+1}$ .

The following lemma is drawn from Gilbert and Nocedal [16].

**Lemma 4.5** *Suppose that  $x_1$  is a starting point for which Assumption 4.1 holds. Consider any method in the form (1.2)–(1.3) with the following three properties: (i)  $\beta_k \geq 0$ ; (ii) the Wolfe conditions (1.4) and (1.13), and the sufficient descent condition  $g_k^T d_k \leq -c \|g_k\|^2$  hold for all  $k$  and some positive constant  $c$ ; (iii) Property (\*) holds, namely, there exist constants  $b > 1$  and  $\lambda > 0$  such that for all  $k$ :  $|\beta_k| \leq b$  and if  $\|x_k - x_{k-1}\| \leq \lambda$ ,  $|\beta_k| \leq (2b)^{-1}$ . Then the method converges in the sense that (4.5) holds.*

By Lemma 4.5, we can prove the following general result for the three-parameter family of nonlinear conjugate gradient methods.

**Theorem 4.6** *Suppose that  $x_1$  is a starting point for which Assumption 4.1 holds. Consider any method in the form (1.2), (1.3) and (2.2), where  $\lambda_k \in [0, 1]$ ,  $\mu_k \in [0, 1]$  and  $\omega_k \in [0, 1 - \mu_k]$ , where the stepsize satisfies the strong Wolfe conditions (1.4)–(1.5), and where the restart criterion*

$$-\xi \|g_k\|^2 \leq g_k^T g_{k-1} \leq \|g_k\|^2 \quad (4.17)$$

*is used. If the parameters are such that*

$$(1 + \xi)\sigma < \frac{1}{2} \quad (4.18)$$

*and*

$$\lambda_k \geq 1 - c_1 \|x_k - x_{k-1}\|, \quad (4.19)$$

*where  $c_1 > 0$  is constant, the method converges in the sense that (4.5) holds.*

**Proof** We proceed by contradiction. Assume that

$$\liminf_{k \rightarrow \infty} \|g_k\| \neq 0. \quad (4.20)$$

Then there exists a positive constant  $\gamma$  such that

$$\|g_k\| \geq \gamma, \quad \text{for all } k \geq 1. \quad (4.21)$$

Using (4.17) and (4.18), we can see from (3.13) in the proof of Theorem 3.1 that for all  $k \geq 1$ :

$$r_k \geq \frac{1 - 2(1 + \xi)\sigma}{1 - (1 + \xi)\sigma} \triangleq c_2, \quad (4.22)$$

where  $r_k$  is given in (3.2). Thus the sufficient descent condition holds. From (1.5), (3.7), (4.1), (4.2), (4.21) and the fact that  $\lambda_k \leq 1$ , we can show

$$\begin{aligned} |\beta_k| &\leq \frac{(c_1 \|g_k\|^2 + L \|g_k\|) \|x_k - x_{k-1}\|}{\|g_{k-1}\|^2 [(1 - \mu_k - \omega_k) + \mu_k(1 - \sigma)c_2 + \omega_k c_2]} \\ &\leq \frac{(c_1 \|g_k\|^2 + L \|g_k\|) \|x_k - x_{k-1}\|}{c_2 \|g_{k-1}\|^2} \\ &\leq c_3 \|x_k - x_{k-1}\|, \end{aligned} \quad (4.23)$$

where

$$c_3 = \frac{c_1 \bar{\gamma}^2 + L \bar{\gamma}}{c_2 \gamma^2}. \quad (4.24)$$

Since Assumption 4.1 implies that there exists a positive constant  $\rho$  such that

$$\|x\| \leq \rho, \quad \text{for all } x \in \mathcal{L}, \quad (4.25)$$

for  $b = 2c_3\rho$  and  $\lambda = (4c_3^2\rho)^{-1}$ , we have from (4.23) and (4.25) that

$$|\beta_k| \leq b \quad (4.26)$$

and if  $\|x_k - x_{k-1}\| \leq \lambda$ ,

$$|\beta_k| \leq (2b)^{-1}. \quad (4.27)$$

Thus Property (\*) holds. In addition, (3.7) and (4.17) imply that  $\beta_k \geq 0$ . Therefore the conditions of Lemma 4.5 are all satisfied and hence (4.5) holds.  $\square$

Theorems 4.3, 4.4 and 4.6 provide some general convergence results for the three-parameter family of nonlinear conjugate gradient methods. If the parameters in (2.2) are specifically chosen, then better global convergence results can be expected. We will end this section with such an illustrative example.

Such an example stems from the following fact. For any method of the form (1.2)–(1.3) with  $d_k$  satisfying  $g_k^T d_k < 0$  and with the strong Wolfe line search (1.4)–(1.5), Corollary 2.4 in [5] tells us that the method converges globally provided that the norm of  $d_k$  does not increase faster than linearly. Exactly, the method gives (4.5) if the following condition holds:

$$\sum_{k \geq 1} \frac{1}{\|d_k\|^2} = +\infty. \quad (4.28)$$

Therefore among the three-parameter family of nonlinear conjugate gradient methods, to shorten the length of  $d_k$ , it is reasonable to choose the parameters such that  $|\beta_k|$  reaches its smallest value, namely, let

$$(\bar{\lambda}_k, \bar{\mu}_k, \bar{\omega}_k) = \operatorname{argmin}_{\lambda_k, \mu_k \in [0,1], \omega_k \in [0,1-\mu_k]} |\beta_k|, \quad (4.29)$$

where  $\beta_k$  is defined in (2.2). In this case, we can obtain the following global convergence result, in which the line search only requires to satisfy the Wolfe conditions.

**Theorem 4.7** *Suppose that  $x_1$  is a starting point for which Assumption 4.1 holds. Consider the method of the form (1.2), (1.3) and*

$$\bar{\beta}_k = \frac{(1 - \bar{\lambda}_k) \|g_k\|^2 + \bar{\lambda}_k g_k^T y_{k-1}}{(1 - \bar{\mu}_k - \bar{\omega}_k) \|g_{k-1}\|^2 + \bar{\mu}_k d_{k-1}^T y_{k-1} - \bar{\omega}_k d_{k-1}^T g_{k-1}}, \quad (4.30)$$

where  $\bar{\lambda}_k, \bar{\mu}_k, \bar{\omega}_k$  are given in (4.29). If the stepsize satisfies the Wolfe conditions (1.4) and (1.13), then each  $d_k$  generated by the method is a descent direction. Further, the method converges in the sense that (4.5) holds.

**Proof** From the choices of  $\bar{\lambda}_k, \bar{\mu}_k$  and  $\bar{\omega}_k$ , it is easy to see that

$$0 \leq \bar{\beta}_k \leq \beta_k^{DY}, \quad (4.31)$$

where  $\beta_k^{DY}$  is given in (1.12). Thus the statements follow by Theorem 3.1 in [6].

## 5. Conclusions and discussions

In this paper, we have proposed a three-parameter family of nonlinear conjugate gradient methods, and studied the global convergence of these methods. The three-parameter family not only includes the already-existing six simple and practical conjugate gradient methods, but has some other families of conjugate gradient methods as its subfamilies. The three-parameter family also includes some hybrid methods as its special cases.

With Powell's restart criterion, we proved that the three-parameter family can ensure a descent search direction at every iteration. Then under suitable conditions, we established some general convergence results, namely, Theorems 4.3, 4.4 and 4.6, for the three-parameter family

of nonlinear conjugate gradient methods. If the parameters are specifically chosen, better global convergence results could be achieved.

It has been pointed out that condition (3.6) is not strict, because the parameter  $\sigma$  in (1.5) is generally chosen to be a relatively small value, which implies that  $\xi$  in (3.1) could be relatively large. However, it still remains to be explored how to find new and efficient conjugate gradient methods among the three-parameter family. Specifically it would be interesting to find the practical performance of the method (4.29).

**Acknowledgements.** The authors would like to thank an anonymous referee for his comments and suggestions which improved this paper.

## 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), pp. 121-124.
- [2] A. Buckley and A. Lenir, *QN-like variable storage conjugate gradients*, Math. Prog. 27 (1983), 155-175.
- [3] Y. H. Dai, *Analyses of nonlinear conjugate gradient method*, Ph.D. thesis, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1997.
- [4] Y. H. Dai, *Some new properties of a nonlinear conjugate gradient method*, Research report ICM-98-010, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1998.
- [5] Y. H. Dai, J. Y. Han, G. H. Liu, D. F. Sun, H. X. Yin, and Y. Yuan, *Convergence properties of nonlinear conjugate gradient methods*, Research report ICM-98-024, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1998 (accepted by SIAM J. Optimization).
- [6] Y. H. Dai and Y. Yuan, *A Nonlinear Conjugate Gradient Method with Nice Global Convergence Properties*, Research report ICM-95-038, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1995 (accepted by SIAM J. Optimization).
- [7] Y. H. Dai and Y. Yuan, *Convergence properties of the Fletcher-Reeves method*, IMA J. Numer. Anal. Vol. 16 No. 2 (1996), 155-164.
- [8] Y. H. Dai and Y. Yuan, *Convergence properties of the conjugate descent method*, Mathematical Advances Vol. 25 No. 6 (1996), 552-562.
- [9] Y. H. Dai and Y. Yuan, *Some properties of a new conjugate gradient method*, in: Y. Yuan ed., *Advances in Nonlinear Programming* (Kluwer, Boston, 1998), pp. 251-262.
- [10] Y. H. Dai and Y. Yuan, *Convergence properties of Beale-Powell restart method*, Sciences in China (series A), Vol. 28, No. 5, pp. 424-432.

- [11] Y. H. Dai and Y. Yuan, *A class of globally convergent conjugate gradient methods*, Research report ICM-98-030, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1998. (submitted to IMA J. Numer. Anal.)
- [12] Y. H. Dai and Y. Yuan, *Extension of a class of conjugate gradient methods*, Research report ICM-98-049, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1998. (submitted to BIT)
- [13] J. W. Daniel, *The conjugate gradient method for linear and nonlinear operator equations*, SIAM J. Numer. Anal., 4 (1967), 10-26.
- [14] R. Fletcher, *Practical Methods of Optimization vol. 1: Unconstrained optimization*, John Wiley & Sons (New York), 1987.
- [15] R. Fletcher and C. Reeves, *Function minimization by conjugate gradients*, Comput. J. 7 (1964), pp. 149-154.
- [16] J. C. Gilbert and J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM. J. Optimization. Vol. 2 No. 1 (1992), pp. 21-42.
- [17] L. Grippo and S. Lucidi, *A globally convergent version of the Polak-Ribière conjugate gradient method*, Math. Prog. 78 (1997), pp. 375-391.
- [18] M. R. Hestenes and E. L. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Standards Sect. 5, 49 (1952), 409-436.
- [19] Y. F. Hu and C. Storey, *Global convergence result for conjugate gradient methods*, JOTA: Vol. 71 No. 2 (1991) 399-405.
- [20] K. M. Khoda, Y. Liu, and C. Storey, *Generalized Polak-Ribière Algorithm*, Journal of Optimization Theory and Applications, Vol. 75, No. 2 (1992), 345-354.
- [21] G. H. Liu, J. Y. Han, and H. X. Yin, *Global convergence of the Fletcher-Reeves algorithm with an inexact line search*, Report, Institute of Applied Mathematics, Chinese Academy of Sciences, 1993.
- [22] Y. Liu and C. Storey, *Efficient Generalized Conjugate Gradient Algorithms, Part 1: Theory*, Journal of Optimization Theory and Applications, Vol. 69 (1991), 129-137.
- [23] L. Nazareth, *Conjugate-gradient methods*, to appear in: Encyclopedia of Optimization (C. Floudas and P. Pardalos, eds.), Kluwer Academic Publishers, Boston, USA and Dordrecht, The Netherlands (1999).
- [24] M. J. D. Powell, *Restart procedures of the conjugate gradient method*, Math. Program. 2 (1977), pp. 241-254.
- [25] M. J. D. Powell, *Nonconvex minimization calculations and the conjugate gradient method*, in: Lecture Notes in Mathematics vol. 1066, Springer-Verlag (Berlin) (1984), pp. 122-141.
- [26] E. Polak and G. Ribière, *Note sur la convergence de directions conjuguées*, Rev. Francaise Informat Recherche Operationelle, 3e Année 16 (1969), pp. 35-43.

- [27] B. T. Polyak, *The conjugate gradient method in extreme problems*, USSR Comp. Math. and Math. Phys. 9 (1969), pp. 94-112.
- [28] H. D. Qi, J. Y. Han and G. H. Liu, *A modified Hestenes-Stiefel conjugate gradient algorithm*, Chinese Annals of Mathematics 17A : 3 (1996), 277-284. (in Chinese)
- [29] D. F. Shanno, *Conjugate gradient methods with inexact searches*, Math. Oper. Res. 3 (1978), 244-256.
- [30] D. Touati-Ahmed and C. Storey, *Efficient hybrid conjugate gradient techniques*, J. Optimization Theory Appl. 64 (1990) 379-397.
- [31] C. Y. Wang and Y. Z. Zhang, *Global convergence properties of s-related conjugate gradient methods*, Report, Qufu Normal University, 1996.