MATHEMATICS OF COMPUTATION
Volume 72, Number 243, Pages 1317–1328
S 0025-5718(03)01491-1
Article electronically published on February 3, 2003

A FAMILY OF HYBRID CONJUGATE GRADIENT METHODS FOR UNCONSTRAINED OPTIMIZATION

YU-HONG DAI

ABSTRACT. Conjugate gradient methods are an important class of methods for unconstrained optimization, especially for large-scale problems. Recently, they have been much studied. This paper proposes a three-parameter family of hybrid conjugate gradient methods. Two important features of the family are that (i) it can avoid the propensity of small steps, namely, if a small step is generated away from the solution point, the next search direction will be close to the negative gradient direction; and (ii) its descent property and global convergence are likely to be achieved provided that the line search satisfies the Wolfe conditions. Some numerical results with the family are also presented.

1. INTRODUCTION

Consider the unconstrained optimization problem

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

where f is smooth and its gradient is available. Conjugate gradient methods are very useful for solving (1.1), especially if the dimension n is large. The methods are of the form

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

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

where g_k denotes $\nabla f(x_k)$, α_k is a steplength obtained by a line search, and β_k is a scalar. The strong Wolfe line search is to find a steplength α_k such that

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

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

where $\delta \in (0, \frac{1}{2})$ and $\sigma \in (\delta, 1)$. In the conjugate gradient field, it is also possible [4, 10, 11] to use the Wolfe line search, which calculates an α_k satisfying (1.4) and

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k. \tag{1.6}$$

Received by the editor May 19, 2000 and, in revised form, November 9, 2001.
2000 *Mathematics Subject Classification.* Primary 49M37, 65K05, 90C30.
Key words and phrases. Unconstrained optimization, conjugate gradient method, line search, descent property, global convergence.
Research partly supported by the Chinese NSF grants 19801033, 10171104 and a Youth Innovation Fund of the Chinese Academy of Sciences.

For the scalar β_k, many formulas have been proposed. Some of them are called the FR [13], PRP [21, 22], DY [10], HS [15], CD [12], and LS [16] ones, and are given by

$$\begin{aligned} \beta_k^{FR} &= \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, & \beta_k^{PRP} &= \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \\ \beta_k^{DY} &= \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}, & \beta_k^{HS} &= \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \\ \beta_k^{CD} &= \frac{\|g_k\|^2}{-g_{k-1}^T d_{k-1}}, & \beta_k^{LS} &= \frac{g_k^T y_{k-1}}{-g_{k-1}^T d_{k-1}}, \end{aligned}$$

where $y_{k-1} = g_k - g_{k-1}$ and $\|\cdot\|$ is the two norm.

Although all nonlinear conjugate gradient methods should reduce to the linear conjugate gradient method when f is a convex quadratic and the line search is exact, their convergence properties may be quite different for nonquadratic functions. For example, the FR method is globally convergent if the steplength α_k satisfies (1.4)-(1.5) with $\sigma \le \frac{1}{2}$ (for example, see [6]). The DY method converges globally provided that the Wolfe line search (with any $\sigma < 1$) is used [10]. In contrast, the PRP and HS methods need not converge even with the exact line search [20]. Consequently, nonlinear conjugate gradient methods were often analyzed individually. However, it is well known that some quasi-Newton methods can be expressed in a unified way and their properties can be analyzed uniformly (for example, see [1, 2]). Thus, similarly to quasi-Newton methods, we wonder whether there exists a family of conjugate gradient methods, and whether its properties can be analyzed uniformly.

Motivated by the above question, Dai and Yuan [7] proposed a family of conjugate gradient methods, in which

$$\beta_k = \frac{||g_k||^2}{\lambda ||g_{k-1}||^2 + (1-\lambda) d_{k-1}^T y_{k-1}}, \quad \lambda \in [0,1]. \tag{1.7}$$

This family can be regarded as some kind of convex combination of the FR and DY methods. Dai and Yuan [8] further extended the family to the case $\lambda \in (-\infty, +\infty)$ and presented some unified convergence results. Almost simultaneously, Nazareth [18] regarded the FR, PRP, HS, and DY formulas as the four leading contenders for the scalar β_k and proposed a two-parameter family:

$$\beta_k = \frac{\lambda_k ||g_k||^2 + (1-\lambda_k) g_k^T y_{k-1}}{\mu_k ||g_{k-1}||^2 + (1-\mu_k) d_{k-1}^T y_{k-1}}, \quad \lambda_k, \mu_k \in [0,1]. \tag{1.8}$$

Later, based on the six formulas in (1.7), Dai and Yuan [9] proposed a three-parameter family:

$$\beta_k = \frac{||g_k||^2 - \lambda_k g_k^T g_{k-1}}{||g_{k-1}||^2 + \mu_k g_k^T d_{k-1} - \omega_k \beta_{k-1} g_{k-1}^T d_{k-2}}, \tag{1.9}$$

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

In this paper, by analyzing how to keep the descent property of the method (1.2)-(1.3) with the Wolfe line search, we will propose a three-parameter family of hybrid conjugate gradient methods (see §2). One advantage of the family is that it can avoid the propensity of small steps; namely, if a small step is produced far away from the solution, the next search direction is automatically close to the negative gradient direction. Under mild conditions, we prove that the family of methods with the Wolfe line search produce a descent search direction at each iteration (see §3). Convergence properties of the family are analyzed in §4, and some numerical results are reported in §5. A brief discussion is given in the last section.

2. A family of hybrid conjugate gradient methods

Special attention must be paid to how to keep the descent property of conjugate gradient methods. Let us consider the method (1.2)–(1.3) with the steplength α_k satisfying the Wolfe conditions (1.4) and (1.6). Assume that the search direction d_{k-1} is downhill, namely,

$$g_{k-1}^T d_{k-1} < 0. \tag{2.1}$$

It follows from (1.3) that

$$d_k^T g_k = -\|g_k\|^2 + \beta_k g_k^T d_{k-1}. \tag{2.2}$$

Then the descent property of d_k requires

$$\beta_k g_k^T d_{k-1} < \|g_k\|^2. \tag{2.3}$$

Assuming that

$$\beta_k = \|g_k\|^2 / (g_k^T d_{k-1} + b_k), \tag{2.4}$$

where b_k satisfies

$$g_k^T d_{k-1} + b_k > 0, \tag{2.5}$$

we find that (2.3) is equivalent to

$$b_k > 0. \tag{2.6}$$

Thus if β_k is given by (2.4) with b_k satisfying (2.5) and (2.6), we must have that $d_k^T g_k < 0$. Therefore by $d_1 = -g_1$ and the induction principle, all search directions $\{d_k\}$ are downhill.

To be such that the method (1.2), (1.3) and (2.4) is a nonlinear conjugate gradient method, we still need b_k to reduce to $\|g_{k-1}\|^2$ when f is a convex quadratic and the line search is exact. From (2.2) with k replaced by $k-1$, we see that the terms $\|g_{k-1}\|^2$, $d_{k-1}^T y_{k-1}$ and $-d_{k-1}^T g_{k-1}$ all have this property. The three terms are positive if (2.1) and (1.6) hold. Hence we may choose b_k as any convex combination of the three terms:

$$b_k = \mu_k \|g_{k-1}\|^2 + \omega_k d_{k-1}^T y_{k-1} + (1 - \mu_k - \omega_k)(-d_{k-1}^T g_{k-1}), \tag{2.7}$$

where $\mu_k \in [0, 1]$ and $\omega_k \in [0, 1 - \mu_k]$. Consequently, by (2.4) and (2.7),

$$\beta_k = \frac{\|g_k\|^2}{(1 + \omega_k) g_k^T d_{k-1} + \mu_k \|g_{k-1}\|^2 + (1 - \mu_k)(-d_{k-1}^T g_{k-1})}. \tag{2.8}$$

If $\mu_k = \omega_k = 0$, (2.8) reduces to the DY formula. The descent property and global convergence of the DY method are achieved with the Wolfe line search (with any $\sigma < 1$). For the family of methods (2.8), using the Wolfe line search, we can show that if $\sigma \le \frac{1}{4}$, then (2.5) holds, and hence $g_k^T d_k < 0$ for all k.

Although we would be satisfied with its descent property, the family of methods (2.8) has the same drawback as the FR method. Powell [19] observed that the FR method with exact line searches may produce many small steps continuously; namely, if a small step is generated away from the solution, its subsequent steps may also be very short. Since (2.8) reduces to the FR method in the case of exact line searches, we know that the argument applies to the family of methods (2.8).

However, in the same case, the PRP method generates a search direction close to $-g_k$ and hence can avoid the propensity of small steps [19]. Combining FR and PRP, Touati-Ahmed and Storey [23] proposed the hybrid method

$$\beta_k = \max\{0, \min\{\beta_k^{PRP}, \beta_k^{FR}\}\}. \tag{2.9}$$

Like the PRP method, the hybrid method can avoid the propensity of small steps. In addition, its global convergence can be proved under the same assumptions as for the FR method. Hybrid conjugate gradient methods are further considered in [14] and [11]. Gilbert and Nocedal [14] considered the method

$$\beta_k = \max\{-\beta_k^{FR}, \min\{\beta_k^{PRP}, \beta_k^{FR}\}\}, \tag{2.10}$$

which allows negative values of β_k. Dai and Yuan [11] studied the hybrid methods of DY and HS. The numerical results in [11] show that the method

$$\beta_k = \max\{0, \min\{\beta_k^{HS}, \beta_k^{DY}\}\} \tag{2.11}$$

with the Wolfe line search is better than the PRP method with the strong Wolfe line search.

For the above reason, instead of (2.8), we consider the formula

$$\beta_k = \frac{\max\{0, \min\{g_k^T y_{k-1}, \tau_k \|g_k\|^2\}\}}{(\tau_k + \omega_k) g_k^T d_{k-1} + \mu_k \|g_{k-1}\|^2 + (1-\mu_k)(-d_{k-1}^T g_{k-1})}, \tag{2.12}$$

where $\mu_k \in [0,1]$, $\omega_k \in [0, 1-\mu_k]$ and $\tau_k \in [1, +\infty)$ are parameters. If f is a convex quadratic and the line search is exact, then (2.12) reduces to the FR formula, since in this case $g_k^T g_{k-1} = 0$ and $g_k^T d_{k-1} = 0$. So the methods (1.2), (1.3), (2.12) with different values of $\mu_k \in [0,1]$, $\omega_k \in [0, 1-\mu_k]$ and $\tau_k \in [1, +\infty)$ form a three-parameter family of hybrid conjugate gradient methods. Such a family can also avoid the propensity of small steps (a formal description will be given in §4). In addition, it reduces to (2.11) if $\tau_k = 1$, $\mu_k = 0$, $\omega_k = 0$.

3. Descent property of the family of methods (2.12)

In this section, we provide a condition that ensures the descent property of the three-parameter family of hybrid conjugate gradient methods (2.12) with the Wolfe line search. To begin our analyses, define

$$\xi_k = \max\left\{0, \min\left\{\frac{g_k^T y_{k-1}}{\tau_k \|g_k\|^2}, 1\right\}\right\}. \tag{3.1}$$

It is obvious that $\xi_k \in [0,1]$. By (3.1), we write (2.12) as

$$\beta_k = \frac{\xi_k \tau_k \|g_k\|^2}{(\tau_k + \omega_k) g_k^T d_{k-1} + \mu_k \|g_{k-1}\|^2 + (1-\mu_k)(-d_{k-1}^T g_{k-1})}. \tag{3.2}$$

Also define

$$r_k = -\frac{g_k^T d_k}{\|g_k\|^2} \quad \text{and} \quad l_k = \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}}. \tag{3.3}$$

Dividing (2.2) by $-\|g_k\|^2$ and substituting (3.2), we can get that

$$r_k = \frac{[(1-\xi_k)\tau_k + \omega_k] g_k^T d_{k-1} + \mu_k \|g_{k-1}\|^2 + (1-\mu_k)(-d_{k-1}^T g_{k-1})}{(\tau_k + \omega_k) g_k^T d_{k-1} + \mu_k \|g_{k-1}\|^2 + (1-\mu_k)(-d_{k-1}^T g_{k-1})}. \tag{3.4}$$

Using the definitions of r_k and l_k in (3.4), we obtain

$$r_k = \frac{\mu_k + [1-\mu_k - ((1-\xi_k)\tau_k + \omega_k)l_k]r_{k-1}}{\mu_k + [1-\mu_k - (\tau_k + \omega_k)l_k]r_{k-1}}. \tag{3.5}$$

Theorem 3.1. *Consider the family of methods* (1.2), (1.3), (2.12) *with* $\mu_k \in [0,1]$, $\omega_k \in [0, 1-\mu_k]$, $\tau_k \in [1, +\infty)$, *and with* α_k *satisfying* (1.6). *If*

$$\tau_k l_k \le \frac{1}{4}, \tag{3.6}$$

the formula (2.12) *is well defined. Further, for all* $k \ge 1$,

$$0 < r_k \le 2. \tag{3.7}$$

Proof. By (3.5), denote

$$r_k = \frac{t_k}{h_k}, \tag{3.8}$$

where

$$t_k = \mu_k + [1-\mu_k - ((1-\xi_k)\tau_k + \omega_k)l_k]r_{k-1} \tag{3.9}$$

and

$$h_k = \mu_k + [1-\mu_k - (\tau_k + \omega_k)l_k]r_{k-1}. \tag{3.10}$$

Since $d_1 = -g_1$ and $r_1 = 1$, (3.7) holds for $k = 1$. Assume that (3.7) holds for $k-1$, namely,

$$0 < r_{k-1} \le 2. \tag{3.11}$$

It follows from (3.6) and $\tau_k \ge 1$ that

$$l_k \le \frac{1}{4}. \tag{3.12}$$

If $[1-\mu_k - (\tau_k + \omega_k)l_k] > 0$, we have by (3.10), (3.11) and $\mu_k \ge 0$ that $h_k > 0$. If $[1-\mu_k - (\tau_k + \omega_k)l_k] \le 0$, by (3.10), (3.11), (3.6), (3.12) with $\mu_k \in [0,1]$ and $\omega_k \in [0, 1-\mu_k]$, we have

$$\begin{aligned} h_k &\ge \mu_k + 2[1-\mu_k - (\tau_k + \omega_k)l_k] \\ &\ge \mu_k + 2[1-\mu_k - \frac{1}{4}(1+\omega_k)] \\ &\ge \frac{3}{2} - \mu_k - \frac{1}{2}\omega_k \ge \frac{1}{2}. \end{aligned} \tag{3.13}$$

Thus we always have $h_k > 0$. This with (3.8) implies that β_k is well defined. Similarly, we can prove that

$$t_k > 0 \tag{3.14}$$

and

$$2h_k - t_k = \mu_k + [1-\mu_k - ((1+\xi_k)\tau_k + \omega_k)l_k]r_{k-1} \ge 0. \tag{3.15}$$

It follows from (3.8), $h_k > 0$, (3.14) and (3.15) that $0 < r_k \le 2$. Thus, by induction, $\{\beta_k\}$ is well defined and (3.7) is true for all $k \ge 1$. □

Note by (1.6) that $l_k \le \sigma$. Since it is preferred to set σ equal to a small value in the implementations of conjugate gradient methods (a typical value of σ is 0.1, see [11, 14]), we see that the condition (3.6) is not strict and allows relatively large values of τ_k.

4. Global convergence

Assume that $g_k \neq 0$ for all k, for otherwise a stationary point has been found. We give the following basic assumptions on the objective function.

Assumption 4.1. *(i) The level set* $\mathcal{L} = \{x \in R^n : f(x) \leq f(x_1)\}$ *is bounded, where* x_1 *is the initial point.*

(ii) In some neighborhood $\mathcal{N}$ *of* $\mathcal{L}$, f *is differentiable and its gradient* ∇f *is Lipschitz continuous in* $\mathcal{N}$*, namely, there exists a constant* $L > 0$ *such that*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \qquad \text{for any } x,\, y \in \mathcal{N}. \tag{4.1}$$

Denote $s_{k-1} = x_k - x_{k-1}$ and suppose that

$$0 < \gamma \leq \|g_k\| \leq \bar{\gamma}, \quad \text{for all } k \geq 1. \tag{4.2}$$

We say [3] that a method (1.2)–(1.3) has Property (#) if there exist a positive and uniformly bounded sequence $\{\psi_k\}$ and constants $b \geq 1$ and $\lambda > 0$ such that $|\beta_k| \leq b\,\frac{\psi_k}{\psi_{k-1}}$ for all k, and if $\|s_{k-1}\| \leq \lambda$, then $|\beta_k| \leq \frac{1}{b}\frac{\psi_k}{\psi_{k-1}}$.

Under Assumption 4.1 on f, we state a general lemma for any method (1.2)–(1.3) having Property (#).

Lemma 4.2 ([3]). *Suppose that Assumption 4.1 holds. Consider any method* (1.2)–(1.3) *with* $\beta_k \geq 0$ *and Property (#). If the steplength* α_k *satisfies the Wolfe conditions* (1.4), (1.6) *and the descent condition* $g_k^T d_k < 0$*, then either*

$$\liminf_{k\to\infty} \|d_k\| < +\infty, \tag{4.3}$$

or the method converges in the sense that

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{4.4}$$

For the family of methods (2.12) that satisfies (3.6), we can check that Property (#) holds. In fact, define $\psi_k = -g_k^T d_k$. It follows by (3.7) and (4.2) that $\{\psi_k\}$ is positive and uniformly bounded. By (3.2)-(3.4), we write

$$\beta_k = \frac{\xi_k \tau_k}{\eta_k}\,\frac{\psi_k}{\psi_{k-1}}, \tag{4.5}$$

where

$$\eta_k = 1 - \mu_k + \mu_k r_{k-1}^{-1} - [(1 - \xi_k)\tau_k + \omega_k] l_k. \tag{4.6}$$

By (3.7), (3.6), (3.12), $\xi_k \in [0,1]$, $\mu_k \in [0,1]$ and $\omega_k \in [0, 1 - \mu_k]$, we can show that

$$\eta_k \geq \frac{3}{4} - \frac{1}{2}\mu_k - \frac{1}{4}\omega_k \geq \frac{1}{4}. \tag{4.7}$$

It follows from (4.5), (4.7), $\xi_k \tau_k \geq 0$ and $\psi_k > 0$ that

$$\beta_k \geq 0. \tag{4.8}$$

Denote $b = \frac{8\bar{\gamma}}{\gamma}$ and $\lambda = \frac{\gamma}{4Lb}$. Noting by (3.1) and the Schwarz inequality that

$$\xi_k \tau_k \leq \frac{|g_k^T y_{k-1}|}{\|g_k\|^2} \leq \frac{\|y_{k-1}\|}{\|g_k\|}, \tag{4.9}$$

we have by (4.5), (4.7), (4.9) and (4.2) that

$$\beta_k \leq \frac{4(\|g_{k-1}\| + \|g_k\|)}{\|g_k\|}\,\frac{\psi_k}{\psi_{k-1}} \leq \frac{8\bar{\gamma}}{\gamma}\,\frac{\psi_k}{\psi_{k-1}} = b\,\frac{\psi_k}{\psi_{k-1}}. \tag{4.10}$$

If $\|s_{k-1}\| \le \lambda$, then by (4.5), (4.7), (4.9), (4.1) and (4.2),

$$\beta_k \le \frac{4L\|s_{k-1}\|}{\|g_k\|}\frac{\psi_k}{\psi_{k-1}} \le \frac{4L\lambda}{\gamma}\frac{\psi_k}{\psi_{k-1}} = \frac{1}{b}\frac{\psi_k}{\psi_{k-1}}. \tag{4.11}$$

Relations (4.8), (4.10) and (4.11) indicate that the family of methods (2.12) that satisfies (3.6) has Property (#).

Now we are ready to give our main convergence result.

Theorem 4.3. *Suppose that Assumption 4.1 holds. Consider the family of methods* (1.2)*,* (1.3)*,* (2.12) *with* $\mu_k \in [0,1]$*,* $\omega_k \in [0, 1-\mu_k]$*,* $\tau_k \in [1,\infty)$*, and with the Wolfe line search* (1.4) *and* (1.6)*. Assume that* (3.6) *holds. Then we have* (4.4) *if one of the following conditions holds: (i)* l_k *is uniformly bounded; (ii)* ω_k *is bounded away from zero; (iii)* μ_k *is bounded away from zero.*

Proof. We proceed by contradiction, assuming that

$$\|g_k\| \ge \gamma, \quad \text{for some } \gamma > 0 \text{ and all } k \ge 1. \tag{4.12}$$

According to the previous discussions, we know that Property (#) holds and $\beta_k \ge 0$. We now prove

$$\lim_{k\to\infty} \|d_k\| = +\infty \tag{4.13}$$

for (i), (ii) and (iii), in turn.

(i) It follows from (4.12), the uniform boundness of l_k and Corollary 2.4 in [5] that

$$\sum_{k\ge 1} \frac{1}{\|d_k\|^2} < +\infty, \tag{4.14}$$

which gives (4.13).

(ii) Assume that

$$\omega_k \ge \epsilon, \quad \text{for some } \epsilon > 0 \text{ and all } k \ge 1. \tag{4.15}$$

(3.5) implies that r_k is monotonically decreasing as $l_k \to -\infty$. Hence

$$\begin{aligned} r_k \quad &> \quad \lim_{l_k\to-\infty} \frac{\mu_k + [1-\mu_k - ((1-\xi_k)\tau_k + \omega_k)l_k]r_{k-1}}{\mu_k + [1-\mu_k - (\tau_k+\omega_k)l_k]r_{k-1}} \\ &= \quad \frac{(1-\xi_k)\tau_k + \omega_k}{\tau_k + \omega_k}. \end{aligned} \tag{4.16}$$

In addition, Assumption 4.1 implies that

$$\|g_k\| \le \bar\gamma, \quad \text{for some } \bar\gamma > 0 \text{ and all } k \ge 1. \tag{4.17}$$

The definition of ξ_k, (4.12) and (4.17) show that

$$0 \le \xi_k \le \min\{1, \frac{\|y_{k-1}\|}{\tau_k\|g_k\|}\} \le \min\{1, \frac{\|g_{k-1}\| + \|g_k\|}{\tau_k\|g_k\|}\} \le \min\{1, \frac{2\bar\gamma}{\gamma\tau_k}\}. \tag{4.18}$$

If $\tau_k \le \frac{4\bar\gamma}{\gamma}$, we have by (4.16) and (4.15) that $r_k \ge \epsilon/(4\bar\gamma\gamma^{-1} + \epsilon)$. If $\tau_k > \frac{4\bar\gamma}{\gamma}$, it follows by (4.18) that $\xi_k \le \frac{1}{2}$. This and (4.16) indicate that $r_k \ge \frac{1}{2}$. Thus we always have

$$r_k \ge \min\{\frac{1}{2}, \frac{\epsilon\gamma}{4\bar\gamma + \epsilon\gamma}\}. \tag{4.19}$$

Since, under Assumption 4.1 on f, any descent method (1.2) with the Wolfe line search gives the Zoutendijk condition [24],

$$\sum_{k\geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty, \tag{4.20}$$

we know from this, (4.19) and (4.12) that (4.13) holds.

(iii) Assume that

$$\mu_k \geq \epsilon, \quad \text{for some } \epsilon > 0 \text{ and all } k \geq 1. \tag{4.21}$$

If (4.13) is false, there exists an infinite subsequence $\{k_i\}$ such that

$$\|d_{k_i}\| \leq M, \quad \text{for some } M < +\infty \text{ and all } i \geq 1. \tag{4.22}$$

It follows from this and (4.17) that

$$|g_{k_i+1}^T d_{k_i}| \leq \bar{\gamma} M. \tag{4.23}$$

Using (3.4), (4.12), (4.17), (4.18), (4.21), (4.23) and $\omega_k \in [0,1]$, we can similarly to (4.19) prove that

$$r_{k_i+1} \geq \min\{\frac{1}{2}, \frac{\epsilon\gamma^3}{(4\bar{\gamma}+\gamma)\bar{\gamma}M + \epsilon\gamma^3}\}. \tag{4.24}$$

By this, the Zoutendijk condition (4.20) and (4.12), we get

$$\liminf_{i\to\infty} \|d_{k_i+1}\| = +\infty. \tag{4.25}$$

Still define h_k as in (3.10). If $1 - \mu_k - (\tau_k + \omega_k) l_k > 0$, we have by (4.21) and (3.7) that $h_k \geq \epsilon$. If $1 - \mu_k - (\tau_k + \omega_k) l_k \leq 0$, it follows by (3.13) that $h_k \geq \frac{1}{2}$. Thus we always have

$$h_k \geq \min\{\epsilon, \frac{1}{2}\}. \tag{4.26}$$

By (3.2), (4.26), (4.9), (4.17) and (4.12), we obtain

$$\beta_{k_i+1} = \frac{\xi_{k_i+1}\tau_{k_i+1}\|g_{k_i+1}\|^2}{\zeta_{k_i+1}\|g_{k_i}\|^2} \leq c, \tag{4.27}$$

where $c = 2\bar{\gamma}^3/(\gamma^3 \min\{\epsilon, \frac{1}{2}\})$ is a positive constant. It follows by (1.3), the triangle inequality, (4.17) and (4.27) that

$$\|d_{k_i+1}\| \leq \bar{\gamma} + c\,\|d_{k_i}\|. \tag{4.28}$$

Letting $i \to \infty$ in (4.28), we find that (4.22) and (4.25) give a contradiction. So (4.13) also holds.

Thus for each case (i), (ii) and (iii), (4.13) is true. By Lemma 4.2, we must have (4.4), contradicting (4.12). Therefore this theorem is true. □

Since the condition (1.5) implies the bound $|l_k| \leq \sigma$, one direct corollary of Theorem 4.3 is that the family of hybrid conjugate gradient methods (2.12) with the strong Wolfe line search converges globally for general functions. Assume that the objective function f is uniformly convex and there exists some positive constant $\eta > 0$ such that

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq \eta \|x - y\|^2, \quad \text{for all } x, y \in \mathcal{L}. \tag{4.29}$$

Then by (1.4), (4.29) and Taylor's series expansion, it is easy to show that

$$\alpha_k \|d_k\|^2 \le 2(1-\delta)\eta^{-1}|g_k^T d_k|, \quad \text{for all } k \ge 1. \tag{4.30}$$

In addition, by the triangle inequality and (4.1),

$$|g_{k+1}^T d_k| \le |g_k^T d_k| + |(g_{k+1}-g_k)^T d_k| \le |g_k^T d_k| + \alpha_k L \|d_k\|^2. \tag{4.31}$$

The above two relations imply that $|l_k|$ is also uniformly bounded. Thus, by Theorem 4.3, the family with the Wolfe line search is globally convergent for uniformly convex functions.

5. Numerical results

In this section, we present some numerical results for the family of hybrid conjugate gradient methods (2.12). Our tests were done on an SGI Indigo workstation with double precision. All the codes are written in FORTRAN. For each method, we use the Wolfe line search (1.4) and (1.5) with $\delta = 0.01$ and some value of σ. The initial value of α_k is always set equal to 1. Our test problems are drawn from Moré *et al.* [17]. See Tables 5.1 and 5.2. The first column "P" denotes the problem number in [17], and the second gives the name of the problem. We tested each problem with two different values of n ranging from $n = 20$ to $n = 10000$. The numerical results are given in the form of I/F/G, where I, F, G denote numbers of iterations, function evaluations, and gradient evaluations. The stopping condition is

$$\|g_k\| \le 10^{-6}. \tag{5.1}$$

Our numerical experiments were divided into two parts. First, we tested the family (2.12) with $\tau_k \equiv \tau \in \{1, 2, 4\}$. The parameter σ corresponding to τ is set equal to $\frac{1}{4\tau}$, which is the largest that ensures the condition (3.6). This part of the numerical results are listed in Table 5.1, where the column (2.11) means the hybrid method (2.11) and the other stand for the method (3.6) with $\mu_k = \omega_k = 0$ and some values of (τ, σ). Second, we tested the family (2.12) with variable τ_k. Specifically, we are interested in the following choice of τ_k:

$$\tau_k = \max\{1, \min\{\nu |l_{k-1}|^{-1}, 4\}\}, \tag{5.2}$$

where ν is some positive constant. The idea behind (5.2) is that we force the method to be closer to (2.11) if the line search is more inexact; otherwise, we use a relatively large value of τ_k such that the conjugacy quantity $d_k^T y_{k-1}$ tends to zero. See Table 5.2 for the numerical results of the method (2.12) with $\mu_k = \omega_k = 0$, τ_k given by (5.2), and different values of (σ, ν).

We compared each method with the method (2.11). Denote by F_a and G_a the numbers of function evaluations and gradient evaluations required by method (a) for some problem. Then we say that method (a) beats method (b) if $F_a < F_b$ and $G_a \le G_b$ or if $F_a \le F_b$ and $G_a < G_b$. If it happens that $(F_a - F_b)(G_a - G_b) < 0$, we decide who is the winner by their CPU times (Since this seldom occurs, we do not list the CPU times in the tables). The numbers of wins for each method comparing with the method (2.11) are given at the bottom of the tables. We can see that the method $(4, \frac{1}{16})$ in Table 5.1 and the method $(0.25, 0.05)$ in Table 5.2 perform similarly to or even slightly better than the hybrid method (2.11). Further, if we only consider the test problems whose dimensions are not less than 100, then the numbers of wins of the methods $(4, \frac{1}{16})$ and $(0.25, 0.05)$ compared with the method (2.11) are both 7 : 3. This means that the two methods perform better than the hybrid method (2.11) for relatively large problems. To sum up, although we do not know yet what are the best choices for the parameters in (2.12), our numerical results indicate that the introduction of the hybrid family (2.12) is worthwhile.

TABLE 5.1. Comparing different conjugate gradient methods

P	Name	n	(2.11)	$(1, 1/4)$	$(2, 1/8)$	$(4, 1/16)$
24	Penalty 2	20	135/419/228	159/446/224	125/378/205	301/926/477
		40	122/366/177	137/399/176	418/1127/489	186/559/275
25	Variably	20	5/30/10	5/30/10	5/30/10	5/30/10
	dimensioned	50	9/51/17	9/51/17	9/51/16	8/45/14
35	Chebyquad	20	100/321/119	124/376/138	140/434/154	155/484/171
		50	350/1156/406	437/1353/460	375/1229/427	438/1407/487
30	Broyden	50	50/158/58	50/156/56	51/161/59	60/186/66
	tridiagonal	500	58/183/67	64/199/71	52/161/57	54/170/62
31	Broyden	50	30/113/49	19/62/23	18/58/21	31/112/46
	banded	500	23/74/27	20/62/22	20/63/23	24/78/29
22	Extended	100	66/203/87	63/179/73	83/237/102	65/195/81
	Powell	1000	66/203/87	72/202/82	86/250/110	65/195/81
26	Trigonome-	100	58/97/95	56/83/82	57/91/90	56/98/96
	tric	1000	52/87/87	54/80/80	58/94/94	56/99/99
21	Extended	1000	28/87/39	27/97/44	39/151/68	21/80/39
	Rosenbrock	10000	28/87/39	27/97/44	39/151/68	22/82/41
23	Penalty 1	1000	54/154/110	74/178/132	62/162/119	27/86/59
		10000	35/111/66	30/121/59	43/137/89	22/63/42
	winners		-	8:8	6:11	9:8

TABLE 5.2. Comparing different conjugate gradient methods

P	Name	n	(0.1,0.05)	(0.1,0.25)	(0.25,0.05)	(0.25,0.25)
24	Penalty 2	20	165/497/278	135/443/249	159/464/254	170/488/260
		40	141/426/215	248/690/310	149/443/213	158/462/222
25	Variably	20	5/30/10	5/30/10	5/30/10	5/30/10
	dimensioned	50	7/44/15	8/45/14	8/46/15	9/55/18
35	Chebyquad	20	103/323/119	138/432/158	135/422/152	114/361/129
		50	372/1200/416	389/1244/429	490/1505/507	374/1169/398
30	Broyden	50	55/173/63	57/179/65	50/156/56	52/163/59
	tridiagonal	500	59/186/68	63/198/72	66/205/73	56/174/62
31	Broyden	50	31/115/49	35/128/53	19/62/23	18/58/21
	banded	500	23/74/27	23/74/27	20/62/22	19/59/21
22	Extended	100	72/209/88	87/254/108	61/171/69	99/287/116
	Powell	1000	85/245/105	83/242/106	73/201/81	115/334/136
26	Trigonome-	100	58/99/97	62/97/96	60/91/89	63/96/94
	tric	1000	55/95/95	60/105/105	54/83/82	63/93/93
21	Extended	1000	28/113/55	28/99/45	29/101/46	31/113/50
	Rosenbrock	10000	28/113/55	29/102/46	29/101/46	31/113/50
23	Penalty 1	1000	42/116/76	36/104/67	45/112/71	62/143/102
		10000	25/96/52	28/95/55	28/114/54	33/98/58
	winners		4:12	3:13	9:8	7:10

6. Discussions

This paper presents a three-parameter family of hybrid conjugate gradient methods for unconstrained optimization. As mentioned in Section 2, this family of methods has the hybrid method (2.11) as a special case. It is known [11] that the hybrid method (2.11) with the Wolfe line search is globally convergent for general functions. However, our main convergence theorem, Theorem 4.3, does not cover this result. We wonder whether Theorem 4.3 holds for all the methods in the family.

Although we do not know yet what are the best choices for the parameters in (2.12), the numerical results of this paper show that the family of hybrid conjugate gradient methods is very promising. Both the theoretical analyses and numerical results with the family again show that it is possible to use the Wolfe line search in the nonlinear conjugate gradient field.

Acknowledgments

The author would like to thank the associate editor and an anonymous referee, whose comments and suggestions improved this paper greatly.

References

[1] C. G. Broyden, *The convergence of a class of double-rank minimization algorithms 1. general considerations* , J. Inst. Math. Appl. 6 (1970) 76–90. MR **55:**6841

[2] R. Byrd, J. Nocedal and Y. Yuan, *Global convergence of a class of quasi-Newton methods on convex problems*, SIAM J. Numer. Anal. 24 (1987), 1171-1190. MR **88m:**65100

[3] Y. H. Dai, *Convergence properties of nonlinear conjugate gradient methods (II)*, Research report AMSS-1999-082, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, 1999.

[4] Y. H. Dai, *New Properties of A Nonlinear Conjugate Gradient Method*, Numerische Mathematics 89 : 1 (2001), pp. 83-98. MR **2002f:**90118

[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*, SIAM Journal on Optimization 10 : 2 (1999), pp. 345–358. MR **2001a:**90059

[6] Y. H. Dai and Y. Yuan, *Convergence properties of the Fletcher-Reeves method*, IMA Journal of Numerical Analysis 16 : 2 (1996), pp. 155-164. MR **97i:**65099

[7] Y. H. Dai and Y. Yuan, *A class of globally convergent conjugate gradient methods*, 1998 (to appear in: Sciences in China, Series A).

[8] Y. H. Dai and Y. Yuan, *An extended class of nonlinear conjugate gradient methods.* In: (D. Li, ed.) Proceedings of the 5th International Conference on Optimization: Techniques and Applications (December 2001, Hong Kong), pp. 778–785.

[9] Y. H. Dai and Y. Yuan, *A three-parameter family of conjugate gradient methods*, Math. Comp. 70 (2001), 1155–1167. MR **2002b:**90128

[10] Y. H. Dai and Y. Yuan, *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM Journal on Optimization 10 : 1 (1999), pp. 177–182. MR **2000i:**90074

[11] Y. H. Dai and Y. Yuan, *An efficient hybrid conjugate gradient method for unconstrained optimization*, Ann. Oper. Res. 103 (2001), 33–47.

[12] R. Fletcher, *Practical Methods of Optimization vol. 1: Unconstrained optimization,* John Wiley & Sons (New York), 1987. MR **89j:**65050

[13] R. Fletcher and C. Reeves, *Function minimization by conjugate gradients*, Comput. J. 7 (1964), pp. 149-154. MR **32:**4827

[14] J. C. Gilbert and J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM J. Optimization 2:1 (1992), pp. 21-42. MR **92k:**90089

[15] 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. MR **15:**651a

[16] Y. Liu and C. Storey, *Efficient Generalized Conjugate Gradient Algorithms, Part 1: Theory*, Journal of Optimization Theory and Applications, Vol. 69 (1991), 129-137. MR **92e:**90077
[17] J. J. Moré, B. S. Garbow, and K. E. Hillstrom, *Testing unconstrained optimization software*, ACM Transactions on Mathematical Software 7 (1981) 17-41. MR **83b:**90144
[18] L. Nazareth, *Conjugate-gradient methods*, Encyclopedia of Optimization (C. Floudas and P. Pardalos, eds.), Vol. I, Kluwer, Dordrecht, 2001. MR **2002h:**00006
[19] M. J. D. Powell, *Restart procedures for the conjugate gradient method*, Math. Program. 2 (1977), 241-254. MR **57:**18099
[20] 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. MR **85g:**49035
[21] E. Polak and G. Ribière, *Note sur la convergence de directions conjugées*, Rev. Francaise Informat Recherche Opertionelle, 3e Année 16 (1969), pp. 35-43. MR **40:**8232
[22] B. T. Polyak, *The conjugate gradient method in extreme problems*, USSR Comp. Math. and Math. Phys. 9 (1969), pp. 94-112. MR **41:**2899
[23] D. Touati-Ahmed and C. Storey, *Efficient hybrid conjugate gradient techniques*, J. Optim. Theory Appl. 64 (1990), pp. 379-397. MR **91c:**65046
[24] G. Zoutendijk, *Nonlinear Programming, Computational Methods*, in: Integer and Nonlinear Programming (J. Abadie, ed.), 1970, pp. 37-86. MR **55:**10015

State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and System Sciences, Chinese Academy of Sciences, P. O. Box 2719, Beijing 100080, Peoples Republic of China

E-mail address: dyh@lsec.cc.ac.cn