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

\[(1.1) \quad \min f(x), \quad x \in \mathbb{R}^n,\]

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

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

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

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

\[(1.4) \quad f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k,\]

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

where \(\delta \in (0, \frac{1}{2})\) and \(\sigma \in (\delta, 1)\). In the conjugate gradient field, it is also possible to use the Wolfe line search, which calculates an \(\alpha_k\) satisfying \((1.4)\) and \((1.5)\) and

\[(1.6) \quad g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k.\]
For the scalar $\beta_k$, many formulas have been proposed. Some of them are called the FR [13], PRP [21, 22], DY [10], HS [15], CD [12], and LS [10] ones, and are given by

\[
\begin{align*}
\beta_k^{FR} &= \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \\
\beta_k^{DY} &= \frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}}, \\
\beta_k^{CD} &= \frac{\|g_k\|^2}{-g_{k-1}^T d_{k-1}},
\end{align*}
\]

where $g_{k-1} = g_k - g_{k-1}$ 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 $\alpha_k$ satisfies $1.4$-$1.5$ with $\sigma \leq \frac{1}{2}$ (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

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

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 $\beta_k$ and proposed a two-parameter family:

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

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

\[
\beta_k = \frac{||g_k||^2 - \lambda_k g_k^T d_{k-1}}{||g_{k-1}||^2 + \mu_k g_{k-1}^T d_{k-1} - \omega_k d_{k-1}^T g_{k-1}}, \quad \lambda_k, \mu_k \in [0, 1] \text{ and } \omega_k \in [0, 1 - \mu_k]
\]

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 $\alpha_k$ satisfying the Wolfe conditions (1.4) and (1.6). Assume that the search direction $d_{k-1}$ is downhill, namely,

$$g_k^T d_{k-1} < 0.$$  

(2.1)

It follows from (1.3) that

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

(2.2)

Then the descent property of $d_k$ requires

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

(2.3)

Assuming that

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

(2.4)

where $b_k$ satisfies

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

(2.5)

we find that (2.3) is equivalent to

$$b_k > 0.$$  

(2.6)

Thus if $\beta_k$ is given by (2.4) with $b_k$ satisfying (2.5) and (2.6), we must have that $d_1 = -g_1$ 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:

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

(2.7)

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

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

(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 $\sigma \leq \frac{1}{4}$, then (2.8) 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 $\beta_k$. 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

$$\beta_k = \max\{0, \min\{g_k^T g_{k-1}, \tau_k \|g_k\|^2\}\} \tag{3.1}$$

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 [12], [13], [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

$$\xi_k = \max\{0, \min\{g_k^T g_{k-1}, \tau_k \|g_k\|^2\}\}. \tag{3.2}$$

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

$$\beta_k = \frac{\xi_k \tau_k \|g_k\|^2}{(\tau_k + \omega_k)g_k^T d_{k-1} + \mu_k \|g_{k-1}\|^2 + (1 - \mu_k)(-d_{k-1}^T g_{k-1})}. \tag{3.3}$$

Also define

$$r_k = -\frac{g_k^T d_k}{\|g_k\|^2} \quad \text{and} \quad l_k = \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}}. \tag{3.4}$$

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

$$r_k = \frac{(1 - \xi_k)\tau_k + \omega_k)g_k^T d_{k-1} + \mu_k \|g_{k-1}\|^2 + (1 - \mu_k)(-d_{k-1}^T g_{k-1})}{(\tau_k + \omega_k)g_k^T d_{k-1} + \mu_k \|g_{k-1}\|^2 + (1 - \mu_k)(-d_{k-1}^T g_{k-1})}.$$
Using the definitions of $r_k$ and $l_k$ in (3.4), we obtain
\begin{equation}
(3.5)\quad 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}}.
\end{equation}

**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 $\alpha_k$ satisfying (1.6). If
\begin{equation}
(3.6)\quad \tau_k l_k \leq \frac{1}{4},
\end{equation}
the formula (2.12) is well defined. Further, for all $k \geq 1$,
\begin{equation}
(3.7)\quad 0 < r_k \leq 2.
\end{equation}

**Proof.** By (3.5), denote
\begin{equation}
(3.8)\quad r_k = \frac{t_k}{h_k},
\end{equation}
where
\begin{equation}
(3.9)\quad t_k = \mu_k + [1 - \mu_k - ((1 - \xi_k)\tau_k + \omega_k)l_k]r_{k-1}
\end{equation}
and
\begin{equation}
(3.10)\quad h_k = \mu_k + [1 - \mu_k - (\tau_k + \omega_k)l_k]r_{k-1}.
\end{equation}
Since $d_1 = -g_1$ and $r_1 = 1$, (3.7) holds for $k = 1$. Assume that (3.7) holds for $k - 1$, namely,
\begin{equation}
(3.11)\quad 0 < r_{k-1} \leq 2.
\end{equation}
It follows from (3.6) and $\tau_k \geq 1$ that
\begin{equation}
(3.12)\quad l_k \leq \frac{1}{4}.
\end{equation}
If $[1 - \mu_k - (\tau_k + \omega_k)l_k] > 0$, we have by (3.10), (3.11) and $\mu_k \geq 0$ that $h_k > 0$. If $[1 - \mu_k - (\tau_k + \omega_k)l_k] \leq 0$, by (3.10), (3.11), (3.9), (3.12) with $\mu_k \in [0, 1]$, and $\omega_k \in [0, 1 - \mu_k]$, we have
\begin{align}
(3.13)\quad h_k & \geq \mu_k + 2[1 - \mu_k - (\tau_k + \omega_k)l_k] \\
& \geq \mu_k + 2[1 - \mu_k - \frac{1}{4}(1 + \omega_k)] \\
& \geq \frac{3}{2} - \mu_k - \frac{1}{2}\omega_k \geq \frac{1}{2}.
\end{align}
Thus we always have $h_k > 0$. This with (3.8) implies that $\beta_k$ is well defined. Similarly, we can prove that
\begin{equation}
(3.14)\quad t_k > 0
\end{equation}
and
\begin{equation}
(3.15)\quad 2h_k - t_k = \mu_k + [1 - \mu_k - ((1 + \xi_k)\tau_k + \omega_k)l_k]r_{k-1} \geq 0.
\end{equation}
It follows from (3.8), $h_k > 0$, (3.12) and (3.13) that $0 < r_k \leq 2$. Thus, by induction, \{\beta_k\} is well defined and (3.7) is true for all $k \geq 1$. \hfill \square

Note by (1.8) that $l_k \leq \sigma$. Since it is preferred to set $\sigma$ equal to a small value in the implementations of conjugate gradient methods (a typical value of $\sigma$ is 0.1, see [11, 14]), we see that the condition (3.6) is not strict and allows relatively large values of $\tau_k$. 

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use
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 \( \mathcal{L} = \{ x \in \mathbb{R}^n : f(x) \leq f(x_1) \} \) is bounded, where \( x_1 \) is the initial point.

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

\[
\| \nabla f(x) - \nabla f(y) \| \leq L \| x - y \|, \quad \text{for any } x, y \in \mathcal{N}.
\]

Denote \( s_{k-1} = x_k - x_{k-1} \) and suppose that

\[
0 < \gamma \leq \| g_k \| \leq \bar{\gamma}, \quad \text{for all } k \geq 1.
\]

We say 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 \( |\beta_k| \leq \frac{\lambda}{b} \frac{\psi_k}{\psi_{k-1}} \).

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 \( \alpha_k \) satisfies the Wolfe conditions (1.3), (1.6) and the descent condition \( g_k^T d_k < 0 \), then either

\[
\liminf_{k \to \infty} \| d_k \| < +\infty,
\]

or the method converges in the sense that

\[
\liminf_{k \to \infty} \| g_k \| = 0.
\]

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

\[
\beta_k = \frac{\xi_k \tau_k}{\eta_k} \frac{\psi_k}{\psi_{k-1}},
\]

where

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

By (3.4), (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

\[
\eta_k \geq \frac{3}{4} - \frac{1}{4} \mu_k - \frac{1}{4} \omega_k \geq \frac{1}{4}.
\]

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

\[
\beta_k \geq 0.
\]

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

\[
\xi_k \tau_k \leq \frac{\| y_{k-1} \|}{\| g_k \|^2} \leq \frac{\| y_{k-1} \|}{\| g_k \|},
\]

we have by (4.5), (4.7), (1.9) and (1.2) that

\[
\beta_k \leq \frac{4(\| g_{k-1} \| + \| g_k \|)}{\| g_k \|} \frac{\psi_k}{\psi_{k-1}} \leq \frac{8\bar{\gamma}}{\gamma} \frac{\psi_k}{\psi_{k-1}} = b \frac{\psi_k}{\psi_{k-1}}.
\]
If \( \|s_{k-1}\| \leq \lambda \), then by (4.5), (4.7), (4.9), (4.1) and (4.2),

\[
\beta_k \leq \frac{4L\|s_{k-1}\|}{\|g_k\|} \leq \frac{4L\lambda}{\gamma} \frac{\psi_k}{\psi_{k-1}} = \frac{1}{b} \frac{\psi_k}{\psi_{k-1}}.
\]

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) \( \omega_k \) is bounded away from zero; (iii) \( \mu_k \) is bounded away from zero.

**Proof.** We proceed by contradiction, assuming that

\[
\|g_k\| \geq \gamma, \text{ for some } \gamma > 0 \text{ and all } k \geq 1.
\]

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

\[
\lim_{k \to \infty} \|d_k\| = +\infty
\]

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

\[
\sum_{k \geq 1} \frac{1}{\|d_k\|^2} < +\infty,
\]

which gives (4.13).

(ii) Assume that

\[
\omega_k \geq \epsilon, \text{ for some } \epsilon > 0 \text{ and all } k \geq 1.
\]

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

\[
r_k > \lim_{l_k \to -\infty} \frac{\mu_k + [1 - \mu_k - ((1 - \xi_k)\tau_k + \omega_k)l_k]r_{k-1}}{\mu_k + [1 - \mu_k - (\tau_k + \omega_k)l_k]r_{k-1}} = \frac{(1 - \xi_k)\tau_k + \omega_k}{\tau_k + \omega_k}.
\]

In addition, Assumption 4.1 implies that

\[
\|g_k\| \leq \bar{\gamma}, \text{ for some } \bar{\gamma} > 0 \text{ and all } k \geq 1.
\]

The definition of \( \xi_k \), (4.12) and (4.17) show that

\[
0 \leq \xi_k \leq \min\{1, \frac{\|g_{k-1}\|}{\tau_k\|g_k\|}\} \leq \min\{1, \frac{\|g_{k-1}\|}{\tau_k\|g_k\|} + \frac{\|g_k\|}{\tau_k\|g_k\|}\} \leq \min\{1, \frac{2\bar{\gamma}}{\bar{\gamma}\tau_k}\}.
\]

If \( \tau_k \leq \frac{4\epsilon}{\bar{\gamma}} \), we have by (4.10) and (4.11) that \( r_k \geq \epsilon/(4\bar{\gamma}\gamma^{-1} + \epsilon) \). If \( \tau_k > \frac{4\epsilon}{\bar{\gamma}} \), it follows by (4.18) that \( \xi_k \leq \frac{1}{2} \). This and (4.10) indicate that \( r_k \geq \frac{1}{2} \). Thus we always have

\[
r_k \geq \min\{\frac{1}{2}, \frac{\epsilon\gamma}{4\bar{\gamma} + \epsilon}\}.
\]
Since, under Assumption 4.1 on \( f \), any descent method (1.2) with the Wolfe line search gives the Zoutendijk condition \([24]\),

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

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

(iii) Assume that

\[
\mu_k \geq \epsilon, \quad \text{for some } \epsilon > 0 \text{ and all } k \geq 1.
\]

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

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

It follows from this and (4.17) that

\[
|g_{k_i+1}^T d_{k_i}| \leq \bar{\gamma} M.
\]

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

\[
r_{k_i+1} \geq \min\left\{ 1, \frac{\epsilon \gamma^3}{4 \gamma + \gamma M + \epsilon \gamma^3} \right\}.
\]

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

\[
\liminf_{i \to \infty} \|d_{k_i+1}\| = +\infty.
\]

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 (4.19) that \( h_k \geq \frac{1}{2} \). Thus we always have

\[
h_k \geq \min\{\epsilon, \frac{1}{2}\}.
\]

By (3.2), (4.26), (3.9), (4.17) and (4.12), we obtain

\[
\beta_{k_i+1} = \frac{\zeta_{k_i+1} \tau_{k_i+1} \|g_{k_i+1}\|^2}{\zeta_{k_i+1} \|g_{k_i}\|^2} \leq c,
\]

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

\[
\|d_{k_i+1}\| \leq \bar{\gamma} + c \|d_{k_i}\|.
\]

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 (4.14), contradicting (4.12). Therefore this theorem is true. \( \square \)

Since the condition (1.9) 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

\[
(\nabla f(x) - \nabla f(y))^T (x - y) \geq \eta \|x - y\|^2, \quad \text{for all } x, y \in \mathcal{L}.
\]
Then by (1.4), (4.29) and Taylor’s series expansion, it is easy to show that
\[ \alpha_k \| d_k \|^2 \leq 2(1 - \delta)\eta^{-1}|g_k^T d_k|, \quad \text{for all } k \geq 1. \]
In addition, by the triangle inequality and (4.1),
\[ |g_{k+1}^T d_k| \leq |g_k^T d_k| + |(g_{k+1} - g_k)^T d_k| \leq |g_k^T d_k| + \alpha_k L \| d_k \|^2. \]
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 \( \sigma \). The initial value of \( \alpha_k \) 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
\[ \| g_k \| \leq 10^{-6}. \]

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 \( \sigma \) corresponding to \( \tau \) is set equal to \( \frac{1}{\tau} \), 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 \( (\tau, \sigma) \). Second, we tested the family (2.12) with variable \( \tau_k \). Specifically, we are interested in the following choice of \( \tau_k \):
\[ \tau_k = \max\{1, \min\{\nu|l_{k-1}|^{-1}, 4\}\}, \]
where \( \nu \) 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 \( \tau_k \) 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 \( \mu_k = \omega_k = 0 \), \( \tau_k \) given by (5.2), and different values of \( (\sigma, \nu) \).

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 \( G_a \leq G_b \) or if \( F_a \leq F_b \) 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, 1/17) 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

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

Table 5.2. Comparing different conjugate gradient methods

<table>
<thead>
<tr>
<th>P</th>
<th>Name</th>
<th>n</th>
<th>(0.1,0.05)</th>
<th>(0.1,0.25)</th>
<th>(0.25,0.05)</th>
<th>(0.25,0.25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>Penalty 2</td>
<td>20</td>
<td>165/407/278</td>
<td>135/443/249</td>
<td>159/464/254</td>
<td>170/488/260</td>
</tr>
<tr>
<td></td>
<td></td>
<td>40</td>
<td>141/246/215</td>
<td>248/690/310</td>
<td>149/443/213</td>
<td>158/462/222</td>
</tr>
<tr>
<td>25</td>
<td>Variably dimensioned</td>
<td>20</td>
<td>5/30/10</td>
<td>5/30/10</td>
<td>5/30/10</td>
<td>5/30/10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>7/44/15</td>
<td>8/48/14</td>
<td>8/46/15</td>
<td>9/55/18</td>
</tr>
<tr>
<td>35</td>
<td>Chebyquad</td>
<td>20</td>
<td>103/323/119</td>
<td>138/432/158</td>
<td>135/422/152</td>
<td>114/361/129</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>372/1200/416</td>
<td>389/1244/429</td>
<td>490/1505/507</td>
<td>374/1169/398</td>
</tr>
<tr>
<td>30</td>
<td>Broyden tridiagonal</td>
<td>50</td>
<td>55/173/63</td>
<td>57/179/65</td>
<td>50/156/56</td>
<td>52/163/59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>51/186/68</td>
<td>63/198/72</td>
<td>66/205/73</td>
<td>56/178/62</td>
</tr>
<tr>
<td>31</td>
<td>Broyden banded</td>
<td>50</td>
<td>31/115/49</td>
<td>35/128/53</td>
<td>19/62/23</td>
<td>18/58/21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>23/74/27</td>
<td>23/74/27</td>
<td>20/62/22</td>
<td>19/59/21</td>
</tr>
<tr>
<td>22</td>
<td>Extended Powell</td>
<td>100</td>
<td>72/209/88</td>
<td>87/254/108</td>
<td>61/171/69</td>
<td>99/287/116</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>85/245/105</td>
<td>83/242/106</td>
<td>73/201/81</td>
<td>115/343/136</td>
</tr>
<tr>
<td>26</td>
<td>Trigonometric</td>
<td>100</td>
<td>58/99/97</td>
<td>62/97/96</td>
<td>60/91/89</td>
<td>63/96/94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>55/95/95</td>
<td>60/105/105</td>
<td>54/83/82</td>
<td>63/93/93</td>
</tr>
<tr>
<td>21</td>
<td>Extended Rosenbrock</td>
<td>1000</td>
<td>28/113/55</td>
<td>28/99/45</td>
<td>29/101/46</td>
<td>31/113/50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1000</td>
<td>28/113/55</td>
<td>29/102/46</td>
<td>29/101/46</td>
<td>31/113/50</td>
</tr>
<tr>
<td>23</td>
<td>Penalty 1</td>
<td>1000</td>
<td>25/96/52</td>
<td>28/95/55</td>
<td>28/114/54</td>
<td>33/98/58</td>
</tr>
</tbody>
</table>

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


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