SIAM J. OPTIM.
Vol. 10, No. 2, pp. 345–358
© 1999 Society for Industrial and Applied Mathematics

CONVERGENCE PROPERTIES OF NONLINEAR CONJUGATE GRADIENT METHODS*

YUHONG DAI†, JIYE HAN‡, GUANGHUI LIU§, DEFENG SUN¶, HONGXIA YIN‡, AND YA-XIANG YUAN†

Abstract. Recently, important contributions on convergence studies of conjugate gradient methods were made by Gilbert and Nocedal [*SIAM J. Optim.*, 2 (1992), pp. 21–42]. They introduce a "sufficient descent condition" to establish global convergence results. Although this condition is not needed in the convergence analyses of Newton and quasi-Newton methods, Gilbert and Nocedal hint that the sufficient descent condition, which was enforced by their two-stage line search algorithm, may be crucial for ensuring the global convergence of conjugate gradient methods. This paper shows that the sufficient descent condition is actually not needed in the convergence analyses of conjugate gradient methods. Consequently, convergence results on the Fletcher–Reeves- and Polak–Ribière-type methods are established in the absence of the sufficient descent condition.

To show the differences between the convergence properties of Fletcher–Reeves- and Polak–Ribière-type methods, two examples are constructed, showing that neither the boundedness of the level set nor the restriction $\beta_k \geq 0$ can be relaxed for the Polak–Ribière-type methods.

Key words. conjugate gradient method, descent condition, global convergence

AMS subject classifications. 65, 49

PII. S1052623494268443

1. Introduction.

We consider the global convergence of conjugate gradient methods for the unconstrained nonlinear optimization problem

$$\min f(x), \tag{1.1}$$

where $f : R^n \to R^1$ is continuously differentiable and its gradient is denoted by g. We consider only the case where the methods are implemented without regular restarts. The iterative formula is given by

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

where λ_k is a step-length and d_k is the search direction defined by

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

where β_k is a scalar and g_k denotes $g(x_k)$.

*Received by the editors May 26, 1994; accepted for publication (in revised form) July 9, 1998; published electronically December 22, 1999. This research was partially supported by Chinese NSF grants 19525101 and 19731010 and by the Australian Research Council.

http://www.siam.org/journals/siopt/10-2/26844.html

†LSEC, Institute of Computational Mathematics, Chinese Academy of Sciences, Beijing 100080, People's Republic of China (dyh@lsec.cc.ac.cn, yyx@lsec.cc.ac.cn).

‡Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing 100080, People's Republic of China (jyhan@amath3.amt.ac.cn, hxyin@public.fhnet.cn.net).

§Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston IL, 60208 (guanghui@pluto.ece.nwu.edu).

¶School of Mathematics, University of New South Wales, Sydney 2052, Australia (sun@maths.unsw.edu.au).

The best-known formulas for β_k are the following Fletcher–Reeves, Polak–Ribière, and Hestenes–Stiefel formulas:

$$\beta_k^{\mathrm{FR}} = \|g_k\|^2/\|g_{k-1}\|^2, \tag{1.4}$$

$$\beta_k^{\mathrm{PR}} = g_k^T(g_k - g_{k-1})/\|g_{k-1}\|^2, \tag{1.5}$$

$$\beta_k^{\mathrm{HS}} = g_k^T(g_k - g_{k-1})/d_{k-1}^T(g_k - g_{k-1}), \tag{1.6}$$

where $\|\cdot\|$ denotes the l_2-norm. The Fletcher–Reeves [4] method with an exact line search was proved to be globally convergent on general functions by Zoutendijk [18]. However, the Polak–Ribière [13] and Hestenes–Stiefel [8] methods with the exact line search are not globally convergent; see the counterexample of Powell [14]. Conjugate gradient methods (1.2)–(1.3) with exact line searches satisfy the equality

$$-g_k^T d_k = \|g_k\|^2, \tag{1.7}$$

which directly implies the *sufficient descent condition*

$$-g_k^T d_k \geq c\|g_k\|^2 \tag{1.8}$$

for some positive constant $c > 0$. This condition has been used often in the literature to analyze the global convergence of conjugate gradient methods with inexact line searches. For instance, Al-Baali [1], Touati-Ahmed and Storey [15], Hu and Storey [9], and Gilbert and Nocedal [5] analyzed the global convergence of algorithms related to the Fletcher–Reeves method with the strong Wolfe line search. Their convergence analyses used the sufficient descent condition, which is implied by the strong Wolfe line search and Fletcher–Reeves-type β_k formulas. For algorithms related to the Polak–Ribière methods, Gilbert and Nocedal [5] investigated wide choices of β_k that resulted in globally convergent methods. In particular, they first gave the global convergence result for the Polak–Ribière-type methods $\beta_k = \max\{0, \beta_k^{\mathrm{PR}}\}$ with inexact line searches. In order for the sufficient descent condition to hold, they modified the strong Wolfe line search to the two-stage line search: the first stage is to find a point using the strong Wolfe line search, and the second stage is, when the sufficient descent condition does not hold, to do more line search iterations until a new point satisfying the sufficient descent condition is found. They hinted that the sufficient descent condition may be crucial for conjugate gradient methods.

It is noted that the sufficient descent condition is not needed in the convergence analyses of Newton and quasi-Newton methods. This motivates us to investigate whether the sufficient descent condition is necessary, as it seemed to be, for the global convergence of conjugate gradient methods. In [11], Liu, Han, and Yin have proved the global convergence properties of the Fletcher–Reeves method under weaker conditions than those of [1]. In [3], Dai and Yuan have proved that the Fletcher–Reeves method using the strong Wolfe line search is globally convergent as long as each search direction is downhill. In the next section, we will provide some basic results for general conjugate gradient methods with a descent condition, instead of the sufficient descent condition. In section 3, we will establish the convergence results for the Fletcher–Reeves- and Polak–Ribière-type methods without assuming the sufficient descent condition. To show the differences between the convergence of Fletcher–Reeves-type methods and Polak–Ribière-type methods, two nonconvergence examples are constructed in section 4 for the Polak–Ribière-type methods, showing that neither the boundedness of the level set nor the restriction $\beta_k \geq 0$ can be relaxed in some sense. A brief discussion is given in the last section.

2. Results for general conjugate gradient methods. Throughout this section, we assume that every search direction d_k satisfies the *descent condition*

$$g_k^T d_k < 0 \tag{2.1}$$

for all $k \geq 1$.

We make the following basic assumptions on the objective function.

ASSUMPTION 2.1. (i) *f is bounded below on the level set $\mathcal{L} = \{x | f(x) \leq f(x_1)\}$, where x_1 is the starting point.* (ii) *In some neighborhood $\mathcal{N}$ of $\mathcal{L}$, f is continuously differentiable, and its gradient is Lipschitz continuous; namely, there exists a constant $L > 0$ such that*

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

The step-length λ_k in (1.2) is computed by carrying out a line search. The *Wolfe line search* [16] consists of finding a positive step-length λ_k such that

$$f(x_k + \lambda_k d_k) \leq f(x_k) + \rho \lambda_k g_k^T d_k, \tag{2.3}$$

$$g(x_k + \lambda_k d_k)^T d_k \geq \sigma g_k^T d_k, \tag{2.4}$$

where $0 < \rho < \sigma < 1$. In order to prove global convergence for the Fletcher–Reeves method, [1], [5] and [9] used the *strong Wolfe line search*, which requires λ_k to satisfy (2.3) and

$$|g(x_k + \lambda_k d_k)^T d_k| \leq -\sigma g_k^T d_k. \tag{2.5}$$

The following important result was obtained by Zoutendijk [18] and Wolfe [16, 17].

LEMMA 2.2. *Suppose that Assumption 2.1 holds. Consider any iteration method of the form (1.2)–(1.3), where d_k satisfies (2.1) and λ_k is obtained by the Wolfe line search. Then*

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty. \tag{2.6}$$

The following theorem is a general and positive result for conjugate gradient methods with the strong Wolfe line search.

THEOREM 2.3. *Suppose that Assumption 2.1 holds. Consider any method of the form (1.2)–(1.3) with d_k satisfying (2.1) and with the strong Wolfe line search (2.3) and (2.5). Then either*

$$\liminf_{k \to \infty} \|g_k\| = 0 \tag{2.7}$$

or

$$\sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty. \tag{2.8}$$

Proof. (1.3) indicates that for all $k \geq 2$,

$$d_k + g_k = \beta_k d_{k-1}. \tag{2.9}$$

Squaring both sides of (2.9), we obtain

$$\|d_k\|^2 = -\|g_k\|^2 - 2g_k^T d_k + \beta_k^2 \|d_{k-1}\|^2. \tag{2.10}$$

It follows from this relation and (2.1) that

$$\|d_k\|^2 \geq \beta_k^2 \|d_{k-1}\|^2 - \|g_k\|^2. \tag{2.11}$$

Definition (1.3) implies the following relation:

$$g_k^T d_k - \beta_k g_k^T d_{k-1} = -\|g_k\|^2, \tag{2.12}$$

which, with the line search condition (2.5), shows that

$$|g_k^T d_k| + \sigma |\beta_k| \, |g_{k-1} d_{k-1}| \geq \|g_k\|^2. \tag{2.13}$$

The above inequality and the Cauchy–Schwartz inequality yield

$$(g_k^T d_k)^2 + \beta_k^2 (g_{k-1}^T d_{k-1})^2 \geq c_1 \|g_k\|^4, \tag{2.14}$$

where $c_1 = (1 + \sigma^2)^{-1}$ is a positive constant. Therefore, it follows from (2.11) and (2.14) that

$$\begin{aligned}
\frac{(g_k^T d_k)^2}{\|d_k\|^2} + \frac{(g_{k-1}^T d_{k-1})^2}{\|d_{k-1}\|^2} &= \frac{1}{\|d_k\|^2} \left[(g_k^T d_k)^2 + \frac{\|d_k\|^2}{\|d_{k-1}\|^2} (g_{k-1}^T d_{k-1})^2 \right] \\
&\geq \frac{1}{\|d_k\|^2} \left[(g_k^T d_k)^2 + \beta_k^2 (g_{k-1}^T d_{k-1})^2 - \frac{(g_{k-1}^T d_{k-1})^2}{\|d_{k-1}\|^2} \|g_k\|^2 \right] \\
&\geq \frac{1}{\|d_k\|^2} \left[c_1 \|g_k\|^4 - \frac{(g_{k-1}^T d_{k-1})^2}{\|d_{k-1}\|^2} \|g_k\|^2 \right]. \qquad (2.15)
\end{aligned}$$

If (2.7) is not true, relations (2.15) and (2.6) imply that the inequality

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} + \frac{(g_{k-1}^T d_{k-1})^2}{\|d_{k-1}\|^2} \geq \frac{c_1}{2} \frac{\|g_k\|^4}{\|d_k\|^2} \tag{2.16}$$

holds for all sufficiently large k. Now inequality (2.8) follows from (2.16) and (2.6). ☐

The following result is a direct corollary of the above theorem.

COROLLARY 2.4. *Suppose that Assumption* 2.1 *holds. Consider any method of the form* (1.2)–(1.3) *with* d_k *satisfying* (2.1) *and with the strong Wolfe line search* (2.3) *and* (2.5). *If*

$$\sum_{k=1}^{\infty} \frac{\|g_k\|^t}{\|d_k\|^2} = +\infty \tag{2.17}$$

for any $t \in [0, 4]$, *the method converges in the sense that* (2.7) *is true.*

Proof. If (2.7) is not true, it follows from Theorem 2.3 that

$$\sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty. \tag{2.18}$$

Because $\|g_k\|$ is bounded away from zero, and $t \in [0, 4]$, it is easy to see that (2.18) contradicts (2.17). This shows that the corollary is true. ∎

If a conjugate gradient method fails to converge, one can easily see from the above corollary that the length of the search direction will converge to infinity. Results similar to Corollary 2.4 can also be established using the Zoutendijk condition and the sufficient descent condition (1.8). It should be noted that we have not assumed the sufficient descent condition. Hence our results are powerful tools for our analyses in the next section, where we will concentrate on proving the global convergence of some conjugate gradient methods without assuming the sufficient descent condition (1.8). Another point worth mentioning is that we do not assume the boundedness of the level set.

3. Global convergence. In this section, we establish some global convergence results for the Fletcher–Reeves- and Polak–Ribière-type methods. The general outline of the proofs is that, assuming that the convergence relation (2.7) does not hold, we can derive that $\sum_{k=1}^{\infty} \frac{\|g_k\|^2}{\|d_k\|^2} = +\infty$ or $\sum_{k=1}^{\infty} \frac{1}{\|d_k\|^2} = +\infty$, which with Corollary 2.4 in turn implies that (2.7) holds, giving a contradiction.

First, we consider the Fletcher–Reeves-type methods of the form (1.2)–(1.3), where β_k is any scalar satisfying

$$\sigma|\beta_k| \leq \bar{\sigma}\beta_k^{\mathrm{FR}} \tag{3.1}$$

for all $k \geq 2$, where σ is the parameter defined in (2.4) and $\bar{\sigma} \in (0, 1/2]$ is a constant. In order to prove its global convergence, Hu and Storey [9] had to restrict the parameter $\bar{\sigma}$ to be strictly less than 1/2 to derive the sufficient descent condition. The following result shows that such a restriction can be relaxed while preserving the global convergence.

THEOREM 3.1. *Suppose that Assumption* 2.1 *holds. Consider any method of the form* (1.2)–(1.3) *with the strong Wolfe line search* (2.3) *and* (2.5)*, where* β_k *satisfies* (3.1) *with* $\bar{\sigma} \in (0, 1/2]$*, and*

$$\|g_k\|^2 \sum_{j=2}^{k} \prod_{i=j}^{k} \left(\frac{\beta_i}{\beta_i^{\mathrm{FR}}}\right)^2 \leq c_2 k \tag{3.2}$$

for some constant $c_2 > 0$*. Then*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{3.3}$$

Proof. From (1.3), (1.4), (2.5), and (3.1), we deduce that

$$\begin{aligned}
\frac{-g_k^T d_k}{\|g_k\|^2} &= 1 - \beta_k \frac{-g_k^T d_{k-1}}{\|g_k\|^2} = 1 - \left(\frac{\beta_k}{\beta_k^{\mathrm{FR}}}\right) \frac{-g_k^T d_{k-1}}{\|g_{k-1}\|^2} \\
&\leq 1 + \left|\frac{\beta_k}{\beta_k^{\mathrm{FR}}}\right| \frac{-\sigma g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} \\
&\leq 1 + \bar{\sigma}\left(\frac{-g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2}\right) \\
&\leq \cdots
\end{aligned}$$

$$\leq \sum_{j=0}^{k-2} \bar{\sigma}^j + \bar{\sigma}^{k-1} \left(\frac{-g_1^T d_1}{\|g_1\|^2} \right) = \frac{1-\bar{\sigma}^k}{1-\bar{\sigma}} < \frac{1}{1-\bar{\sigma}}. \tag{3.4}$$

Similarly, we have that

$$\frac{-g_k^T d_k}{\|g_k\|^2} \geq 1 - \bar{\sigma} \frac{1-\bar{\sigma}^{k-1}}{1-\bar{\sigma}} > 0 \tag{3.5}$$

because $\bar{\sigma} \leq 1/2$. Thus, d_k is a descent direction.

On the other hand, it follows from (2.10) that

$$\|d_k\|^2 \leq -2g_k^T d_k + \beta_k^2 \|d_{k-1}\|^2. \tag{3.6}$$

Using (3.6) recursively and observing that $d_1 = -g_1$, we get that

$$\|d_k\|^2 \leq -2g_k^T d_k - 2 \sum_{j=2}^{k} \prod_{i=j}^{k} \beta_i^2 g_{j-1}^T d_{j-1}$$

$$= -2g_k^T d_k - 2\|g_k\|^4 \sum_{j=2}^{k} \prod_{i=j}^{k} \left(\frac{\beta_i}{\beta_i^{\mathrm{FR}}} \right)^2 \left(\frac{g_{j-1}^T d_{j-1}}{\|g_{j-1}\|^4} \right). \tag{3.7}$$

If the theorem is not true, (3.2) holds and there exists a positive constant γ such that

$$\|g_k\| \geq \gamma \quad \text{for all } k. \tag{3.8}$$

Thus, it follows from the above inequality, (3.4), and (3.7) that

$$\frac{\|d_k\|^2}{\|g_k\|^2} \leq \frac{2}{1-\bar{\sigma}} \left[1 + \frac{\|g_k\|^2}{\gamma^2} \sum_{j=2}^{k} \prod_{i=j}^{k} \left(\frac{\beta_i}{\beta_i^{\mathrm{FR}}} \right)^2 \right]. \tag{3.9}$$

The above relation and (3.2) imply that

$$\sum_{k=1}^{\infty} \frac{\|g_k\|^2}{\|d_k\|^2} = +\infty. \tag{3.10}$$

This, with Corollary 2.4, implies that $\liminf_k \|g_k\| = 0$. This completes our proof. □

The above theorem extends Hu and Storey's [9] result to the case when $\bar{\sigma} = 1/2$. If $\bar{\sigma} \in (0, 1/2)$, we can see from (3.5) that the sufficient descent condition (1.8) holds. If $\bar{\sigma} = 1/2$, however, we only have that

$$\frac{-g_k^T d_k}{\|g_k\|^2} \geq \frac{1}{2^k}, \tag{3.11}$$

which does not imply the sufficient descent condition.

Now we consider methods that are related to the Polak–Ribière and Hestenes–Stiefel algorithms. We need the following assumption.

Assumption 3.2. *The level set $\mathcal{L} = \{x | f(x) \leq f(x_1)\}$ is bounded.*

Under Assumptions 2.1 and 3.2, there exists a positive constant $\bar{\gamma}$ such that

$$\|g(x)\| \leq \bar{\gamma} \quad \text{for all } x \in \mathcal{L}. \tag{3.12}$$

Denote $s_{k-1} = x_k - x_{k-1}$ and $u_k = d_k/\|d_k\|$. In [5], Gilbert and Nocedal introduced the following property.

PROPERTY $(*)$. *Consider a method of the form* (1.2)–(1.3)*, and suppose that* (3.12) *and* (3.8) *hold. Then we say that the method has Property* $(*)$ *if there exist constants* $b > 1$ *and* $\lambda > 0$ *such that for all* k*,*

$$|\beta_k| \leq b \tag{3.13}$$

and

$$\|s_{k-1}\| \leq \lambda \Longrightarrow |\beta_k| \leq \frac{1}{2b}. \tag{3.14}$$

Let N^* denote the set of positive integers. For $\lambda > 0$ and positive integer Δ, denote

$$\mathcal{K}_{k,\Delta}^{\lambda} := \{i \in N^* : k \leq i \leq k + \Delta - 1, \|s_{i-1}\| > \lambda\}.$$

Let $|\mathcal{K}_{k,\Delta}^{\lambda}|$ denote the number of elements of $\mathcal{K}_{k,\Delta}^{\lambda}$ and let $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ denote, respectively, the floor and ceiling operators. The following lemmas are drawn from [5].

LEMMA 3.3 (see [5]). *Suppose that Assumptions* 2.1 *and* 3.2 *hold. Consider any method of the form* (1.2)–(1.3) *with a descent direction* d_k*. If, at the* k*th step,* $\beta_k \geq 0$*, then* $d_k \neq 0$ *and*

$$\|u_k - u_{k-1}\| \leq 2\frac{\|g_k\|}{\|d_k\|}. \tag{3.15}$$

LEMMA 3.4 (see [5]). *Suppose that Assumptions* 2.1 *and* 3.2 *hold. Consider the method of* (1.2)–(1.3) *with any line search satisfying* (2.1)*. Assume that the method has Property* $(*)$ *and that*

$$\sum_{k=1}^{\infty} \frac{1}{\|d_k\|^2} < +\infty. \tag{3.16}$$

Assume also that (3.8) *holds. Then there exists* $\lambda > 0$ *such that, for any* $\Delta \in N^*$ *and any index* k_0*, there is a greater index* $k > k_0$ *such that*

$$|\mathcal{K}_{k,\Delta}^{\lambda}| > \frac{\Delta}{2}.$$

The conditions used in Lemma 3.4 are not the same as those used in [5]. In particular, the sufficient descent condition (1.8) used in [5] is here replaced by the descent condition (2.1). Under this weaker condition, we can also establish a similar global convergence result as that in [5].

The next theorem is a global convergence result of conjugate gradient methods with Property $(*)$. It is applicable, for example, to the Polak–Ribière-type method

$$\beta_k = \max\{0, \beta_k^{\mathrm{PR}}\}. \tag{3.17}$$

The proof of the theorem is similar to that in [5].

THEOREM 3.5. *Suppose that Assumptions* 2.1 *and* 3.2 *hold. Consider the method* (1.2)–(1.3) *with the following three properties:* (i) $\beta_k \geq 0$*;* (ii) *the strong Wolfe line search conditions* (2.3) *and* (2.5) *and the descent condition* (2.1) *hold for all* k*;* (iii) *Property* $(*)$ *holds. Then the method converges in the sense that* (3.3) *holds.*

Proof. We proceed by contradiction, assuming that the theorem is not true. Then there exists a positive constant γ such that (3.8) holds. Since $\beta_k \geq 0$ and d_k is a descent direction, it follows from Lemma 3.3 that

$$\|u_k - u_{k-1}\| \leq 2\frac{\|g_k\|}{\|d_k\|} \tag{3.18}$$

for all $k \geq 2$. The above inequality, (3.8), and Theorem 2.3 imply that

$$\sum_{k=1}^{\infty} \|u_k - u_{k-1}\|^2 \leq \frac{4}{\gamma^2} \sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty. \tag{3.19}$$

For any two indices l, k, with $l \geq k$, we can write

$$\begin{aligned} x_l - x_{k-1} &= \sum_{i=k}^{l} \|s_{i-1}\| u_{i-1} \\ &= \sum_{i=k}^{l} \|s_{i-1}\| u_{k-1} + \sum_{i=k}^{l} \|s_{i-1}\| (u_{i-1} - u_{k-1}). \end{aligned}$$

This relation and the fact that $\|u_{k-1}\| = 1$ give

$$\sum_{i=k}^{l} \|s_{i-1}\| \leq \|x_1 - x_{k-1}\| + \sum_{i=k}^{l} \|s_{i-1}\| \, \|u_{i-1} - u_{k-1}\|. \tag{3.20}$$

Since f_k decreases with k, we have that $\{x_k\} \subset \mathcal{L}$, which together with Assumption 3.2 implies that there exists a positive constant B such that $\|x_k\| \leq B$ for all $k \geq 1$. Hence

$$\sum_{i=k}^{l} \|s_{i-1}\| \leq 2B + \sum_{i=k}^{l} \|s_{i-1}\| \, \|u_{i-1} - u_{k-1}\|. \tag{3.21}$$

By Corollary 2.4, we can assume that (3.16) holds. Thus the conditions of Lemma 3.4 are satisfied. Let $\lambda > 0$ be given by Lemma 3.4 and define $\Delta := \lceil 8B/\lambda \rceil$. By (3.19), we can find an index $k_0 \geq 1$ such that

$$\sum_{i \geq k_0} \|u_i - u_{i-1}\|^2 \leq \frac{1}{4\Delta}. \tag{3.22}$$

With this Δ and k_0, Lemma 3.4 gives an index $k \geq k_0$ such that

$$|\mathcal{K}_{k,\Delta}^{\lambda}| > \frac{\Delta}{2}. \tag{3.23}$$

Next, for any index $i \in [k, k + \Delta - 1]$, by the Cauchy–Schwartz inequality and (3.22),

$$\begin{aligned} \|u_i - u_{k-1}\| &\leq \sum_{j=k}^{i} \|u_j - u_{j-1}\| \\ &\leq (i - k + 1)^{1/2} \left(\sum_{j=k}^{i} \|u_j - u_{j-1}\|^2 \right)^{1/2} \\ &\leq \Delta^{1/2} \left(\frac{1}{4\Delta} \right)^{1/2} = \frac{1}{2}. \qquad (3.24) \end{aligned}$$

Using this relation and (3.23) in (3.21), with $l = k + \Delta - 1$, we get that

$$2B \geq \frac{1}{2} \sum_{i=k}^{k+\Delta-1} \|s_{i-1}\| > \frac{\lambda}{2} |\mathcal{K}_{k,\Delta}^{\lambda}| > \frac{\lambda \Delta}{4}. \tag{3.25}$$

Thus $\Delta < 8B/\lambda$, which contradicts the definition of Δ. Therefore, the theorem is true. □

4. Nonconvergence examples. In the previous section, we have proved two convergence theorems, namely, Theorems 3.1 and 3.5, for the Fletcher–Reeves- and Polak–Ribière-type methods. Neither of the theorems needs the line search to satisfy the sufficient descent condition (1.8). In this section, we will present two nonconvergence examples for the Polak–Ribière methods.

It can be seen from Theorem 3.1 that the boundedness of the level set is not required in analyzing the Fletcher–Reeves-type methods. Therefore, the convergence results for the Fletcher–Reeves-type methods also apply to noncoercive objective function. In contrast, we are able to construct an example, as included in the following theorem, to show that the boundedness of the level set is necessary for the convergence of Polak–Ribière methods even if line searches are exact. It is easy to see that the theorem is also true for the Polak–Ribière-type method (3.17).

THEOREM 4.1. *Consider the Polak–Ribière method* (1.2), (1.3), *and* (1.5) *with* λ_k *chosen to be any local minimizer of* $\Phi_k(\lambda) = f(x_k + \lambda d_k)$, $\lambda > 0$. *Then there exists a starting point* x_1 *and a function* $f(x)$ *satisfying Assumption* 2.1 *such that the iterations generated by the method satisfy, for all* $k \geq 1$,

$$\beta_{k+1}^{\mathrm{PR}} \geq 0 \tag{4.1}$$

and

$$\|g_k\| = 1. \tag{4.2}$$

Proof. We define

$$\theta_k = \begin{cases} -\dfrac{\pi}{2} & \text{for } k = 0, \\ 0 & \text{for } k = 1, \\ \dfrac{1}{6}\left[1 - \left(-\dfrac{1}{2}\right)^{k-1}\right]\pi & \text{for } k \geq 2 \end{cases} \tag{4.3}$$

and consider the gradients and the search directions given by

$$g_k = (-1)^k \begin{pmatrix} \sin\theta_{k-1} \\ -\cos\theta_{k-1} \end{pmatrix} \tag{4.4}$$

and

$$d_k = \csc\frac{\pi}{2^k} \begin{pmatrix} \cos\theta_k \\ \sin\theta_k \end{pmatrix}, \tag{4.5}$$

where

$$\csc\frac{\pi}{2^k} = \frac{1}{\sin\frac{\pi}{2^k}}.$$

It follows that (4.2) holds for all $k \geq 1$. In addition, (4.4) and (4.5) clearly satisfy the equality

$$g_{k+1}^T d_k = 0. \tag{4.6}$$

Because

$$|\theta_k - \theta_{k-1}| = \frac{\pi}{2^k} \tag{4.7}$$

holds for all $k \geq 1$, it follows from (1.5), (4.2), and (4.4) that

$$\beta_{k+1}^{\mathrm{PR}} = 1 - g_{k+1}^T g_k = 1 + \cos(\theta_k - \theta_{k-1}) = 1 + \cos\frac{\pi}{2^k} = 2\cos^2\frac{\pi}{2^{k+1}}. \tag{4.8}$$

Thus (4.1) also holds for all $k \geq 1$. Further, direct calculations show that

$$\begin{aligned}
-g_{k+1} + \beta_{k+1}^{\mathrm{PR}} d_k &= (-1)^{k+1}\begin{pmatrix} -\sin\theta_k \\ \cos\theta_k \end{pmatrix} + 2\cos^2\frac{\pi}{2^{k+1}}\csc\frac{\pi}{2^k}\begin{pmatrix} \cos\theta_k \\ \sin\theta_k \end{pmatrix} \\
&= \csc\frac{\pi}{2^{k+1}}\left[\sin\frac{\pi}{(-2)^{k+1}}\begin{pmatrix} -\sin\theta_k \\ \cos\theta_k \end{pmatrix} + \cos\frac{\pi}{2^{k+1}}\begin{pmatrix} \cos\theta_k \\ \sin\theta_k \end{pmatrix}\right] \\
&= \csc\frac{\pi}{2^{k+1}}\begin{pmatrix} \cos\left(\theta_k + (-1)^{k+1}\frac{\pi}{2^{k+1}}\right) \\ \sin\left(\theta_k + (-1)^{k+1}\frac{\pi}{2^{k+1}}\right) \end{pmatrix} \\
&= \csc\frac{\pi}{2^{k+1}}\begin{pmatrix} \cos\theta_{k+1} \\ \sin\theta_{k+1} \end{pmatrix} = d_{k+1}. \qquad (4.9)
\end{aligned}$$

This together with $d_1 = -g_1$ imply that if the gradients are given by (4.4), then the Polak–Ribière method will produce the search directions as in (4.5).

Now, we let $\lambda_k = 1/\|d_k\|$ and define

$$x_k = \sum_{i=0}^{k-1}\begin{pmatrix} \cos\theta_i \\ \sin\theta_i \end{pmatrix} \tag{4.10}$$

and

$$f_k = -\sum_{i=0}^{k-1}\sin\frac{\pi}{2^i}. \tag{4.11}$$

Then (1.2) holds and since $\|d_k\| = csc\frac{\pi}{2^k}$ and $g_k^T d_k = -1$, (2.3) and (2.5) hold. Because

$$\lim_{k\to\infty}\theta_k = \frac{\pi}{6} \tag{4.12}$$

and

$$\|x_{k+1} - x_k\| = 1, \tag{4.13}$$

we can see that $\{x_k\}$ has no cluster points and hence that it is easy to construct a function f satisfying Assumption 2.1 such that for all $k \geq 1$,

$$f(x_k) = f_k, \qquad g(x_k) = g_k, \tag{4.14}$$

and λ_k is a local minimizer of $\Phi_k(\lambda)$. Therefore, for the starting point $x_1 = (0, -1)^T$ and the function f, the iterations generated by the Polak–Ribière method satisfy (4.1) and (4.2) for all $k \geq 1$. ▯

As opposed to Theorem 3.1, Theorem 3.5 does not allow any negative values of β_k. However, as pointed out in Gilbert and Nocedal [5], the Polak–Ribière method can produce negative values of β_k^{PR} even for strong convex objective functions. Therefore, it is interesting to investigate in what range the restriction $\beta_k \geq 0$ in Theorem 3.5 can be relaxed. After further studies of the $n = 2$, $m = 8$ example of Powell [14], we obtain the following result.

THEOREM 4.2. *For any given positive constant ε, consider the method* (1.2)–(1.3) *with*

$$\beta_k = \max\{\beta_k^{\mathrm{PR}}, -\varepsilon\} \tag{4.15}$$

and with λ_k chosen to be any local minimizer of $\Phi_k(\lambda) = f(x_k + \lambda d_k)$, $\lambda > 0$. There exists a starting point x_1 and a function $f(x)$ satisfying Assumptions 2.1 *and* 3.2 *such that the sequence of the gradient norms $\{\|g_k\|\}$ generated by the method is bounded away from zero.*

Proof. For any positive constant $\phi \in (0, 1)$, let the steps of the method have the form

$$s_{8j+i} = a_i \begin{pmatrix} 1 \\ b_i \phi^{2j} \end{pmatrix}, \quad s_{8j+4+i} = a_i \begin{pmatrix} -1 \\ b_i \phi^{2j+1} \end{pmatrix}, \quad j \geq 0, \quad i = 1, 2, 3, 4, \tag{4.16}$$

where the numbers $\{a_i; i = 1, 2, 3, 4\}$ are all positive, and consider the values

$$b_1 = -2, \quad b_2 = \frac{6 - 2\phi - 2\phi^2}{2 + 5\phi}, \quad b_3 = -\phi, \quad b_4 = -2.$$

To satisfy the line search condition

$$g_{k+1}^T d_k = 0, \tag{4.17}$$

we assume that the gradients have the form

$$\begin{aligned} g_{8j+1} &= c_1 \begin{pmatrix} b_4 \phi^{2j-1} \\ 1 \end{pmatrix}, \quad g_{8j+i} = c_i \begin{pmatrix} -b_{i-1} \phi^{2j} \\ 1 \end{pmatrix}, \quad i = 2, 3, 4; \\ g_{8j+5} &= c_1 \begin{pmatrix} -b_4 \phi^{2j+1} \\ 1 \end{pmatrix}, \quad g_{8j+4+i} = c_i \begin{pmatrix} b_{i-1} \phi^{2j+1} \\ 1 \end{pmatrix}, \quad i = 2, 3, 4, \end{aligned} \tag{4.18}$$

where $\{c_i; i = 1, 2, 3, 4\}$ are constants. To ensure the conjugacy condition

$$s_k^T (g_{k+1} - g_k) = 0 \tag{4.19}$$

for all $k \geq 1$, we choose each c_i as follows:

$$c_1 = 3\phi(1-\phi)(5-\phi), \quad c_2 = -3(1+\phi)(2+\phi^2),$$
$$c_3 = (1+\phi)(2-\phi)(2+5\phi), \quad c_4 = 2(5-\phi)(1-\phi^2). \tag{4.20}$$

Because $n = 2$, relations (4.17) and (4.19) ensure that each d_k is produced by the Polak–Ribière method. In addition, direct calculations show that $g_k^T s_k < 0$ holds for all $k \geq 1$; namely, each d_k is a descent direction.

Due to symmetry, we can reduce the objective function at every iteration if the following relations hold:

$$f(x_{8j+1}) > f(x_{8j+2}) > f(x_{8j+3}) > f(x_{8j+4}) > f(x_{8j+5}). \tag{4.21}$$

Now, when the first component of x is equal to the first component of x_k, where k is any positive integer, then the values in (4.18) allow the second component of $g(x)$ to be constant, provided that the first components of the points $\{x_{8j+i}; i = 1, 2, \ldots, 8\}$ are all different. Thus, the equation

$$f(x_k) - f^* = (x_k)_2 (g_k)_2 \tag{4.22}$$

is satisfied, where f^* is the limit of f_k. Given the limit point $\hat{x}_1 = \lim_{j\to\infty} x_{8j+1}$, we can compute x_{8j+1} in the following way:

$$x_{8j+1} = \hat{x}_1 - \sum_{k=j}^{\infty} \sum_{i=1}^{8} s_{8k+i} = \begin{pmatrix} 0 \\ h\phi^{2j}/(\phi - 1) \end{pmatrix} \tag{4.23}$$

and

$$x_{8j+i+1} = x_{8j+i} + s_{8j+i}, \qquad i = 1, 2, \ldots, 7, \tag{4.24}$$

where $h = a_1 b_1 + a_2 b_2 + a_3 b_3 + a_4 b_4$. It follows that expression (4.21) is equivalent to the inequalities

$$\begin{aligned} -c_1(a_1b_1 + a_2b_2 + a_3b_3 + a_4b_4) &> -c_2(a_1b_1\phi + a_2b_2 + a_3b_3 + a_4b_4) \\ &> -c_3(a_1b_1\phi + a_2b_2\phi + a_3b_3 + a_4b_4) \\ &> -c_4(a_1b_1\phi + a_2b_2\phi + a_3b_3\phi + a_4b_4) \\ &> -c_1\phi(a_1b_1 + a_2b_2 + a_3b_3 + a_4b_4). \end{aligned} \tag{4.25}$$

These inequalities are consistent because, if

$$a_1 = 10, \quad a_2 = 35\phi, \quad a_3 = 38, \quad a_4 = \phi, \tag{4.26}$$

and if ϕ is small, then the dominant terms of the five lines of (4.25) are 300ϕ, 270ϕ, 240ϕ, 220ϕ, and 300ϕ, respectively. Now, as in Powell [14], we can construct a function satisfying Assumptions 2.1 and 3.2 such that the gradient conditions (4.18) hold.

By direct estimations, we can obtain that the dominant terms of $\{\beta_{4j+1}^{\mathrm{PR}}; i = 1, 2, 3, 4\}$ are

$$-\frac{3}{2}\phi, \quad \frac{4}{25\phi^2}, \quad \frac{10}{9}, \quad \frac{9}{4},$$

respectively, when ϕ is small and j is large. Therefore, for any positive number $\varepsilon > 0$, we have that $\beta_k^{\mathrm{PR}} \geq -\varepsilon$ for all large j, provided that $\phi \in (0, 1)$ is sufficiently small. This completes our proof. $\square$

In [2], the above theorem is proved by using a three-dimensional example, in which line searches choose the first local minimum in every iteration.

5. Discussions. In this paper we have presented some global convergence results for nonlinear conjugate gradient methods, where the step-length is computed by the strong Wolfe conditions under the assumption that all the search directions are descent directions. The sufficient descent condition (1.8) has not been used in our convergence proofs and we have established convergence results for Fletcher–Reeves- and Polak–Ribière-type methods.

We have also provided two examples for which Polak–Ribière-type methods fail to converge. From these examples, we can see that the Fletcher–Reeves-type methods have better convergence properties than the Polak–Ribière-type methods, even though the latter perform better in practice. We believe that the results given in this paper will lead to a deeper understanding of the behavior of nonlinear conjugate gradient methods with inexact line searches.

This paper is a combination of two research reports, [6] and [2]; readers can find a more extensive discussion on the subject of this paper in those reports. See also [7], [10], and [11]. Some recent advances can be found in [7] and [10].

Acknowledgments. The authors are very grateful to Professor Jorge Nocedal for his suggestions in revising this paper. The third author is much indebted to Professor Jorge Nocedal for his grant support and supervision. The authors thank the referees for their extensive comments, which improved the paper.

REFERENCES

[1] M. Al-Baali, *Descent property and global convergence of the Fletcher-Reeves method with inexact linesearch*, IMA J. Numer. Anal., 5 (1985), pp. 121–124.

[2] Y. H. Dai and Y. Yuan, *Further Studies on the PRP Method*, Research Report ICM-95-040, Inst. of Comp. Math. and Sci. Eng. Computing, Chinese Academy of Sciences, Beijing, 1995.

[3] Y. H. Dai and Y. Yuan, *Convergence properties of the Fletcher-Reeves method*, IMA J. Numer. Anal., 16 (1996), pp. 155–164.

[4] R. Fletcher and C. M. Reeves, *Function minimization by conjugate gradients*, Comput. J., 7 (1964), pp. 149–154.

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

[6] J. Y. Han, G. H. Liu, D. F. Sun, and H. X. Yin, *On the Global Convergence of Nonlinear Conjugate Gradient Methods*, Technical Report 94-011, Inst. of Applied Math., Chinese Academy of Sciences, Beijing, 1994; Acta Math. Appl. Sinica, to appear.

[7] J. Y. Han, G. H. Liu, and H. X. Yin, *Convergence properties of conjugate gradient methods with strong Wolfe linesearch*, Systems Sci. Math. Sci., 11 (1998), pp. 112–116.

[8] M. R. Hestenes and E. Stiefel, *Method of conjugate gradient for solving linear system*, J. Res. Nat. Bur. Standards, 49 (1952), pp. 409–436.

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

[10] G. H. Liu, J. Y. Han, H. D. Qi, and Z. L. Xu, *Convergence analysis on a class of conjugate gradient methods*, Acta Math. Scientia, 18 (1998), pp. 11–16.

[11] G. H. Liu, J. Y. Han, and H. X. Yin, *Global convergence of the Fletcher-Reeves algorithm with inexact linesearch*, Appl. Math. J. Chinese Univ. Ser. B, 10 (1995), pp. 75–82.

[12] J. J. Moré and D. J. Thuente, *On line search algorithms with guaranteed decrease*, ACM Trans. Math. Software, 20 (1994), pp. 286–307.

[13] E. Polak and G. Ribière, *Note sur la convergence de méthodes de directions conjugées*, Revue Française d'Informatique et de Recherche Opérationnelle, 16 (1969), pp. 35–43.

[14] M. J. D. Powell, *Nonconvex minimization calculations and the conjugate gradient method*, in Numerical Analysis, Lecture Notes in Mathematics 1066, D. F. Griffiths, ed., Springer-Verlag, Berlin, 1984, pp. 122–141.

[15] D. Touati-Ahmed and C. Storey, *Efficient hybrid conjugate gradient techniques*, J. Optim. Theory Appl., 64 (1990), pp. 379–397.
[16] P. Wolfe, *Convergence conditions for ascent methods*, SIAM Rev., 11 (1969), pp. 226–235.
[17] P. Wolfe, *Convergence conditions for ascent methods* II*: Some corrections*, SIAM Rev., 13 (1971), pp. 185–188.
[18] G. Zoutendijk, *Nonlinear programming, computational methods*, in Integer and Nonlinear Programming, J. Abadie, ed., North-Holland, Amsterdam, 1970, pp. 37–86.