Convergence Properties of Beale-Powell Restart Algorithm *

Y. H. Dai and Y. Yuan
(State Key Laboratory of Scientific and Engineering Computing, ICMSEC, Chinese Academy of Sciences, Beijing 100080, China)

Abstract The Beale-Powell restart algorithm is highly useful for large-scale unconstrained optimization. In this paper, we use an example to show that the algorithm may fail to converge. The global convergence of a slightly modified algorithm is proved.

Keywords: unconstrained optimization, conjugate gradient, restart, line search, global convergence.

The standard conjugate gradient method for solving the unconstrained optimization problem

$$\min f(x), \quad x \in R^n \tag{0.1}$$

has the following form

$$x_{k+1} = x_k + \alpha_k d_k \tag{0.2}$$

$$d_k = \begin{cases} -g_k, & \text{for } k = 1; \\ -g_k + \beta_k d_{k-1}, & \text{for } k \geq 2, \end{cases} \tag{0.3}$$

where $g_k = \nabla f(x_k)$, α_k is a steplength obtained by a one-dimensional line search, and β_k is a scalar. If β_k is chosen to be

$$\beta_k^{FR} = \|g_k\|^2 / \|g_{k-1}\|^2, \tag{0.4}$$

where and below $||\cdot||$ stands for the Euclidean norm, the corresponding method is called as the Fletcher-Reeves ([1]) method, abbreviated to the FR method. If β_k equals

$$\beta_k^{PRP} = g_k^T y_{k-1} / \|g_{k-1}\|^2, \tag{0.5}$$

*Partially supported by Chinese NSF grants 19525101,19731001.

where $y_{k-1} = g_k - g_{k-1}$, the corresponding method is the Polak-Ribière-Polyak method ([2, 3]), abbreviated to the PRP method. The FR, and PRP methods are two of the well-known nonlinear conjugate gradient methods. Some recent results on them can be seen in references [4] and [5].

It is proved in [6] that the standard conjugate gradient method without restart is at most linearly convergent [1]. Practical conjugate gradient algorithms therefore include a periodic restart strategy, namely, set

$$d_k = -g_k, \quad \text{for } k = in + 1,\ i = 1, 2, \ldots \tag{0.6}$$

If the line search is asymptotically exact, [8, 9] proved that this restart strategy leads to the $n-$step quadratic convergence rate of the algorithms. However, as pointed out by Powell [10], this restart strategy has the following drawbacks: (a) the frequency of restart that should depend on the objective function can not be simply set to n; (b) a restart along $-g_k$ abandons the second derivative information that is found by the search along d_{k-1}; and (c) the immediate reduction in the objective function is usually less than it would be without the restart. Therefore it seems more reasonable to use $-g_k + \beta_k d_{k-1}$ as the restart direction.

Beale [11] studied such a restart strategy, which uses $-g_k + \beta_k d_{k-1}$ as the restart direction and extends the non-restart direction from two terms to three terms (see Step 3 in Algorithm 1) such that all search directions are conjugate to one another if f is convex quadratic and the line search is exact. McGuire and Wolfe[2] tried this algorithm, but disappointed numerical results were reported. By introducing a new restart criterion, namely, (1.1), Powell [10] overcame the difficulties that McGuire and Wolfe encountered and obtained satisfactory numerical results. The current general subroutine of the algorithm is VE04 in Harwell subroutine library. In this paper, we call it as the Beale-Powell restart algorithm.

From Algorithm 1 we see that the Beale-Powell restart algorithm only needs to store six n-dimensional vectors. As a result, the Beale-Powell restart algorithm is still available for solving (0.1) even if its dimension n is very large. One large-scale practical problem that uses the Beale-Powell restart algorithm to minimize can be seen in [13].

Despite its good numerical performances and adaptability for large-scale unconstrained optimization, it is not clear yet whether or not the Beale-

[1] [7] showed that the convergence rate of the conjugate gradient method is exactly linear for uniformly convex quadratics.

[2] see McGuire M F, Wolfe P. Evaluating a restart procedure for conjugate gradients. Report RC-4382, IBM Research Center, Yorktown Heights, 1973

Powell restart algorithm converges theoretically. Ref. [12] considered general three-term conjugate gradient method in which d_k has the form

$$d_k = -g_k + \beta_k d_{k-1} + \gamma_k d_{t(p)}, \quad t(p) < k \le t(p+1) \tag{0.7}$$

where $t(p)$ stands for the p-th restart iteration, and established convergence results for the general method under quite a few restrictions on the scalars β_k and γ_k. In this paper, one of the examples in [14] is used to show that the Beale-Powell restart algorithm may fail to converge. The global convergence of a slightly modified algorithm is proved.

1. Beale-Powell restart algorithm

The Beale-Powell restart algorithm is described as follows.

Algorithm 1.

Step **1** Given $x_1 \in R^n$; c_1,$c_2 \in (0,1)$, $c_3 \in (1,\infty)$, $\varepsilon \in [0,1)$;
set $d_1 = -g_1$; $k = t = 1$;

Step **2** If $k = 1$, go to Step 5;
If $k - t \ge n$, set $t = k-1$; otherwise if $k \ge 2$ and

$$|g_{k-1}^T g_k| > c_1 \|g_k\|^2, \tag{1.1}$$

also set $t = k-1$;

Step **3** If $k > t+1$, compute d_k as follows and go to Step 4:

$$d_k = -g_k + \beta_k d_{k-1} + \gamma_k d_t, \tag{1.2}$$

where β_k and γ_k is defined by

$$\begin{aligned} \beta_k &= g_k^T y_{k-1} / d_{k-1}^T y_{k-1}, & (1.3)\\ \gamma_k &= g_k^T y_{t-1} / d_{t-1}^T y_{t-1}; & (1.4) \end{aligned}$$

if $k = t+1$, compute d_k by setting $\gamma_k = 0$ in (1.2) and go to Step 5;

Step **4** If the following relation is not satisfied

$$-c_3 \|g_k\|^2 \le d_k^T g_k \le -c_2 \|g_k\|^2, \tag{1.5}$$

then set $t = k-1$, go to Step 3;

Step **5** Carry out a line search along d_k, getting α_k;
set $x_{k+1} = x_k + \alpha_k d_k$;

Step **6** If $\|g_{k+1}\| \le \epsilon$, stop; otherwise set $k = k + 1$ and go to Step 2.

In the above algorithm, we still denote $y_{k-1} = g_k - g_{k-1}$. Ref. [10] suggested that $\{c_i; i = 1, 2, 3\}$ can take the values $c_1 = 0.2$, $c_2 = 0.8$, and $c_3 = 1.2$ respectively, and the line search in Step 5 satisfies the strong Wolfe conditions, namely,

$$f(x_k) - f(x_k + \alpha_k d_k) \ge -\delta \alpha_k g_k^T d_k, \tag{1.6}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \le -\sigma g_k^T d_k, \tag{1.7}$$

where $0 < \delta < \sigma < 1$.

2. Non-convergence example and a modified algorithm

In this section, we first show that Algorithm 1 may fail to converge. Consider the $n = 2, m = 8$ example for the PRP method in [14]. The example shows that if the steplength can be chosen to be any local minimizer of the function

$$\Phi_k(\alpha) = f(x_k) + \alpha d_k, \qquad \alpha > 0 \tag{2.1}$$

then the PRP method can cycle near eight points without approaching a solution point. Since this example is such that $g_{k+1}^T d_k = 0$, we have that

$$d_k^T g_k = -\|g_k\|^2. \tag{2.2}$$

Besides it, direct calculations show that, when $j \to \infty$, $\{|g_{4j+i}^T g_{4j+i-1}|/\|g_{4j+i}\|^2; i = 1, 2, 3, 4\}$ tend to

$$1, \qquad 2, \qquad 1/3, \qquad 3/2$$

respectively, which implies that (1.1) is satisfied for all large k provided that the parameter in (1.1) is such that $c_1 < 1/3$. As a result, if Algorithm 1 is used to minimize the function, a restart will be done at every iteration and hence by (2.2),

$$\beta_k = \beta_k^{PRP}, \qquad \gamma_k = 0 \tag{2.3}$$

holds for all large k. Therefore Algorithm 1 produces the same iterates as the PRP method does, which means that Algorithm 1 may fail to converge.

In the above example, due that $\|d_k\|$ tends to infinity with k, we can see that if $\beta_k^{PRP} < 0$ for some k, then the two consecutive directions d_{k-1}

and d_k tend to be opposite directions. In the case that $\beta_k \geq 0$ but $\gamma_k < 0$, such a phenomenon may also happen and lead to the non-convergence of the Beale-Powell restart algorithm.

Therefore, to establish convergence results for Algorithm 1, we restrict the values of β_k and γ_k to be nonnegative, namely, set

$$\beta_k^+ = \max\{\beta_k, 0\}, \qquad \gamma_k^+ = \max\{\gamma_k, 0\}. \tag{2.4}$$

In addition, we see that Algorithm 1 tests the sufficient descent condition

$$d_k^T g_k \leq -c_2 \|g_k\|^2 \tag{2.5}$$

for non-restart iterates, but not for restart iterates. However, the restart direction needs not be downhill if relations (2.5) is not satisfied. In this case, since the restart direction has the form

$$d_k = -g_k + \beta_k^+ d_{k-1}, \tag{2.6}$$

and since $\beta_k^+ \geq 0$, we can use the line search strategy in [5] to ensure (2.5) for non-restart iterates. The basic idea of the line search strategy in [5] is that, (a) find a point by the strong Wolfe line search, and (b) if at that point (2.5) is not satisfied, more line search iterates will proceed by the one-dimensional optimization algorithm in [15] until a new point satisfying (2.5) is found.

A modified Beale-Powell restart algorithm is then given as follows.

Algorithm 2.

Step **1** Given $x_1 \in R^n$; $c_1,c_2 \in (0,1)$, $c_3 \in (1,\infty)$, $\varepsilon \in [0,1)$;
set $d_1 = -g_1$; $k = t = 1$;

Step **2** If $k = 1$, go to Step 5;
If $k - t \geq n$, set $t = k - 1$; otherwise if $k \geq 2$ and

$$|g_{k-1}^T g_k| > c_1 \|g_k\|^2, \tag{2.7}$$

also set $t = k - 1$;

Step **3** If $k > t + 1$, compute d_k as follows and go to Step 4:

$$d_k = -g_k + \beta_k^+ d_{k-1} + \gamma_k^+ d_t, \tag{2.8}$$

where β_k^+ and γ_k^+ are computed by (2.4); otherwise, if $k = t + 1$, compute d_k by setting $\gamma_k^+ = 0$ in (2.8) and go to Step 5;

Step **4** If the following relation is not satisfied

$$-c_3 \|g_k\|^2 \leq d_k^T g_k \leq -c_2 \|g_k\|^2, \tag{2.9}$$

then set $t = k - 1$, go to Step 3;

Step **5** Carry out a line search along d_k, getting α_k;
if $k = t - 1$, perform more line searches such that (2.5) is also satisfied;
set $x_{k+1} = x_k + \alpha_k d_k$;

Step **6** If $\|g_{k+1}\| \leq \epsilon$, stop; otherwise set $k = k + 1$ and go to Step 2.

3. Convergence of the modified Algorithm

We always suppose that the objective function satisfies

Assumption 1. (i) *The level set* $\mathcal{L} = \{x \in R^n : f(x) \leq f(x_1)\}$ *is bounded;* (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*

$$\|g(x) - g(\tilde{x})\| \leq L\|x - \tilde{x}\|, \quad \forall x, \tilde{x} \in \mathcal{N}. \tag{3.1}$$

For any descent algorithm, due to the fact that $f(x_k) < f(x_{k-1})$ for all k, we know that $\{x_k\} \in \mathcal{L}$. This, with (i) in Assumption 1, implies that there exists a positive constant $B > 0$ such that for all $k \geq 1$,

$$\|x_k\| \leq B. \tag{3.2}$$

Thus by (ii) in Assumption 1, there exists a positive constant $\overline{\gamma} > 0$ such that for all $k \geq 1$,

$$\|g_k\| \leq \overline{\gamma}. \tag{3.3}$$

The line search is supposed to satisfy the standard Wolfe conditions, namely, (1.6) and

$$g(x_k + \alpha_k d_k)^T \geq \sigma g_k^T d_k, \tag{3.4}$$

where also $0 < \delta < \sigma < 1$. The following result was obtained in [16, 17, 18].

Lemma A *Suppose that x_1 is a starting point for which Assumption 1 is satisfied. Consider any method in the form (0.2) where d_k is a descent direction and α_k satisfies (1.6) and (3.4). Then*

$$\sum_{k \geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{3.5}$$

In the following, we will prove the global convergence of Algorithm 2 by contradiction. It should be noted that the proof here follows the same line as but is more difficult than that for the PRP method in [5]. Assume that

$$\liminf_{k \to \infty} \|g_k\| \neq 0. \tag{3.6}$$

Then there exists a positive constant $\gamma > 0$ such that

$$\|g_k\| \geq \gamma \tag{3.7}$$

holds for all $k \geq 1$. In this case, it follows by Lemma A and (2.5) that

$$\sum_{k \geq 1} \frac{1}{\|d_k\|^2} < \infty. \tag{3.8}$$

Lemma 3.1 *Suppose that $u, v \in R^n$ are two vectors satisfying $\|u\| = \|v\| = 1$. If*

$$w = u - \theta v \tag{3.9}$$

holds some positive number $\theta \geq 0$, then we have that

$$\|u - v\| \leq 2\|w\|. \tag{3.10}$$

Proof It follows from the definitions of u and v that

$$\|u - v\|^2 = 2 - 2u^T v. \tag{3.11}$$

Besides it, we have by (3.9) and $\theta \geq 0$ that

$$\begin{aligned} 4\|w\|^2 &= 4(1 - 2\theta u^T v + \theta^2) \\ &\geq \min_{\eta \geq 0} 4(1 - 2\eta u^T v + \eta^2) \\ &= \begin{cases} 4(1 - (u^T v)^2), & u^T v \geq 0; \\ 4, & u^T v < 0. \end{cases} \end{aligned} \tag{3.12}$$

Combining the above relations, we know that (3.10) holds. □

Lemma 3.2 *Suppose that x_1 is a starting point for which Assumption 1 is satisfied. Consider Algorithm 2, where $\epsilon = 0$, and where the line search conditions are (1.6) and (3.4). Then if (3.7) holds, $d_k \neq 0$. Further, denoting $u_k = d_k/\|d_k\|$, we have that*

$$\sum_{k \geq 2} \|u_k - u_{k-1}\|^2 < \infty. \tag{3.13}$$

Proof If (3.7) holds, we clearly have that $d_k \neq 0$ for otherwise we have from (2.5) that $\|g_k\| = 0$. Hence u_k is well-defined. Assume that $t, \bar{t}\,(\bar{t} > t)$ are any consecutive restart iterates. Since

$$d_{t+1} = -g_{t+1} + \beta_{t+1}^{+} d_t, \tag{3.14}$$

defining

$$r_k = \frac{-g_k}{\|d_k\|} \quad \text{and} \quad \delta_k = \frac{\beta_k^{+}\|d_{k-1}\|}{\|d_k\|}, \tag{3.15}$$

we have that

$$u_{t+1} = r_{t+1} + \delta_{t+1} u_t. \tag{3.16}$$

Noting the facts that $\|u_{t+1}\| = \|u_t\| = 1$ and that $\delta_{t+1} \geq 0$, we get from Lemma 3.1 and (3.3) that

$$\|u_{t+1} - u_t\| \leq 2\|r_{t+1}\| \leq \frac{2\overline{\gamma}}{\|d_{t+1}\|}. \tag{3.17}$$

If $\bar{t} \geq t + 2$, then

$$d_{t+2} = -g_{t+2} + \beta_{t+2}^{+} d_{t+1} + \gamma_{t+2}^{+} d_t. \tag{3.18}$$

Thus if we define

$$\overline{r}_{t+2} = \frac{-g_{t+2}}{\|d_{t+2}\|} + \beta_{t+2}^{+} \frac{\|d_{t+1}\|}{\|d_{t+2}\|}(u_{t+1} - u_t) \tag{3.19}$$

and

$$\overline{\delta}_{t+2} = \beta_{t+2}^{+} \frac{\|d_t\|}{\|d_{t+2}\|} + \gamma_{t+2}^{+} \frac{\|d_{t+1}\|}{\|d_{t+2}\|}, \tag{3.20}$$

(3.18) can be written as

$$u_{t+2} = \overline{r}_{t+2} + \overline{\delta}_{t+2} u_t, \tag{3.21}$$

which, with Lemma 3.1, implies that

$$\|u_{t+2} - u_t\| \leq 2\|\overline{r}_{t+2}\|. \tag{3.22}$$

In this case, applying (3.3), (3.7), (3.4), (2.9) and (2.5) in the definitions of β_k^+ and γ_k^+, we deduce that

$$\beta_k^+ \leq b, \qquad \gamma_k^+ \leq b, \tag{3.23}$$

holds for constant $b = 2\bar{\gamma}^2/(1-c_2)\gamma^2$. Hence by (3.22), (3.19), (3.17), (3.23) and (3.3), it follows that

$$\begin{aligned} \|u_{t+2} - u_t\| &\leq \frac{2\|g_{t+2}\|}{\|d_{t+2}\|} + \frac{2\beta_{t+2}^{+}\|d_{t+1}\|}{\|d_{t+2}\|}(\|u_{t+1} - u_t\|) \\ &\leq \frac{2\|g_{t+2}\|}{\|d_{t+2}\|} + \frac{4\beta_{t+2}^{+}\|d_{t+1}\|\|r_{t+1}\|}{\|d_{t+2}\|} \\ &\leq \frac{(2+4b)\bar{\gamma}}{\|d_{t+2}\|}. \end{aligned} \tag{3.24}$$

Further, by induction, it can be proved that

$$\|u_{t+i} - u_t\| \leq \frac{2[1-(2b)^i]}{1-2b} \cdot \frac{\bar{\gamma}}{\|d_{t+i}\|} \tag{3.25}$$

holds for any $0 \leq i \leq \bar{t} - t$. From (3.25) and the fact that $\bar{t} - t \leq n$, we have

$$\|u_{t+i} - u_t\| \leq \frac{c}{\|d_{t+i}\|}, \quad \forall\, 1 \leq i \leq \bar{t} - t, \tag{3.26}$$

where $c = 2[1-(2b)^n]\bar{\gamma}/(1-2b)$ is constant. (3.26) also holds for $i = 0$. For any k, let $\hat{t}$ be the last restart iterate such that $\hat{t} \leq k$, we have that

$$\begin{aligned} \|u_{k+1} - u_k\| &\leq \|u_{k+1} - u_{\hat{t}}\| + \|u_k - u_{\hat{t}}\| \\ &\leq \frac{c}{\|d_{k+1}\|} + \frac{c}{\|d_k\|}. \end{aligned} \tag{3.27}$$

It follows that

$$\sum_{k\geq 1} \|u_{k+1} - u_k\|^2 \leq 2\left(\sum_{k\geq 1} \frac{c^2}{\|d_{k+1}\|^2} + \sum_{k\geq 1} \frac{c^2}{\|d_k\|^2}\right) \leq 4c^2 \sum_{k\geq 1} \frac{1}{\|d_k\|^2}, \tag{3.28}$$

from which and (3.8) we know that (3.13) holds. $\square$

Now, we denote Z^+ to be the set of all positive integers and define

$$\mathcal{K}_{k,l}^{\lambda} = \{i \in Z^+ : k \leq i \leq k+l-1, \|s_i\| = \|x_{i+1} - x_i\| > \lambda\}. \tag{3.29}$$

In addition, we use $|\mathcal{K}_{k,l}^{\lambda}|$ to stand for the number of all elements in $\mathcal{K}_{k,l}^{\lambda}$.

Lemma 3.3 *Suppose that x_1 is a starting point for which Assumption 1 is satisfied. Consider Algorithm 2, where $\epsilon = 0$, and where the line search conditions are (1.6) and (3.4). Then if (3.7) holds, there exist a constant $\lambda > 0$ and an integer k, such that for any $l \in Z^+$,*

$$|\mathcal{K}_{k,l}^{\lambda}| > \frac{l}{2}. \tag{3.30}$$

Proof (3.8) implies that

$$\|d_k\| \to \infty, \quad k \to \infty, \tag{3.31}$$

which with (3.3) means that there exists an integer k_0 such that

$$\|g_k\| \leq \frac{1}{2}\|d_k\|, \quad \text{for all } k \geq k_0. \tag{3.32}$$

Let $t, \bar{t}\, (\bar{t} > t \geq k_0)$ to be any consecutive restart iterates and denote

$$\tilde{\beta}_{t+1} = \max\{\beta_{t+1}^+; \gamma_{t+2}^+, \ldots, \gamma_{\bar{t}}^+\}. \tag{3.33}$$

Then we have by (3.23) that

$$\tilde{\beta}_{t+1} \leq b. \tag{3.34}$$

In addition, if we denote

$$\lambda = \min\{\frac{\gamma^2}{4b^3 L\bar{\gamma}}, \frac{(1-c_1)\gamma^2}{L\bar{\gamma}}\}, \tag{3.35}$$

where L is the Lipschitz constant in (3.1), then when $\|s_t\| \le \lambda$, we have by (3.1), (3.3) and (3.7) that

$$\tilde{\beta}_{t+1} \le \frac{L\lambda\bar{\gamma}}{\gamma^2} \le \frac{1}{4b^3}. \tag{3.36}$$

Now, we prove by induction that for all $1 \le i \le \bar{t} - t$,

$$\|d_{t+i}\| \le \frac{2[1-(2b)^i]}{1-2b}\tilde{\beta}_{t+1}\|d_t\|. \tag{3.37}$$

In fact, for $i = 1$, we have from (3.14) that

$$\|d_{t+1}\| \le \|g_{t+1}\| + \beta_{t+1}^{+}\|d_t\|. \tag{3.38}$$

The above inequality, (3.32) and the definition of $\tilde{\beta}_{t+1}$ imply that (3.37) holds for $i = 1$. Suppose that (3.37) is true for some i satisfying $1 \le i < \bar{t}-t$. Then by the definition of d_{t+i+1}, (3.32), (3.23), (3.33) and the induction supposition, we have that

$$\begin{aligned}
\|d_{t+i+1}\| &\le \|g_{t+i+1}\| + \beta_{t+i+1}^{+}\|d_{t+i}\| + \gamma_{t+i+1}^{+}\|d_t\| \\
&\le \frac{1}{2}\|d_{t+i+1}\| + \left(2b\frac{1-(2b)^i}{1-2b}\tilde{\beta}_{t+1} + \gamma_{t+i+1}^{+}\right)\|d_t\| \\
&\le \frac{1}{2}\|d_{t+i+1}\| + \frac{1-(2b)^{i+1}}{1-2b}\tilde{\beta}_{t+1}\|d_t\|,
\end{aligned} \tag{3.39}$$

which implies that (3.37) is also true for $i+1$. Thus by induction, (3.37) holds for all $1 \le i \le \bar{t} - t$.

Assume without loss of generality that $b \ge 3/2$. In this case, it follows by (3.37) that

$$\|d_{t+i}\| \le (2b)^i \tilde{\beta}_{t+1}\|d_t\| \tag{3.40}$$

holds for all $1 \le i \le \bar{t}-t$. For convenience, we also assume that k_0 is exactly a restart iterate. Denote

$$T_{k_0,l} = \{t : k_0 \le t \le k_0 + l - 1, t \text{ is a restart iterate}\}. \tag{3.41}$$

Then it follows from (3.40) that for any $l \geq 1$,

$$\|d_{k_0+l}\| \leq (2b)^l \left(\prod_{t \in T_{k_0,l}} \tilde{\beta}_{t+1} \right) \|d_{k_0}\|. \tag{3.42}$$

We now proceed by contradiction and assume that for any $\lambda > 0$ and any integer $k \geq 1$, there exists an integer $l \in Z^+$, such that

$$|\mathcal{K}_{k,l}^{\lambda}| \leq \frac{l}{2}. \tag{3.43}$$

In this case, since k is arbitrary, there exists a sequence $l(j) \to \infty$, such that

$$|\mathcal{K}_{k_0,l(j)}^{\lambda}| \leq \frac{l(j)}{2}, \quad \forall j = 1, 2, \ldots \tag{3.44}$$

For any fixed j, if $i \in [k_0, k_0 + l(j) - 1]$ but $i \notin T_{k_0,l(j)}$, we claim that

$$\|s_i\| > \lambda. \tag{3.45}$$

This is because, if $i \notin T_{k_0,l(j)}$, namely, if i is a non-restart iterate, Algorithm 2 satisfies

$$|g_{i+1}^T g_i| \leq c_1 \|g_{i+1}\|^2. \tag{3.46}$$

Hence if $\|s_i\| \leq \lambda$ and if b is so large that $\lambda < \frac{(1-c_1)\gamma^2}{L\bar{\gamma}}$, we get by (3.35), (3.1), (3.3), (3.7) and the definition of λ that

$$\begin{aligned} |g_{i+1}^T g_i| &= |\|g_{i+1}\|^2 - g_{i+1}^T (g_{i+1} - g_i)| \\ &> \|g_{i+1}\|^2 - L\lambda\bar{\gamma} \\ &\geq \|g_{i+1}\|^2 - (1 - c_1)\gamma^2 \\ &\geq c_1 \|g_{i+1}\|^2, \end{aligned} \tag{3.47}$$

which contradicts (3.46). Thus (3.45) holds for all $i \notin T_{k_0,l(j)}$. Consequently, if we denote

$$\Gamma_{k_0,l(j)} = \{i \in T_{k_0,l(j)} : \|s_i\| \leq \lambda\}, \tag{3.48}$$

we have by (3.44) and (3.45) that

$$p \triangleq |\Gamma_{k_0,l(j)}| \geq \frac{l(j)}{2}. \tag{3.49}$$

Furthermore, it is clear that

$$q \triangleq |T_{k_0,l(j)}| \leq l(j). \tag{3.50}$$

In this case, if in relation (3.42), we use (3.36) when $\|s_i\| > \lambda$ and (3.34) when $\|s_i\| \le \lambda$, then

$$\begin{aligned}\|d_{k_0,l(j)}\| &\le (2b)^{l(j)} b^{q-p} (4b^3)^{-p} \|d_{k_0}\| = 2^{l(j)-2p} b^{l(j)+q-4p} \|d_{k_0}\| \\ &\le (2b^2)^{l(j)-2p} \|d_{k_0}\| \le \|d_{k_0}\|, \end{aligned} \tag{3.51}$$

where (3.49) and (3.50) are also used. By letting $j \to \infty$ in (3.51), we obtain a contradiction to (3.31). Therefore (3.30) must hold. $\square$

The following is our main theorem of this paper.

Theorem 1 *Suppose that x_1 is a starting point for which Assumption 1 is satisfied. Consider Algorithm 2, where $\varepsilon = 0$, and where the line search conditions are (1.6) and (3.4). Then the following convergence relation holds*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{3.52}$$

Proof We proceed by contradiction and assume that (3.52) is not satisfied. Then (3.7) holds and hence the conditions of Lemmas 3.2 and 3.3 are satisfied. For any integers l, k, we write

$$\begin{aligned} x_{k+l} - x_k &= \sum_{i=k}^{k+l-1} \|s_i\| u_i \\ &= \sum_{i=k}^{k+l-1} \|s_i\| u_k + \sum_{i=k}^{k+l-1} \|s_i\| (u_i - u_k). \end{aligned} \tag{3.53}$$

Taking norms in (3.53) and using (3.2), we get that

$$\sum_{i=k}^{k+l-1} \|s_i\| \le 2B + \sum_{i=k}^{k+l-1} \|s_i\| \|u_i - u_k\|. \tag{3.54}$$

Define λ by (3.35) and $l = \{8B/\lambda\}$, where $\{T\}$ stands for the smallest integer not less than T. Lemma 2 implies that

$$\|u_{i+1} - u_i\| \to 0, \quad i \to \infty. \tag{3.55}$$

Then if k_0 is so large that

$$\|u_{j+1} - u_j\| \le \frac{1}{2l} \tag{3.56}$$

for all $k_0 \leq j \leq k_0 + l - 1$, we have that

$$\|u_{k_0+i} - u_{k_0}\| \leq \sum_{j=k_0}^{k_0+i-1} \|u_{j+1} - u_j\| \leq \frac{i}{2l} \leq \frac{1}{2} \tag{3.57}$$

for all $1 \leq i \leq l$. Using (3.54), (3.2), (3.57) and Lemma 3, we obtain

$$2B \geq \frac{1}{2} \sum_{i=k_0}^{k_0+l-1} \|s_i\| \geq \frac{\lambda}{2} |\mathcal{K}_{k_0,l}^{\lambda}| > \frac{\lambda l}{4}. \tag{3.58}$$

So $l < 8B/\lambda$, which contradicts the definition of l. Therefore (3.52) must hold. □

Acknowledgements. The authors thank Doctor Zheng-hui Xie for providing reference [15], and the referees and editors for their many valuable suggestions on the draft of this paper.

References

[1] Fletcher R, Reeves C. Function minimization by conjugate gradients. Comput J, 1964, 7: 149-154

[2] Polak E, Ribière G. Note sur la convergence de directions conjugées. Rev Francaise Informat Recherche Opertionelle, 1969, 3(16): 35-43

[3] Polyak B T. The conjugate gradient method in extremem problems. USSR Comp Math and Math Phys, 1969, 9: 94-112

[4] Dai Y H, Yuan Y. Convergence properties of the Fletcher-Reeves method. IMA J Numer Anal, 1996, 16(2): 155-164

[5] Gilbert J C, Nocedal J. Global convergence properties of conjugate gradient methods for optimization. SIAM J Optimization, 1992, 2(1): 21-42

[6] Crowder H P, Wolfe P. Linear convergence of the conjugate gradient method. IBM J Res Dev, 1969, 16: 431-433

[7] Yuan Y. Analysis on the conjugate gradient method. Optimization Methods and Software, 1993, 2: 19-29

[8] Cohen A. Rate of convergence of several conjugate gradient algorithms. SIAM J Numer Anal, 1972, 9: 248-259

[9] McCormick G P, Ritter K. Alternative Proofs of the convergence properties of the conjugate-gradient method. JOTA, 1975, 13(5): 497-518

[10] Powell M J D. Restart procedures of the conjugate gradient method. Math Program, 1977, 2: 241-254

[11] Beale E M L. A derivative of conjugate gradients. in: Lootsma F A, eds. Numerical methods for nonlinear optimization, London, Academic Press, 1972, 39-43

[12] Deng N, Li Z. Global convergence of three terms conjugate gradient methods. Optimization Methods and Software, 1995, 4: 273-282

[13] Liao D, Liu C. Some new techniques in numerical weather prediction. Beijing, Meteorological Publishers, 1995.

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

[15] Morè J J, Thuente D J. Line search algorithms with guaranteed sufficient decrease. ACM Transactions on Mathematics Software, 1994, 20(3): 286-307

[16] Wolfe P. Convergence conditions for ascent methods. SIAM Review, 1969, 11: 226-235

[17] Wolfe P. Convergence conditions for ascent methods. II: Some corrections. SIAM Review, 1969, 13: 185-188

[18] Zoutendijk G. Nonlinear Programming, Computational Methods. in: Abadie J, eds. Integer and Nonlinear Programming, North-Holland, Amsterdam, 1970, 37-86