*IMA Journal of Numerical Analysis* (2002) **22**, 1–10

# $R$-linear convergence of the Barzilai and Borwein gradient method

YU-HONG DAI†
*State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, PO Box 2719, Beijing 100080, People's Republic of China*

AND

LI-ZHI LIAO‡
*Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong, People's Republic of China*

[Received on 6 September 1999; revised on 4 April 2001]

Combined with non-monotone line search, the Barzilai and Borwein (BB) gradient method has been successfully extended for solving unconstrained optimization problems and is competitive with conjugate gradient methods. In this paper, we establish the $R$-linear convergence of the BB method for any-dimensional strongly convex quadratics. One corollary of this result is that the BB method is also locally $R$-linear convergent for general objective functions, and hence the stepsize in the BB method will always be accepted by the non-monotone line search when the iterate is close to the solution.

*Keywords*: unconstrained optimization; gradient method; $R$-linear convergence; strictly convex.

## 1. Introduction

Consider the problem of minimizing a strictly convex quadratic,

$$\min f(x) = \tfrac{1}{2}x^t Ax - b^t x, \tag{1.1}$$

where $A \in R^{n\times n}$ is a real symmetric positive definite matrix and $b \in R^n$. The Barzilai and Borwein (BB) gradient method for solving (1.1) has the form

$$x_{k+1} = x_k - \alpha_k^{-1} g_k, \tag{1.2}$$

where $g_k = \nabla f(x_k)$ and $\alpha_k$ is determined by the information achieved at the points $x_{k-1}$ and $x_k$. Denote $s_{k-1} = x_k - x_{k-1}$ and $y_{k-1} = g_k - g_{k-1}$. Since the matrix $D_k = \alpha_k I$ is an approximation to the Hessian of $f$ at $x_k$, Barzilai & Borwein (1988) so chose the stepsize $\alpha_k$ such that $D_k$ has a certain quasi-Newton property

$$D_k = \arg\min_{D=\alpha I} \|Ds_{k-1} - y_{k-1}\|_2, \tag{1.3}$$

†Email: dyh@lsec.cc.ac.cn
‡Email: liliao@hkbu.edu.hk

yielding

$$\alpha_k = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}}. \tag{1.4}$$

Compared with the classical steepest descent method, which can be dated to Cauchy (1847), the BB method often requires less computational work and speeds up the convergence greatly (see Akaike, 1959 and Fletcher, 1990). A direct application of the Barzilai and Borwein method in chemistry can be found in Glunt *et al.* (1993).

To extend the BB method to minimize a general smooth function,

$$\min f(x), \quad x \in R^n, \tag{1.5}$$

Raydan (1997) considered the use of the non-monotone line search proposed by Grippo *et al.* (1986). The resulting algorithm, called the global Barzilai and Borwein algorithm, is proved to be globally convergent for general functions and is competitive with some standard conjugate gradient codes (see Raydan, 1997). A successful application of the global Barzilai and Borwein algorithm can be found in Birgin *et al.* (1999). The idea of Raydan (1997) was further extended by Birgin *et al.* (2000) for minimizing differentiable functions on closed convex sets, resulting in more efficient projected gradient methods. Liu & Dai (1999) recently provided a powerful scheme for unconstrained optimization problems with strong noises by combining the BB method and the stochastic approximation method. Other work related to the BB method can be found in Birgin *et al.* (2000) and Friedlander *et al.* (1999). To sum up, due to its simplicity and numerical efficiency, the Barzilai and Borwein gradient method has now received a good deal of attention in the optimization community.

In this paper, we are concerned with the convergence rate of the BB method. The convergence analysis of the method is difficult and non-standard so that convergence results are often provided for convex quadratics. For two-dimensional convex quadratics, Barzilai & Borwein (1988) established the $R$-superlinear convergence of the method. Raydan (1993) proved that the method can always give the unique solution $x^* = A^{-1}b$ of problem (1.1) for any-dimensional quadratics. Under a restrictive assumption, Molina & Raydan (1996) established the $Q$-linear rate of convergence of the (preconditioned) BB method. Assume that $\lambda_1$ and $\lambda_n$ are the minimal and maximal eigenvalues of $A$, respectively. Their assumption says that

$$\lambda_n < 2\lambda_1, \tag{1.6}$$

that is very strong. By refining the analysis in Raydan (1993), we will establish in the paper the $R$-linear convergence of the BB method for any-dimensional strictly convex quadratics (see Section 2). No additional restriction is required by our result. One corollary of this result is that, the method is also locally $R$-linearly convergent for general objective functions, and hence the stepsize in the method will always be accepted when the iterate is close to the solution if a non-monotone line search with suitable parameters are used (see the discussions in Section 3).

## 2. Convergence analysis

In this section, we establish the $R$-linear convergence of the BB method applied to any quadratic function

$$f(x) = \tfrac{1}{2}x^t Ax - b^t x,$$

where $A \in R^{n\times n}$ is symmetric and positive definite. Our analysis will proceed with the gradient sequence $\{g_k\}$, whereas the analysis in Raydan (1993) proceeds with the sequence $\{x^* - x_k\}$ with $x^* = A^{-1}b$.

Let $\{x_k\}$ be the sequence generated by the BB method from initial vectors $x_0$ and $x_1$. Then, using (1.2) and the fact that $g_k = Ax_k - b$, we have for all $k \geqslant 1$,

$$g_{k+1} = \frac{1}{\alpha_k}(\alpha_k I - A)g_k. \tag{2.1}$$

Assume that the eigenvalues of $A$ are $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$ and $\{v_1, v_2, \ldots, v_n\}$ are associated orthonormal eigenvectors. For any initial vector $x_1$, there exist constants $d_1^1$, $d_2^1, \ldots, d_n^1$ such that

$$g_1 = \sum_{i=1}^{n} d_i^1 v_i.$$

Using (2.1) we can obtain for any $k \geqslant 1$,

$$g_{k+1} = \sum_{i=1}^{n} d_i^{k+1} v_i, \tag{2.2}$$

where

$$d_i^{k+1} = \left(\frac{\alpha_k - \lambda_i}{\alpha_k}\right) d_i^k = \prod_{j=1}^{k}\left(\frac{\alpha_j - \lambda_i}{\alpha_j}\right) d_i^1. \tag{2.3}$$

First, we have the following lemma.

LEMMA 2.1 For all $i = 1, \ldots, n$ and $k \geqslant 1$, we have that

$$(d_i^{k+1})^2 \leqslant \delta^2 (d_i^k)^2 \tag{2.4}$$

and

$$\|g_{k+1}\|_2 \leqslant \delta \|g_k\|_2, \tag{2.5}$$

where

$$\delta = \frac{\lambda_n}{\lambda_1} - 1. \tag{2.6}$$

*Proof.* By (2.1) and the fact that $s_k = -\frac{1}{\alpha_k} g_k$, we can write $\alpha_k$ in (1.4) as

$$\alpha_k = \frac{g_{k-1}^t A g_{k-1}}{g_{k-1}^t g_{k-1}}. \tag{2.7}$$

Namely, $\alpha_k$ is the Rayleigh quotient of $A$ at vector $g_{k-1}$. Hence,

$$\lambda_1 \leqslant \alpha_k \leqslant \lambda_n \quad \text{for all } k. \tag{2.8}$$

From (2.8) and the first equality in (2.3), it is easy to obtain (2.4) if $\lambda_i \geqslant \alpha_k$. In the case of $\lambda_i < \alpha_k$, from $\frac{\alpha_k}{\lambda_i} \leqslant \frac{\lambda_n}{\lambda_1}$ and $(\lambda_1 - \lambda_n)(\lambda_i - \alpha_k) \geqslant 0$, it is straightforward to get $\frac{\lambda_i}{\alpha_k} - 1 \geqslant 1 - \frac{\lambda_n}{\lambda_1}$. This, together with the first equality in (2.3), implies (2.4). Therefore in both cases, (2.4) holds. From (2.2) and the orthonormality of the eigenvectors $\{v_1, \ldots, v_n\}$, we have

$$\|g_k\|_2^2 = \sum_{i=1}^{n} (d_i^k)^2. \tag{2.9}$$

This, together with (2.4), implies (2.5). □

If $\delta < 1$, then by (2.5) we immediately know that $\|g_k\|_2$ is $Q$-linearly convergent and hence $R$-linearly convergent. Since it follows from (1.6) that $\delta < 1$, we re-obtain the $Q$-linear convergence result of the BB method in Molina & Raydan (1996).

To establish the $R$-linear rate of convergence for the BB method, careful considerations must be given for the case that $\delta \geqslant 1$. For this purpose, we let $c$ be the constant

$$c = 1 - \frac{\lambda_1}{\lambda_n}. \tag{2.10}$$

Then it follows from $\delta \geqslant 1$ that $\frac{1}{2} \leqslant c < 1$. For any number $\varepsilon_l \in (0, \frac{1}{4}]$ and positive integer $m_l$, we denote the positive integer $\Delta_l$ as follows:

$$\Delta_l = \left\lceil \frac{\log(2\varepsilon_l \delta^{-2(m_l+1)})}{2\log c} \right\rceil. \tag{2.11}$$

In addition, we introduce a useful quantity $D(k, l)$:

$$D(k, l) = \sum_{i=1}^{l} (d_i^k)^2. \tag{2.12}$$

Then it is easy to see that $D(k, n) = \|g_k\|_2^2$. We now prove the following two lemmas for the case that $\delta \geqslant 1$.

LEMMA 2.2 Assume that $\delta \geqslant 1$. For any integer $1 \leqslant l < n$ and $k \geqslant 1$, if there exist positive number $\varepsilon_l$ and positive integer $m_l$ such that

$$D(k + j, l) \leqslant \varepsilon_l \|g_k\|_2^2, \quad \text{for all } j \geqslant m_l, \tag{2.13}$$

then there must exist some integer $j_0 \in [m_l, m_l + \Delta_l + 1]$ such that

$$(d_{l+1}^{k+j_0})^2 \leqslant 2\varepsilon_l \|g_k\|_2^2, \tag{2.14}$$

where $\Delta_l$ is given by (2.11).

*Proof.* To prove this lemma, it suffices to show that if

$$(d_{l+1}^{k+j})^2 > 2\varepsilon_l \|g_k\|_2^2, \quad \text{for all } j \in [m_l, m_l + \Delta_l], \tag{2.15}$$

then we must have that

$$(d_{l+1}^{k+m_l+\Delta_l+1})^2 \leqslant 2\varepsilon_l \|g_k\|_2^2. \tag{2.16}$$

Assume that (2.15) is true. For any $k \geqslant 1$, we have by (2.4) and (2.9) that

$$(d_{l+1}^{k+m_l+1})^2 \leqslant \delta^{2(m_l+1)} (d_{l+1}^k)^2 \leqslant \delta^{2(m_l+1)} \|g_k\|_2^2. \tag{2.17}$$

On the other hand, it follows from (2.7) and (2.9) that

$$\alpha_{k+1} = \frac{\sum_{i=1}^n (d_i^k)^2 \lambda_i}{\sum_{i=1}^n (d_i^k)^2}. \tag{2.18}$$

By this, (2.13) and (2.15), we have that, for any $j \in [m_l, m_l + \Delta_l]$,

$$\alpha_{k+j+1} \geqslant \frac{\lambda_{l+1} \sum_{i=l+1}^n (d_i^{k+j})^2}{\varepsilon_l \|g_k\|_2^2 + \sum_{i=l+1}^n (d_i^{k+j})^2} \geqslant \frac{2}{3} \lambda_{l+1}, \tag{2.19}$$

which, together with (2.8) and (2.3), indicates that

$$(d_{l+1}^{k+j+2})^2 \leqslant c^2 (d_{l+1}^{k+j+1})^2, \quad \text{for all } j \in [m_l, m_l + \Delta_l], \tag{2.20}$$

where $c$ is the constant in (2.10). Thus by (2.17), (2.20) and the definition (2.11) of $\Delta_l$, we obtain

$$(d_{l+1}^{k+m_l+\Delta_l+1})^2 \leqslant c^{2\Delta_l} (d_{l+1}^{k+m_l+1})^2 \leqslant c^{2\Delta_l} \delta^{2(m_l+1)} \|g_k\|_2^2 \leqslant 2\varepsilon_l \|g_k\|_2^2.$$

This completes our proof. □

LEMMA 2.3 Assume that $\delta \geqslant 1$. For any integer $1 \leqslant l < n$ and $k \geqslant 1$, assume that there exist positive number $\varepsilon_l$ and positive integer $m_l$ such that (2.13) holds. Denote $\varepsilon_{l+1} = (1 + 2\delta^4)\varepsilon_l$ and $m_{l+1} = m_l + \Delta_l + 1$. Then we must also have that

$$D(k + j, l + 1) \leqslant \varepsilon_{l+1} \|g_k\|_2^2, \quad \text{for all } j \geqslant m_{l+1}. \tag{2.21}$$

*Proof.* Notice that

$$D(k + j, l + 1) = D(k + j, l) + (d_{l+1}^{k+j})^2. \tag{2.22}$$

Thus by (2.13) and the definition of $\varepsilon_{l+1}$, it suffices to prove that

$$(d_{l+1}^{k+j})^2 \leqslant 2\delta^4 \varepsilon_l \|g_k\|_2^2 \tag{2.23}$$

holds for all $j \geqslant m_{l+1}$. By Lemma 2.2, we know that there must exist some integer $j_0 \in [m_l, m_l + \Delta_l + 1]$ such that (2.14) holds. If (2.14) holds for all $j \geqslant j_0$, then (2.23) naturally holds since $\delta \geqslant 1$. Now, let us prescribe $\hat{j} = j_0$ and denote $j_1 \geqslant \hat{j}$ to be the integer for which

$$(d_{l+1}^{k+j})^2 \leqslant 2\varepsilon_l \|g_k\|_2^2, \quad \text{for } j_0 \leqslant j \leqslant j_1, \tag{2.24}$$

but

$$(d_{l+1}^{k+j_1+1})^2 > 2\varepsilon_l \|g_k\|_2^2. \tag{2.25}$$

Similar to (2.19), (2.20), we can get by (2.13) and (2.25) that

$$\alpha_{k+j_1+2} \geqslant \tfrac{2}{3}\lambda_{l+1}, \quad (d_{l+1}^{k+j_1+3})^2 \leqslant c^2 (d_{l+1}^{k+j_1+2})^2.$$

Generally, if $j \geqslant m_l$ and $(d_{l+1}^{k+j+1})^2 > 2\varepsilon_l \|g_k\|_2^2$, we have by this and (2.13) that $(d_{l+1}^{k+j+3})^2 \leqslant c^2 (d_{l+1}^{k+j+2})^2$. Since $c < 1$, we know from this fact and (2.25) that there exists an integer, $j_2$ say, such that

$$(d_{l+1}^{k+j+1})^2 > 2\varepsilon_l \|g_k\|_2^2, \quad \text{for } j_1 \leqslant j \leqslant j_2 - 2 \tag{2.26}$$

but

$$(d_{l+1}^{k+j_2})^2 \leqslant 2\varepsilon_l \|g_k\|_2^2. \tag{2.27}$$

Relations (2.13) and (2.26) imply that

$$(d_{l+1}^{k+j+3})^2 \leqslant c^2 (d_{l+1}^{k+j+2})^2, \quad \text{for } j_1 \leqslant j \leqslant j_2 - 2. \tag{2.28}$$

In addition, by (2.4) and (2.24), it is obvious that

$$(d_{l+1}^{k+j_1+1})^2 \leqslant 2\delta^2 \varepsilon_l \|g_k\|_2^2 \tag{2.29}$$

and

$$(d_{l+1}^{k+j_1+2})^2 \leqslant 2\delta^4 \varepsilon_l \|g_k\|_2^2. \tag{2.30}$$

Since $c < 1$, we obtain from (2.30) and (2.28) that

$$(d_{l+1}^{k+j+3})^2 < 2\delta^4 \varepsilon_l \|g_k\|_2^2, \quad \text{for } j_1 \leqslant j \leqslant j_2 - 2. \tag{2.31}$$

Thus by (2.24), (2.29)–(2.31), we know that (2.23) holds for all $j \in [j_0, j_2]$. Letting $\hat{j} = j_2$ and repeating the above process, we can eventually obtain the result (2.23) for all $j \geqslant j_0$. This, with $j_0 \leqslant m_{l+1}$, completes our proof. □

Considering the cases $\delta < 1$ and $\delta \geqslant 1$ together, we now can provide the following lemma.

LEMMA 2.4 Let $f(x)$ be a strictly convex quadratic function. Let $\{x_k\}$ be the sequence generated by the BB method. Then, there exists a positive integer $m$ which depends only on $\lambda_1$ and $\lambda_n$ such that

$$\|g_{k+m}\|_2 \leqslant \tfrac{1}{2}\|g_k\|_2 \quad \text{for all } k \geqslant 1. \tag{2.32}$$

*Proof.* If $\delta < 1$, we have from Lemma 2.1 that (2.32) holds with

$$m = \left\lceil -\frac{\log 2}{\log \delta} \right\rceil. \tag{2.33}$$

Assume that $\delta \geqslant 1$. For any $1 \leqslant l \leqslant n$, we denote

$$\varepsilon_l = \tfrac{1}{4}(1 + 2\delta^4)^{l-n}.$$

Let $m_1 = \left\lceil \frac{\log \varepsilon_1}{2\log c} \right\rceil$, $m_{l+1} = m_l + \Delta_l + 1$ for $l = 1, \ldots, n-1$ and $m = m_n$, where $c$ and $\Delta_l$ are given by (2.10) and (2.11), respectively. It is easy to see that $m$ depends only on $\lambda_1$ and $\lambda_n$. In addition, note that $\|g_k\|_2^2 = D(k, n)$ and $\varepsilon_n = \frac{1}{4}$. Thus, to prove (2.32) for any $k$, it suffices to show that the following relation holds for $l = 1, \ldots, n$:

$$D(k + j, l) \leqslant \varepsilon_l \|g_k\|_2^2, \quad \text{for all } j \geqslant m_l. \tag{2.34}$$

We show (2.34) by induction. By the first equality in (2.3) and (2.8), we have

$$(d_1^{k+1})^2 = \left(\frac{\alpha_k - \lambda_1}{\alpha_k}\right)^2 (d_1^k)^2 \leqslant c^2 (d_1^k)^2. \tag{2.35}$$

which, with $D(k, 1) = (d_1^k)^2$, gives

$$D(k + j, 1) \leqslant c^{2j} (d_1^k)^2 \leqslant c^{2j} \|g_k\|_2^2.$$

The above relation and the definition of $m_1$ indicate that (2.34) holds for $l = 1$. Now we suppose that (2.34) holds for some $1 \leqslant l \leqslant n-1$. Then we know by Lemma 2.3 that (2.34) also holds for $l+1$. Thus, by induction principle, (2.34) holds for all $1 \leqslant l \leqslant n$. This completes our proof. □

Now we are able to establish the $R$-linear convergence of the BB gradient method applied to any strictly convex quadratic function.

THEOREM 2.5 Let $f(x)$ be a strictly convex quadratic function. Let $\{x_k\}$ be the sequence generated by the BB method. Then, either $g_k = 0$ for some finite $k$, or the sequence $\{\|g_k\|_2\}$ converges to zero $R$-linearly.

*Proof.* We only need to consider the case that $g_k \neq 0$ for all $k$. By Lemma 2.4, we have that there exists a positive integer $m$ such that

$$\|g_{mj+1}\|_2 \leqslant \tfrac{1}{2}\|g_{mj-m+1}\|_2, \quad \text{for all } j \geqslant 1.$$

It follows that

$$\|g_{mj+1}\|_2 \leqslant (\tfrac{1}{2})^j \|g_1\|_2. \tag{2.36}$$

For any $k \geqslant 1$, let $k = mj + i$, where $j \geqslant 0$ and $i \in [1, m]$ are integers. It is obvious that $j \geqslant k/m - 1$. By this, (2.5) and (2.36), we can obtain

$$\begin{aligned}\|g_k\|_2 &\leqslant \delta^{i-1}\|g_{mj+1}\|_2 \leqslant \delta^{i-1}(\tfrac{1}{2})^j \|g_1\|_2 \\ &\leqslant \delta^{m-1}(\tfrac{1}{2})^{\frac{k}{m}-1}\|g_1\|_2 \leqslant c_1 c_2^k \|g_1\|_2,\end{aligned} \tag{2.37}$$

where $c_1 = 2\delta^{m-1}$ and $c_2 = 2^{-\frac{1}{m}} < 1$ are constants. Therefore, the sequence $\{\|g_k\|_2\}$ converges to zero $R$-linearly. This completes our proof. □

## 3. Some discussions

In this paper, by refining the analysis in Raydan (1993), we have established the $R$-linear convergence of the BB gradient method for any-dimensional strictly convex quadratics. No additional restriction is required by this result, whereas a restrictive assumption (1.6) is needed for the result in Molina & Raydan (1996). Similar to Theorem 2.5, one can easily draw the conclusion that the preconditioned BB method studied by Molina & Raydan (1996) is $R$-linearly convergent, since its iterate can be regarded to be generated by the BB method for a quadratic function with its Hessian being $E^{-1}AE^{-t}$ if $C = EE^t$ is the preconditioner.

By regarding $D_k^{-1}$ as an approximation to the inverse Hessian of $f$ at $x_k$ and choosing $D_k^{-1}$ such that

$$D_k^{-1} = \arg \min_{D=\alpha^{-1}I} \|s_{k-1} - D^{-1}y_{k-1}\|_2, \tag{3.1}$$

Barzilai & Borwein (1988) also obtained the following choice for $\alpha_k$:

$$\alpha_k = \frac{s_{k-1}^t y_{k-1}}{y_{k-1}^t y_{k-1}}. \tag{3.2}$$

The above choice often performs worse than (1.4) in practical computations. Nevertheless, in a similar way, we can also prove that the method (1.2)–(3.2) is $R$-linearly convergent for strongly convex quadratics. In addition, the result in the paper can also be easily extended to those gradient methods with retards considered in Friedlander *et al.* (1999).

From a practical point of view, one can extend the result in this paper to explain the fact that, for general unconstrained optimization problems, the stepsize in the BB method can always be accepted by the non-monotone line search, as observed by Raydan (1997). To do so, we should first note that the $R$-linear convergence result of the BB method has been extended for general unconstrained optimization problems. See Liu & Dai (1999) for the sketch of its proof. Suppose that $f$ is a three times continuously differentiable function and $x^*$ is a point for which $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) > 0$, and that the sequence $\{x_k\}$ generated by the BB method converges to $x^*$. The idea of Liu & Dai (1999) is to consider an additional iterate $\{\hat{x}_k\}$ generated by the BB method for the quadratic function

$$\hat{f}(x) = f(x^*) + \tfrac{1}{2}(x - x^*)^t \nabla^2 f(x^*)(x - x^*), \tag{3.3}$$

and then to compare $\|x_k - x^*\|_2$ and $\|\hat{x}_k - x^*\|_2$. They were able to show that there exists an infinite subsequence $\{k_i\} \subset \{1, 2, \dots\}$ such that the relations

$$k_{i+1} - k_i \leqslant M_1, \tag{3.4}$$

$$\|x_{k_{i+1}} - x^*\|_2 \leqslant c'\|x_{k_i} - x^*\|_2 \tag{3.5}$$

hold for some constant $c' \in (0, 1)$ and positive integer $M_1$, and hence the BB method is also locally $R$-linearly convergent for general objective functions.

On the other hand, under the assumption that $f(x)$ is uniformly convex and that the search direction $d_k$ satisfies the conditions

$$g_k^T d_k \leqslant -\tau_1\|g_k\|^2 \quad \text{and} \quad \|d_k\|_2 \leqslant \tau_2\|g_k\|^2, \tag{3.6}$$

Dai (2000) proved that the stepsize $\alpha_k$ is always accepted by the Grippo *et al.* (1986) non-monotone line search with suitable parameters if and only if the iterative method $x_{k+1} = x_k + \alpha_k d_k$ is such that relations (3.5) and (3.4) hold. (The non-monotone line search in Grippo *et al.* (1986) is to choose the first non-negative integer $h_k$ such that the stepsize $\alpha_k = \bar{\alpha_k}\sigma^{h_k}$ satisfies $f(x_k + \alpha_k d_k) \leqslant \max_{0\leqslant j\leqslant m(k)}[f(x_{k-j})] + \delta\alpha_k g_k^T d_k$, where $\bar{\alpha_k}$ is the first trial stepsize, $\sigma \in (0, 1)$, $m(0) = 0$ and $0 \leqslant m(k) \leqslant \min[m(k-1)+1, M-1]$, $k \geqslant 1$.) Since the negative gradient $-g_k$ clearly satisfies (3.6), and since the BB method satisfies (3.4) and (3.5), we then know that the stepsize in the BB method will always be accepted by the Grippo *et al.* non-monotone line search when $x_k$ is close to $x^*$.

Generally, like the steepest descent method and the conjugate gradient method, the BB method becomes slow when a problem is more ill-conditioned: for example, see Friedlander *et al.* (1999). Noting that $\delta = \frac{\lambda_n}{\lambda_1} - 1 = \text{cond}(A) - 1$, it is easy to see that the value of $m$ in Lemma 2.4 is increasing with the condition number of $A$, and hence the value $c_2^{-\frac{1}{m}}$ will tend to 1. This explains to some extent why the BB method is also affected by the problem condition.

Finally, we should note that the analysis in this paper can be further refined such that the value of $c_2$ in (2.37) becomes smaller. For example, in Lemma 2.2, we can obtain from (2.3) and (2.19) that

$$(d_{l+1}^{k+m_l+1})^2 \leqslant \bar{\delta}^{2(m_l+1)}(d_{l+1}^k)^2 \leqslant \bar{\delta}^{2(m_l+1)}\|g_k\|_2^2, \tag{3.7}$$

where

$$\bar{\delta} = \min\left\{\frac{1}{2}, 1 - \frac{\lambda_{l+1}}{\lambda_n}\right\}. \tag{3.8}$$

Hence, we may choose in Lemma 2.2

$$\Delta_l = \left\lceil \frac{\log(2\varepsilon_l \bar{\delta}^{-2(m_l+1)})}{2\log c} \right\rceil, \tag{3.9}$$

decreasing the values of $m$ and $c_2$ in case of $\bar{\delta} < \delta$ (if $\bar{\delta} = \delta$, we know similar to (2.35) that $d_{l+1}^k$ converges $Q$-linearly). However, it is doubtful whether such analyses can lead to a convergence result better than the steepest descent method even for three-dimensional quadratics. The latter is known to be $Q$-linearly convergent. In other words, it still remains under investigation whether there exists any theoretical evidence demonstrating the numerical efficiency of the BB method over the classical steepest descent method.

## Acknowledgements

The authors are much indebted to Professor Roger Fletcher whose comments and suggestions improved this paper greatly. This research was supported in part by the Chinese NSF grant 19801033 and grant FRG/99-00/II-23 of Hong Kong Baptist University.

## REFERENCES

AKAIKE, H. 1959 On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method. *Ann. Inst. Statist. Math. Tokyo*, **11**, 1–17.

BARZILAI, J. & BORWEIN, J. M. 1988 Two-point step size gradient methods. *IMA J. Numer. Anal.*, **8**, 141–148.

BIRGIN, E. G., CHAMBOULEYRON, I. & MARTÍNEZ, J. M. 1999 Estimation of the optical constants and the thickness of thin films using unconstrained optimization. *J. Comput. Phys.*, **151**, 862–880.

BIRGIN, E. G., MARTÍNEZ, J. M. & RAYDAN, M. 2000 Nonmonotone spectral projected gradient methods on convex sets. *SIAM J. Optim.*, **10**, 1196–1211.

CAUCHY, A. 1847 Méthode générale pour la résolution des systèms d'équations simultanées. *Comp. Rend. Sci. Paris*, **25**, 536–538.

DAI, Y. H. 2000 On nonmonotone line search. *Research Report (accepted by J. Optim. Theor. Appl.)*.

FLETCHER, R. 1990 Low storage methods for unconstrained optimization. *Lectures Appl. Math. (AMS)*, **26**, 165–179.

FRIEDLANDER, A., MARTÍNEZ, J. M, MOLINA, B. & RAYDAN, M. 1999 Gradient method with retards and generalizations. *SIAM J. Numer. Anal.*, **36**, 275–289.

GLUNT, W., HAYDEN, T. L. & RAYDAN, M. 1993 Molecular conformations from distance matrices. *J. Comput. Chem.*, **14**, 114–120.

GRIPPO, L., LAMPARIELLO, F. & LUCIDI, S. 1986 A nonmonotone line search technique for Newton's method. *SIAM J. Numer. Anal.*, **23**, 707–716.

LIU, W. B. & DAI, Y. H. 1999 Minimization algorithms based on supervisor and searcher co-operation: I—faster and robust gradient algorithms for minimization problems with stronger noises. *Research Report (to appear in J. Optim. Theor. Appl.)*.

MOLINA, B. & RAYDAN, M. 1996 Preconditioned Barzilai–Borwein method for the numerical solution of partial differential equations. *Numer. Algorithms*, **13**, 45–60.

RAYDAN, M. 1993 On the Barzilai and Borwein choice of steplength for the gradient method. *IMA J. Numer. Anal.*, **13**, 321–326.

RAYDAN, M. 1997 The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM J. Optim.*, **7**, 26–33.