J Optim Theory Appl (2017) 172:179–186
DOI 10.1007/s10957-016-1008-9

A New Modified Barzilai–Borwein Gradient Method for the Quadratic Minimization Problem

Yutao Zheng[1] · **Bing Zheng**[1]

Received: 6 November 2015 / Accepted: 3 September 2016 / Published online: 21 September 2016

Abstract A new modified Barzilai–Borwein gradient method for solving the strictly convex quadratic minimization problem is proposed by properly changing the Barzilai–Borwein stepsize such that some certain multi-step quasi-Newton condition is satisfied. The global convergence is analyzed. Numerical experiments show that the new method can outperform some known gradient methods.

Keywords BB gradient method · Modified BB gradient method · Multi-step method · Global convergence

Mathematics Subject Classification 65K05 · 90C20 · 90C52

1 Introduction

We study the quadratic minimization problem. One of the simplest and most fundamental methods for solving this problem is the steepest descent (SD) method (or gradient method). However, it converges linearly, performs poorly and is badly affected by ill-conditioning. Therefore, many efforts have been made to overcome these defects (see, e.g., [1–4]).

An efficient gradient method (simply called BB method) was proposed by Barzilai and Borwein [1], in which some quasi-Newton property is used. The BB method

✉ Yutao Zheng
zhengyt13@lzu.edu.cn

Bing Zheng
bzheng@lzu.edu.cn

1 School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, People's Republic of China

is shown to be globally and R-linearly convergent [3,5]; it can outperform the SD method in practice. As the efficient modifications for BB method, an alternate stepsize (AS) gradient method through alternately using the SD stepsize and BB stepsize was provided by Dai [6] and an alternate minimization (AM) gradient method by alternately using the two types of BB stepsize was presented by Dai and Yuan [4]. Early before these works, the Cauchy-Barzilar-Borwein (CBB) method by Raydan [2] doubles the BB stepsize in an iteration to reduce the computation. More related works on the gradient method can be found in [7–9]. In this note, we will propose a new strategy for choice of the stepsize in gradient method, which leads to a new modification for BB method. This idea comes from the multi-step secant condition derived by Ford and Moghrabi [10,11]. The numerical experiments show that our method can have much better performance than those of the BB method and its modified versions mentioned previously, including a method proposed by Dai and Fletcher [12].

The remainder of the paper is organized as follows. In Sect. 2, we introduce a modified BB method by deriving a new stepsize. The convergence analysis is made in Sect. 3 and the numerical comparisons between our method and some other gradient methods are given in Sect. 4. Finally, in Sect. 5, we give some conclusions to end the paper.

2 Modified BB Gradient Method

Consider the following quadratic minimization problem

$$\min f(x) = \frac{1}{2}x^T Ax - b^T x, \tag{1}$$

where $x,\ b \in \mathbb{R}^n$ are vectors, $A \in \mathbb{R}^{n\times n}$ is a symmetric and positive definite matrix. Let x_k be the current iteration point, g_k the gradient of $f(x)$ at x_k, and $s_{k-1} = x_k - x_{k-1}$, $y_{k-1} = g_k - g_{k-1}$. The gradient method for (1) is of the form

$$x_{k+1} = x_k - \alpha_k^{-1} g_k, \tag{2}$$

where α_k is a stepsize. The SD stepsize and BB stepsize are, respectively,

$$\alpha_k^{SD} = \frac{g_k^T A g_k}{g_k^T g_k} \quad \text{and} \quad \alpha_k^{BB} = \frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}.$$

Using interpolation, Ford and Moghrabi [10,11] presented a multi-step quasi-Newton method for general unconstrained optimization problem. Based on the experience with a two-step method of this kind, we consider the following secant condition

$$B_{k+1} r_k = w_k, \tag{3}$$

where B_{k+1} is an Hessian approximation of $f(x)$ at x_{k+1}. r_k and w_k have the forms of $r_k = s_k - \psi_k s_{k-1}$ and $w_k = y_k - \psi_k y_{k-1}$, where ψ_k is a scaling number. At the

$(k+2)$-th iteration, let $B_{k+1} = \alpha_{k+1}^{-1} I$ satisfy the certain multi-step quasi-Newton condition (3), that is, we set α_{k+1} in (2) as

$$\alpha_{k+1} = \arg\min_{\alpha} \|\frac{1}{\alpha} r_k - w_k\|_2 = \frac{r_k^T w_k}{r_k^T r_k} = \frac{(s_k - \psi_k s_{k-1})^T (y_k - \psi_k y_{k-1})}{(s_k - \psi_k s_{k-1})^T (s_k - \psi_k s_{k-1})}. \tag{4}$$

Since $r_k = \alpha_k^{-1} g_k - \psi_k \alpha_{k-1}^{-1} g_{k-1}$ and $w_k = A r_k$, α_{k+1} keeps invariant when scaling r_k. Define $\hat{r}_k = g_k - \gamma_k g_{k-1}$ and $\hat{w}_k = A\hat{r}_k$, γ_k is a bounded positive number. Then, an equivalent form of (4) is

$$\alpha_{k+1} = \frac{\hat{r}_k^T A \hat{r}_k}{\hat{r}_k^T \hat{r}_k} = \frac{(g_k - \gamma_k g_{k-1})^T A (g_k - \gamma_k g_{k-1})}{(g_k - \gamma_k g_{k-1})^T (g_k - \gamma_k g_{k-1})}. \tag{5}$$

We take α_{k+1} in (4) or (5), denoted by $\alpha_{k+1}^{\mathrm{MBB}}$, as a new stepsize for the gradient method. The resulting method is simply called "MBB" gradient method in this paper. As a special case of $\gamma_k = 0$ for all k, our MBB gradient method is reduced to the original BB method.

We remark that another stepsize involved the curvature pairs $\{(s_{k-i}, y_{k-i}), i = 0, 1, \ldots, m-1\}$ was given by Dai and Fletcher [12] as follows:

$$\alpha_{k+1}^{\mathrm{DF}} = \frac{\sum_{i=0}^{m-1} s_{k-i}^T y_{k-i}}{\sum_{i=0}^{m-1} s_{k-i}^T s_{k-i}}.$$

Though it was initially provided for solving singly linearly constrained quadratic programs, $\alpha_{k+1}^{\mathrm{DF}}$ will reduce to the original BB formula α_k^{BB} when $m = 1$ and also can be used to solve strictly convex quadratic minimization problem. In Sect. 4, we will show that the gradient algorithm with our new stepsize $\alpha_{k+1}^{\mathrm{MBB}}$ can also outperform that with $\alpha_{k+1}^{\mathrm{DF}}$ of Dai and Fletcher.

3 Convergence Analysis

We consider the global convergence of the MBB method. Our analysis follows that of Raydan [5] for the BB gradient method.

Let x_* be the unique minimizer of the problem (1). Denote $e_k = x_k - x_*$ and $\hat{e}_k = e_k - \gamma_k e_{k-1}$. Then,

$$e_{k+1} = \frac{1}{\alpha_k}(\alpha_k I - A) e_k \quad \text{and} \quad \alpha_{k+1} = \frac{\hat{e}_k^T A^3 \hat{e}_k}{\hat{e}_k^T A^2 \hat{e}_k}. \tag{6}$$

Assume that A has the eigenvalues $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ with the associated orthonormal eigenvectors $v_1, v_2, \ldots, v_n$. Easily we have the relations:

$$e_{k+1} = \sum_{i=1}^{n} d_i^{k+1} v_i, \quad d_i^{k+1} = \left(1 - \frac{\lambda_i}{\alpha_k}\right) d_i^k = \prod_{j=1}^{k} \left(1 - \frac{\lambda_i}{\alpha_j}\right) d_i^1,$$
$$\hat{e}_k = \sum_{i=1}^{n} \hat{d}_i^k v_i, \quad \hat{d}_i^k = d_i^k - \gamma_k d_i^{k-1}. \tag{7}$$

Before giving the convergence theorem for the MBB gradient method, we provide two lemmas which are needed in the later analysis.

Lemma 3.1 [3,5] *The sequence $\{d_1^k\}$ converges to zero Q-linearly with convergence factor $c = 1 - \lambda_1/\lambda_n$. For all $i = 1, \ldots, n$ and $k \geq 1$, $(d_i^{k+1})^2 \leq \delta^2 (d_i^k)^2$, where $\delta = \lambda_n/\lambda_1 - 1$.*

Lemma 3.2 *If the sequences $\{d_1^k\}$, $\{d_2^k\}$, ..., $\{d_l^k\}$ all converge to zero for a fixed integer l, $1 \leq l < n$. Then $\lim_{k\to\infty} \inf |d_{l+1}^k| = 0$ and $\lim_{k\to\infty} \inf |\hat{d}_{l+1}^k| = 0$.*

Proof We prove the lemma by contradiction. Suppose that there exists a constant $\epsilon > 0$ such that for all k,

$$\left(d_{l+1}^k\right)^2 \lambda_{l+1}^2 \geq \epsilon \quad \text{and} \quad \left(\hat{d}_{l+1}^k\right)^2 \lambda_{l+1}^2 \geq \epsilon.$$

By (7) and Lemma 3.1, we know that

$$\left(\hat{d}_i^k\right)^2 = \left(d_i^k\right)^2 + \gamma_k^2 \left(d_i^{k-1}\right)^2 - 2\gamma_k d_i^k d_i^{k-1} \leq (\delta + \gamma_k)^2 \left(d_i^{k-1}\right)^2, \tag{8}$$

which, together with that the sequences $\{d_i^k\}_{i=1}^l$ all converge to zero, implies that $\{\hat{d}_i^k\}$ converges to zero for all $i = 1, \ldots, l$. Then, there exists $\hat{k}$ sufficiently large such that $\sum_{i=1}^{l} (\hat{d}_i^k)^2 \lambda_i^2 \leq \frac{1}{2}\epsilon$ holds for all $k \geq \hat{k}$. By orthogonality of $\{v_i\}$ and the later half part of hypothesis, we have

$$\alpha_{k+1} = \frac{\sum_{i=1}^{n} \left(\hat{d}_i^k\right)^2 \lambda_i^3}{\sum_{i=1}^{n} \left(\hat{d}_i^k\right)^2 \lambda_i^2} \geq \frac{\left(\sum_{l+1}^{n} \left(\hat{d}_i^k\right)^2 \lambda_i^2\right) \lambda_{l+1}}{\frac{1}{2}\epsilon + \sum_{l+1}^{n} \left(\hat{d}_i^k\right)^2 \lambda_i^2}$$
$$\geq \frac{\left(\hat{d}_{l+1}^k\right)^2 \lambda_{l+1}^2}{\frac{1}{2}\epsilon + \left(\hat{d}_{l+1}^k\right)^2 \lambda_{l+1}^2} \lambda_{l+1} \geq \frac{2}{3} \lambda_{l+1},$$

which indicates that

$$|d_{l+1}^{k+1}| = |1 - \frac{\lambda_{l+1}}{\alpha_{k+1}}||d_{l+1}^{k}| \leq \max\left\{\frac{1}{2}, c\right\}|d_{l+1}^{k}| < |d_{l+1}^{k}| \text{ for all } k \geq \hat{k} + 1.$$

This contradicts the assumption, and therefore, the lemma is proved. □

Using Lemmas 3.1 and 3.2, we have the following convergence theorem.

Theorem 3.1 *Let f(x) be a strictly convex quadratic function, $\{x_k\}$ be the sequence generated by the MBB method and x_* the unique minimizer of $f(x)$. Then, either $x_k = x_*$ for some finite k, or the sequence $\{x_k\}$ converges to x_*.*

Proof It suffices to prove $\lim_{k\to\infty} e_k = 0$ when $e_k = x_k - x^* \neq 0$ for all k. To this end, we only show $\lim_{k\to\infty} d_i^k = 0$ for $i = 1, 2, \ldots, n$, where d_i^k are defined as those in (7). We do this work by induction.

Lemma 3.1 shows $\{d_1^k\}$ converges to zero. Now assume that $\{d_1^k\}$, ..., $\{d_{p-1}^k\}$ all tend to zero and so do $\{\hat{d}_1^k\}$, ..., $\{\hat{d}_{p-1}^k\}$ by (8). Then, for any given $\epsilon > 0$, there exists sufficiently large $\hat{k}$ such that $\sum_{i=1}^{p-1}(\hat{d}_i^k)^2\lambda_i^2 \leq \frac{1}{2}\epsilon$ holds for all $k \geq \hat{k}$. Additionally, there exists $\bar{k} \geq \hat{k}$ such that $(\hat{d}_p^{\bar{k}})^2\lambda_p^2 < \epsilon$ holds by Lemma 3.2.

Let $k_0 > \bar{k}$ such that $(\hat{d}_p^{k_0-1})^2\lambda_p^2 < \epsilon$ and $(\hat{d}_p^{k_0})^2\lambda_p^2 \geq \epsilon$. We can prove the following inequalities

$$\frac{2}{3}\lambda_p \leq \alpha_{k+1} \leq \lambda_n, \tag{9}$$

hold for any $k_0 \leq k \leq j - 1$ in the similar manner for Lemma 3.2, where j is the first integer for $(d_p^j)^2\lambda_p^2 < \epsilon$ holding. Thus, by using (9) and (7), we have

$$\left|d_p^{k+2}\right| = \left|1 - \frac{\lambda_p}{\alpha_{k+1}}\right|\left|d_p^k\right| \leq \hat{c}\left|d_p^{k+1}\right| \quad \text{for } k_0 \leq k \leq j - 1.$$

Finally, since $|d_p^{k_0+1}| \leq \delta^2|d_p^{k_0-1}|$, we have

$$(d_p^k)^2 \leq \delta^4(d_p^{k_0-1})^2 \leq \delta^4\frac{\epsilon}{\lambda_p^2}$$

for all k satisfying $k_0 + 1 \leq k \leq j + 1$. Further, we have $(d_p^{k_0})^2 \leq \hat{c}^2(d_p^{k_0-1})^2$ from (7). Hence, $(d_p^k)^2$ is bounded above by a constant multiple of ϵ for all $k \geq k_0 - 1$. This implies $\lim_{k\to\infty} d_p^k = 0$ which completes proof of the theorem. □

4 Numerical Experiments

This section focuses on numerical performance of the MBB method. We stop the iteration when $\|Ax_k - b\|_\infty \leq Tol$ is satisfied. Only the number of iterations are

Table 1 Number of iterations required by MBB method for a real problem

γ_k	1/5	1/4	1/3	1/2	2/3	3/4	$\frac{\alpha_{k-1}}{3\alpha_{k-2}}$	$\frac{\alpha_{k-1}}{2\alpha_{k-2}}$	$\frac{2\alpha_{k-1}}{3\alpha_{k-2}}$	$\frac{3\alpha_{k-1}}{4\alpha_{k-2}}$
IT	5563	7894	12,983	10,188	11,968	11,684	6768	7839	11,898	12,905

Table 2 Numerical comparisons for a real problem

Tol	BB	AS	AM	MBB	SS1	SS2	CG	CBB
10^1	211	225	295	222	259	222	229	352
10^0	386	489	697	351	478	463	486	390
10^{-1}	883	1239	2343	903	1342	1937	501	1058
10^{-2}	1554	2333	6657	1349	2951	3536	501	1778

compared since the calculations in each iteration for different gradient methods are almost the same. In tables below, the number of iterations for each method is the average of five runs from different starting points.

We first compare the MBB method with the original BB method under Tol $= 10^{-6}$ through a real problem [13]. It is a linear system in which matrix A appears frequently in the numerical solution of two-point boundary-value problems. Set $n = 1000$, the number of iterations for BB method is 13,230. Listed in Table 1 are the average number of iterations required by MBB method with different choices of γ_k. MBB method obviously outperforms the original BB method.

We further compare the MBB method with three BB-like methods (the AS, AM and CBB methods), Yuan's stepsize (YM) method [7], the SS1, SS2 methods [4] and conjugate gradient (CG) method. We fix the problem size $n = 800$ and predefine $\gamma_k = 0.2$ in the MBB method, 0.8 and 0.75 for the values of parameter in the SS1 and SS2 methods, respectively. The detailed results are presented in Tables 2 and 3. The testing problems for Table 3 are some random problems generated in the same way of [8]. We skip the results of YM method for the real problem since it requires lot of iterations.

As shown in Table 2, the MBB method outperforms other gradient methods for different tolerances Tol. If Tol $= 10^1, 10^0$, namely if only a solution with a low precision is required, the MBB, BB, AS and SS2 methods are competitive with the CG method. The CG method is clearly a winner when finding a solution with a higher precision. The same conclusions can be achieved from Table 3. Furthermore, the MBB method is better than YM method. If the problem is well conditioned ($\kappa(A) \leq 100$), the MBB method can be competitive with the CG method even for finding a solution with a higher precision.

By setting the tolerance Tol $= 10^{-6}$ and the problem size $n = 800$, Table 4 presents the numerical results when α^{DF} is employed for some random problems which were also generated in the way of [8]. Obviously, our stepsize is superior to the one defined by Dai and Fletcher [12].

Table 3 Numerical comparisons I for random problems

$\kappa(A)$	Tol	BB	AS	AM	MBB	YM	SS1	SS2	CG	CBB
10	10^{-2}	18	22	23	18	21	23	19	18	22
	10^{-4}	26	31	35	26	31	32	26	25	30
	10^{-6}	36	38	49	34	40	41	34	32	38
100	10^{-2}	69	72	87	65	75	108	103	53	74
	10^{-4}	97	96	118	84	101	150	132	76	96
	10^{-6}	138	128	154	108	120	189	163	98	132
1000	10^{-2}	215	245	333	189	237	271	307	163	208
	10^{-4}	374	324	583	315	463	367	458	229	314
	10^{-6}	472	488	671	428	969	676	502	298	376
10000	10^{-2}	1036	805	1667	738	1059	1591	4663	490	812
	10^{-4}	1266	1134	2359	942	1596	2130	8743	697	1088
	10^{-6}	1444	1901	3517	1105	2160	2550	12823	888	1598

Table 4 Numerical comparisons II for random problems

$\kappa(A)$	MBB	DF								
		$m=2$	3	4	5	6	7	8	9	10
10	34	36	35	35	37	41	42	47	49	49
100	115	124	122	122	120	126	124	125	130	141
1000	377	408	397	428	396	392	394	408	414	404
10000	1272	1556	1388	1291	1235	1277	1323	1332	1390	1416

5 Conclusions

In this note, we proposed a new modified BB gradient method by using the multi-step quasi-Newton equation. The new stepsize combines the two gradient directions which are used in the SD and BB methods, respectively. The modified method is shown to be globally convergent. Numerical experiments show it outperforms some existing gradient methods, such as the AS, AM and CBB methods. It is worth mentioning that this stepsize can be used in those algorithms which use the BB stepsize to solve the unconstrained optimization problem.

Acknowledgements The authors are grateful to the anonymous referees and editor for their useful comments. This research was supported by the National Science Foundation of China under Grant Nos. 11171371 and 11571004.

References

1. Barzilai, J., Borwein, J.: Two-point step size gradient methods. IMA J. Numer. Anal. **8**(1), 141–148 (1988)

2. Raydan, M., Svaiter, B.: Relaxed steepest descent and Cauchy–Barzilai–Borwein method. Comput. Optim. Appl. **21**(2), 155–167 (2002)
3. Dai, Y., Liao, L.: R-linear convergence of the Barzilai and Borwein gradient method. IMA J. Numer. Anal. **22**(1), 1–10 (2002)
4. Dai, Y., Yuan, Y.: Alternate minimization gradient method. IMA J. Numer. Anal. **23**, 377–393 (2003)
5. Raydan, M.: On the Barzilai and Borwein choice of steplength for the gradient method. IMA J. Numer. Anal. **13**(3), 321–326 (1993)
6. Dai, Y.: Alternate step gradient method. Optimization **52**(4–5), 395–415 (2003)
7. Yuan, Y.: A new stepsize for the steepest descent method. J. Comput. Math. **24**(2), 149–156 (2006)
8. De Asmundis, R., Di Serafino, D., Riccio, F., Toraldo, G.: On spectral properties of steepest descent methods. IMA J. Numer. Anal. **33**(4), 1416–1435 (2013)
9. Gonzaga, C., Schneider, R.: On the steepest descent algorithm for quadratic functions. Comput. Optim. Appl. **63**(2), 523–542 (2016)
10. Ford, J., Moghrabi, I.: Alternative parameter choices for multi-step quasi-Newton methods. Optim. Methods Softw. **2**(3–4), 357–370 (1993)
11. Ford, J., Moghrabi, I.: Multi-step quasi-Newton methods for optimization. J. Comput. Appl. Math. **50**(93), 305–323 (1994)
12. Dai, Y., Fletcher, R.: New algorithms for singly linearly constrained quadratic programs subject to lower and upper bounds. Math. Progr. **106**(3), 403–421 (2006)
13. Friedlander, A., Martínez, J., Molina, B., Raydan, M.: Gradient method with retards and generalizations. SIAM J. Numer. Anal. **36**(1), 275–289 (1999)