ICIC Express Letters ICIC International ©2016 ISSN 1881-803X
Volume **10**, Number **10**, October **2016** pp. **2383–2389**

A SUFFICIENT DESCENT BARZILAI-BORWEIN GRADIENT METHOD FOR UNCONSTRAINED OPTIMIZATION WITH APPLICATIONS IN IMPULSE NOISE REMOVAL

Min Sun[1], Jing Liu[2,3] and Guangde Liu[1]

[1]School of Mathematics and Statistics
Zaozhuang University
No. 1, Beian Road, Zaozhuang 277160, P. R. China
ziyouxiaodou@163.com; liuguangde2012@126.com

[2]School of Economics and Management
Tongji University
No. 1239, Siping Street, Shanghai 200092, P. R. China
liujingta@126.com

[3]School of Data Sciences
Zhejiang University of Finance and Economics
No. 18, Xueyuan Street, Hangzhou 310018, P. R. China

Received February 2016; accepted May 2016

Abstract. *In this paper, we propose a modified Barzilai-Borwein (BB) gradient method for unconstrained optimization problems. The resulting algorithm has sufficient descent property, which is independent of any line search. Furthermore, under mild conditions, we establish the global convergence of the new method for smooth unconstrained optimization. Some numerical results about impulse noise removal in image processing are given, which indicate that the new method is efficient even for high noise ratio.*
Keywords: Barzilai-Borwein gradient method, Global convergence, Two-phase method, Image processing, PSNR

1. **Introduction.** In this paper, we consider the following unconstrained optimization problem:

$$\min f(x),\ x \in \mathcal{R}^n, \tag{1}$$

where $f : \mathcal{R}^n \to \mathcal{R}$ is a smooth function, bounded from below, whose gradient $\nabla f(x)$ is denoted by $g(x)$, or g is Lipschitz continuous. We are interested in the case that the number n is large. The unconstrained optimization problem (1) is the mathematical model of many scientific problems arising from game theory, transportation, economic equilibrium, image restoration and compressive sensing, etc., see [1-3]. Moreover, the nonlinear equation

$$F(x) = 0, \tag{2}$$

where $F : \mathcal{R}^n \to \mathcal{R}^n$, can also be transformed into (1) by setting $f(x) = \|F(x)\|^2/2$.

In the past few decades, many efficient iterative methods have been proposed to solve (1). Different from quasi-Newton method, conjugate gradient method and Barzilai-Borwein gradient method do not need to compute and store any matrix; therefore, they are attractive, especially when the dimension n is large, and have received a great deal of attention, see [1-7].

In this paper, we intend to study the Barzilai-Borwein gradient method. Now, let us recall this method, whose iterative scheme is

$$x_{k+1} = x_k - \frac{1}{\alpha_k} g_k,\ k = 0, 1, 2, \ldots, \tag{3}$$

where x_k is the kth iterative point, $g_k = g(x_k)$, and α_k is called BB steplength defined by

$$\alpha_k^{\mathrm{I}} = \frac{s_{k-1}^\top y_{k-1}}{\|s_{k-1}\|^2}, \text{ or } \alpha_k^{\mathrm{II}} = \frac{\|y_{k-1}\|^2}{s_{k-1}^\top y_{k-1}},$$

in which $s_{k-1} = x_k - x_{k-1}$ and $y_{k-1} = g_k - g_{k-1}$. Though the search direction of BB gradient method is always the negative gradient, the steplength is obviously different from the choice of the steepest descent method. Due to its high efficiency, a large amount of theoretical results have appeared in the literature. Barzilai and Borwein [1] have proved that BB gradient method converges globally for the convex quadratic optimization with two variables. Then, Raydan [2] has extended the above result to the convex quadratic optimization with any number of variables by incorporating a nonmonotone line search. Recently, the BB gradient method has been successfully generalized to solve the (constrained) nonlinear monotone equations, and the image restoration problems, see [4].

Note that the modified BB gradient methods in [2,3] have the sufficient descent property, that is

$$g_k^\top d_k \le -c\|g_k\|^2,$$

where $c > 0$. This property makes the proof of the global convergence of the corresponding method easy and standard. In addition, the iterative methods with this property are often more efficient than those without this property. However, the modified BB gradient methods in [2,3] can only be used to solve the (constrained) nonlinear monotone equations, and the modified BB gradient method in [4] uses the inverse of α_k^{I} to design the descent direction, which is not reasonable (See the motivations in Section 2). Therefore, in this paper, based on the modified BB gradient methods in [2-4], we shall design a new BB gradient method for (1), which has the sufficient descent property under standard assumptions, and also tends to the classical BB gradient method if the constant $r > 0$ (r will be given later) is sufficiently small.

The rest of the paper is organized as follows. In Section 2, some motivations are introduced and the detailed sufficient descent BB gradient method is outlined. In Section 3, the global convergence of the new method is proved under some standard conditions. In Section 4, numerical results and comparison about the impulse noise removal in image processing are reported to show the efficiency of the new method. Some concluding remarks are summarized in the final section.

2. **Motivations and Algorithm.** For the nonlinear Equation (2), Zhang and Zhou [6] introduced some interesting modifications of BB gradient method such that the modified method has sufficient descent property. More specially, the search direction in [6] is

$$d_k = -\frac{1}{\theta_k^{\mathrm{ZZ}}} F(x_k), \tag{4}$$

where $\theta_1^{\mathrm{ZZ}} = 1$ and $\theta_k^{\mathrm{ZZ}} = s_k^\top y_k / \|s_k\|^2$, if $k \ge 2$. Here $s_k = x_k - x_{k-1}$, $y_k = F(x_k) - F(x_{k-1}) + r s_k$ and $r > 0$ is a constant. By introducing the term $r s_k$ in y_k and under the mapping $F(\cdot)$ is monotone and Lipschitz continuous, Zhang and Zhou [6] proved that the modified BB gradient method possesses the following nice property

$$-\frac{1}{r}\|F(x_k)\|^2 \le F(x_k)^\top d_k \le -\frac{1}{L+r}\|F(x_k)\|^2,$$

where $L > 0$ is the Lipschitz constant of the mapping $F(\cdot)$. Obviously, the right inequality indicates that the search direction d_k satisfies the sufficient descent property. The above inequalities are based on the monotonicity of $F(\cdot)$, and $F(\cdot)$ is equivalent to $g(\cdot)$ in (1). However, the mapping $g(\cdot)$ often does not assume to satisfy the monotone property. Thus, we cannot generalize the modified BB gradient method in [6] to solve (1), though it satisfies the sufficient descent property. As for the method proposed by Liu and Li [4], its search direction is also generated by

$$d_k = -\frac{1}{\theta_k^{\mathrm{LL}}} F(x_k), \tag{5}$$

where $\theta_1^{\mathrm{LL}} = 1$ and $\theta_k^{\mathrm{LL}} = \|s_k\|^2 / s_k^\top y_k$, if $k \geq 2$. Here $s_k = x_k - x_{k-1}$, $y_k = g_k - g_{k-1} + t_k s_k$ and

$$t_k = 1 + \max\left\{ 0, -\frac{s_k^\top (g_k - g_{k-1})}{\|s_k\|^2} \right\}.$$

Obviously, θ_k^{LL} is the inverse of α_k^{I} or θ_k^{ZZ}. It is well-known that α_k^{I} is deduced by using some quasi-Newton property, which can greatly speed up the convergence rate of the gradient method for quadratic functions. The parameter θ_k^{LL} can ensure that the search direction d_k generated by (5) satisfies the sufficient descent property, and this is the main motivation of θ_k^{LL}. However, unlike θ_k^{ZZ}, θ_k^{LL} cannot tend to α_k^{I} except for some special cases. Therefore, in this paper, we intend to design a modified BB gradient method which not only has the sufficient descent property but also can reduce to the classical BB gradient method in some cases.

Throughout this paper, we make the following assumptions.

Assumptions.

(A1) The level set $\Omega = \{x \in \mathcal{R}^n | f(x) \leq f(x_0)\}$ is bounded.

(A2) In some neighborhood N of Ω, the objective function $f(\cdot)$ is continuously differentiable and its gradient $g(\cdot)$ is Lipschitz continuous, i.e., there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \ \forall x, y \in N. \tag{6}$$

We now formally state the steps of the modified BB gradient method for solving problem (1) as follows.

Algorithm 2.1

Step 0. Given an initial point $x_0 \in \mathcal{R}^n$, and choose four constants $r > 0$, $\rho \in (0,1)$, $\delta \in (0,1)$, $\varepsilon > 0$. Set $k := 0$.

Step 1. If $\|g_k\| \leq \varepsilon$, stop.

Step 2. Compute d_k by

$$d_k = \begin{cases} -g_k, & \text{if } k = 0, \\ -\dfrac{1}{\theta_k^{\mathrm{SS}}} g_k, & \text{if } k \geq 1, \end{cases} \tag{7}$$

where $\theta_k^{\mathrm{SS}} = s_k^\top y_k / \|s_k\|^2$ is similar to α_k^{I}, and $s_k = x_k - x_{k-1}$, but y_k is defined by

$$y_k = (g_k - g_{k-1}) + t_k s_k,$$

with

$$t_k = r + \max\left\{ 0, -\frac{s_k^\top (g_k - g_{k-1})}{\|s_k\|^2} \right\}.$$

Step 3. Determine stepsize α_k by the following Armijo line search, $\alpha_k = \rho^{l_k}$ with l_k being the smallest nonnegative integer l such that

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k^2 g_k^\top d_k. \tag{8}$$

Step 4. Let $x_{k+1} = x_k + \alpha_k d_k$. Set $k := k + 1$ and go to Step 1.

Now we give some remarks about Algorithm 2.1.

Remark 2.1. *Obviously, we have the following inequality*

$$s_k^\top y_k = s_k^\top (g_k - g_{k-1}) + t_k \|s_k\|^2 \geq r\|s_k\|^2. \tag{9}$$

On the other hand, if $s_k^\top (g_k - g_{k-1}) \geq 0$, then by (6), we have

$$s_k^\top y_k = s_k^\top (g_k - g_{k-1}) + r\|s_k\|^2 \leq (L + r)\|s_k\|^2.$$

If $s_k^\top (g_k - g_{k-1}) < 0$, we have

$$s_k^\top y_k = s_k^\top (g_k - g_{k-1}) - \frac{s_k^\top (g_k - g_{k-1})}{\|s_k\|^2} \|s_k\|^2 + r\|s_k\|^2 = r\|s_k\|^2 \le (L+r)\|s_k\|^2.$$

By the above two inequalities and (9), we have

$$r \le \theta_k^{\rm SS} \le r + L. \tag{10}$$

This together with (7) implies the following two important inequalities

$$-\frac{1}{r}\|g_k\|^2 \le g_k^\top d_k \le -\frac{1}{L+r}\|g_k\|^2. \tag{11}$$

Obviously, the right inequality indicates that the search direction d_k generated by Algorithm 2.1 satisfies the sufficient descent property. In addition, from (10), we can also get another two important inequalities

$$\frac{1}{L+r}\|g_k\| \le \|d_k\| \le \frac{1}{r}\|g_k\|. \tag{12}$$

Remark 2.2. *Obviously, if $r = 0$ and $s_k^\top (g_k - g_{k-1}) \ge 0$, then $\theta_k^{\rm SS}$ reduces to $\alpha_k^{\rm I}$. Therefore, the descent direction d_k defined by (7) is identical to the direction of the classical BB gradient method.*

3. **Global Convergence.** In this section, we shall prove the global convergence of Algorithm 2.1 under Assumptions (A1) and (A2). First, we need to prove that the line search scheme (8) will terminate in a finite number of steps for every $k \ge 0$ to ensure that Algorithm 2.1 is well-defined.

Lemma 3.1. *Algorithm 2.1 is well-defined, namely, there exists a nonnegative integer l_k satisfying the line search scheme (8) for all k.*

Proof: The proof is standard, and thus is omitted.

Lemma 3.2. *There exists a constant $\varrho > 0$ such that the stepsize α_k involved in Step 3 of Algorithm 2.1 satisfies*

$$\alpha_k \ge \min\left\{1, \varrho\frac{\|g_k\|^2}{\|d_k\|^2}\right\}. \tag{13}$$

Proof: If $\alpha_k \ne 1$, then by the acceptance rule of stepsize α_k in Algorithm 2.1, we have $\alpha'_k = \alpha_k/\rho$ does not satisfy (8), namely $f(x_k + \alpha'_k d_k) > f(x_k) + \delta(\alpha'_k)^2 g_k^\top d_k$. This and (11), (12) give

$$f(x_k + \alpha'_k d_k) > f(x_k) - \frac{\delta(\alpha'_k)^2}{r}\|g_k\|^2 \ge f(x_k) - \frac{\delta(L+r)(\alpha'_k)^2}{r}\|d_k\|^2. \tag{14}$$

On the other hand, from the mean value theorem, (11) and (A2), there exists a constant $\theta_k \in (0,1)$, such that

$$\begin{aligned} f(x_k + \alpha'_k d_k) - f(x_k) &= \alpha'_k g(x_k + \theta_k \alpha'_k d_k)^\top d_k \\ &= \alpha'_k g_k^\top d_k + \alpha'_k (g(x_k + \theta_k\alpha'_k d_k) - g_k)^\top d_k \\ &\le -\frac{\alpha'_k}{L+r}\|g_k\|^2 + (\alpha'_k)^2 L\|d_k\|^2, \end{aligned}$$

which together with (14) shows that (13) holds with $\varrho = \frac{r}{(L+r)(rL+\delta(L+r))}$. This completes the proof.

Lemma 3.3. *Let the sequence $\{d_k\}$ be generated by Algorithm 2.1. Then we have*

$$\lim_{k\to\infty} \alpha_k\|d_k\| = 0. \tag{15}$$

Proof: Elementary.

Now, using the above lemmas, we can obtain the global convergence of Algorithm 2.1 as follows.

Theorem 3.1. *Suppose that (A1) and (A2) hold. Then the sequence $\{x_k\}$ generated by Algorithm 2.1 is global convergence in the sense that*

$$\lim_{k\to\infty} \inf \|g_k\| = 0.$$

Proof: By contradiction, we suppose that the conclusion is not right. Then there exists a constant $\varepsilon > 0$ such that $\|g_k\| \geq \varepsilon$, $\forall k \geq 0$. This together with (12) implies that

$$\|d_k\| \geq \frac{\varepsilon}{L+r}, \ \forall k \geq 0.$$

Combining the above inequality with (15) gives

$$\lim_{k\to\infty} \alpha_k = 0. \tag{16}$$

On the other hand, (A1) and (A2) imply there is a constant $\gamma > 0$ such that $\|g_k\| \leq \gamma$, $\forall k \geq 0$. Then, this and (12) give that

$$\|d_k\| \leq \gamma/r. \tag{17}$$

Thus, from (13) and (17), we get $\alpha_k \geq \min\left\{1, \frac{\varrho\varepsilon^2 r^2}{\gamma^2}\right\}$, $\forall k \geq 0$. This yields a contradiction with (16). The proof is then completed.

4. **Numerical Results.** In this section, we use Algorithm 2.1 to remove the salt-and-pepper impulse noise in the second phase of the two-phase method [5], and present some numerical experiments to evaluate its performance. At the same time, we also give some comparisons with the related algorithms, including the PRP conjugate gradient method in [5], denoted by CCM, and the modified BB gradient method in [4], denoted by LL.

Let $X = [x_{i,j}]_{M\times N}$ be the true image with M-by-N pixels. For each $(i,j) \in A := \{1,2,\ldots,M\} \times \{1,2,\ldots,N\}$, let $V_{i,j}$ be the neighborhood of (i,j), i.e., $V_{i,j} = \{(i,j-1),(i,j+1),(i-1,j),(i+1,j)\}$. In addition, let $y_{i,j}$ be the observed pixel value of the image at the position (i,j). Now let us review the efficient two-phase method for restoring images corrupted with high level salt-and-pepper impulse noise. In the first phase, the salt-and-pepper impulse noise is detected by the adaptive median filter (AMF), and let $N \subseteq A$ denote the set of indices of the noise pixels detected in this phase. Then, the second phase is the recovering of the noise pixels by minimizing the following function:

$$\mathcal{G}_\alpha(u) = \sum_{(i,j)\in N} \left\{ \sum_{(m,n)\in V_{i,j}\setminus N} \varphi_\alpha(u_{i,j} - y_{m,n}) + \frac{1}{2} \sum_{(m,n)\in V_{i,j}\cap N} \varphi_\alpha(u_{i,j} - y_{m,n}) \right\}, \tag{18}$$

where α is the regularization parameter, and φ_α is an edge-preserving function, and $u = [u_{i,j}]_{(i,j)\in N}$ is a column vector of length lexicographically. Here c denotes the number of element of N.

We implemented all the algorithms with codes written in Matlab 7.10. The testing is performed on a PC computer with Pentium(R) Dual-Core CPU T4400@2.2GHz, 4GB of memory. The test images are two 256×256 gray level images (Cameraman and Lena) and two 512×512 gray level images (Barbara and Boat).

Throughout the computational experiments, the parameters used in Algorithm 2.1 are chosen as follows: $r = 0.0001$, $\rho = 0.4$, $\delta = 0.2$, and the parameters used in LL are chosen as follows: $\tau = \sqrt{99}/8$, $\rho = 0.9$, $\delta = 0.5$. The CCM does not use any line search, and its stepsize is computed explicitly by $\alpha_k = \delta\frac{g_k^\top d_k}{\|d_k\|^2}$, at each iteration with $\delta = \sqrt{99}/8$. To assess the restoration performance qualitatively, we use the peak signal to noise ratio (PSNR) defined as

$$\mathrm{PSNR} = 10\log_{10} \frac{255^2}{\frac{1}{MN}\sum_{i,j}\left(u^*_{i,j} - x_{i,j}\right)^2},$$

where $u^*_{i,j}$ is the pixel values of the restored image. Furthermore, the stopping criterion of three algorithms are

$$\frac{|\mathcal{G}_\alpha(u_k) - \mathcal{G}_\alpha(u_{k-1})|}{|\mathcal{G}_\alpha(u_k)|} \le 10^{-4}, \text{ or } \frac{\|u_k - u_{k-1}\|}{\|u_k\|} \le 10^{-4}.$$

The edge-preserving function φ_α is defined as

$$\varphi_\alpha(t) = \begin{cases} t^2/(2\alpha), & \text{if } |t| \le \alpha, \\ |t| - \alpha/2, & \text{if } |t| > \alpha, \end{cases}$$

with $\alpha = 10$.

Table 1 lists some detailed numerical results with different noise levels $r = 30\%$, 50%, 70%, and 90%, respectively. All the experiments are repeated 10 times and the average of the ten results are reported. We report the CPU time (in second), the number of iterations (Niter) required for the whole denoising process and the PSNR of the recovered image. From Table 1, we find that: (1) For the same stopping criterion, three algorithms generate almost the same PRSNs, which indicates that the qualities of restored images from three algorithms are similar; (2) Comparing with the other two criteria: CPU and Niter, we find that Algorithm 2.1 is faster than the other two algorithms, since it needs less CPU and less Niter to reach the same accuracy. Overall, Algorithm 2.1 is an efficient method for the subproblem (18) of the two-phase method.

Table 1. Performance of salt-and-pepper denoising via three algorithms

Image	$r(\%)$	CCM			LL			Algorithm 2.1		
		PSNR	CPU	Niter	PSNR	CPU	Niter	PSNR	CPU	Niter
Cameraman	30	30.34	3.44	49.8	30.48	3.15	46.0	30.53	2.98	30.6
	50	27.30	5.48	61.6	27.38	5.80	59.2	27.37	4.10	32.8
	70	24.57	8.28	81.8	24.65	8.04	81.2	24.74	6.48	44.0
	90	21.11	16.68	155.4	21.14	14.75	155.0	21.15	11.18	67.8
Lena	30	33.34	2.67	23.6	33.38	2.37	31.0	33.39	2.54	30.8
	50	30.03	4.06	44.5	30.04	4.01	43.1	30.07	3.19	26.2
	70	27.02	6.06	64.4	27.04	6.46	64.4	27.11	4.39	32.8
	90	22.71	13.47	131.6	22.68	14.09	129.0	22.68	9.75	59.6
Barbara	70	24.57	26.15	55.0	24.56	25.92	53.2	24.58	21.71	30.0
Boat	70	27.90	29.41	63.0	27.90	28.70	59.4	27.90	24.85	36.2

5. **Conclusions.** In this paper, we proposed a modified Barzilain-Borwein gradient method for the unconstrained optimization. The most desirable property is that the direction generated by the new method is always sufficient descent. In the future, we shall further study the BB gradient method and apply it to other domains, such as compressive sensing, and nonlinear regression.

Acknowledgment. The authors would like to express their thanks to the anonymous reviewer for his valuable suggestions. The first author is supported by the Zaozhuang University's Jiaogai Project titled "Construction and Application of Statistics Specialty Experimental Teaching System", and the second author is supported by the National Science Foundation of China (71371139) and the Key Project of National Natural Science Foundation of China (71532015).

REFERENCES

[1] J. Barzilai and J. M. Borwein, Two-point step size gradient methods, *IMA Journal of Numerical Analysis*, vol.8, no.1, pp.141-148, 1988.
[2] M. Raydan, The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem, *SIAM Journal on Optimization*, vol.7, pp.2-33, 1997.
[3] Z. S. Yu, J. Lin, J. Sun, Y. H. Xiao, L. Y. Liu and Z. H. Li, Spectral gradient projection method for monotone nonlinear equations with convex constraints, *Applied Numerical Mathematics*, vol.59, pp.2416-2423, 2009.
[4] J. K. Liu and S. J. Li, Spectral gradient method for impulse noise removal, *Optimization Letters*, vol.9, no.7, pp.1341-1351, 2015.
[5] J. F. Cai, R. H. Chan and C. D. Fiore, Minimization of a detail-preserving regularization functional for impulse noise removal, *Journal of Mathematical Imaging and Vision*, vol.29, pp.79-91, 2007.
[6] L. Zhang and W. J. Zhou, Spectral gradient projection method for solving nonlinear monotone equations, *Journal of Computational and Applied Mathematics*, vol.196, no.2, pp.478-484, 2006.
[7] R. Fletcher, On the Barzilai-Borwein method, *Applied Optimization*, vol.96, pp.235-256, 2001.