

CONVERGENCE OF THE BARZILAI-BORWEIN METHOD FOR SOLVING SLIGHTLY UNSYMMETRIC LINEAR SYSTEMS

Bayda Ghanim Fathi

Dept. of Mathematics, Faculty of Science, University of Zakho, Kurdistan Region-Iraq.

(Accepted for publication: May 18, 2015)

Abstract:

Due to its simplicity and numerical efficiency, the Barzilai and Borwein (BB) gradient method has received numerous attentions in different scientific fields. In this paper, the sufficient condition for convergence of the BB method when the coefficient matrix of linear algebraic equations is slightly unsymmetric with positive definite symmetric part is presented.

Keywords: Unsymmetric linear algebraic equations, Barzilai and Borwein gradient method, symmetric and skew-symmetric matrices, eigenvalues, condition number.

1. Introduction

A significant development that has completely changed our perspectives on the effectiveness of gradient methods is due to Barzilai and Borwein (Barzilai and Borwein, 1988). They proposed two choices of step size the gradient method. The computational cost and typical behaviour of the algorithm for both choices of step size are quite similar. Their method aimed to accelerate the convergence of the steepest descent (SD) method.

The main idea of Barzilai and Borwein's approach is to use the information in the previous iteration to decide the step size in the current iteration. The Barzilai-Borwein (BB) method requires few storage locations and inexpensive computations. Therefore, several authors have paid attention to the BB method. Their method is strongly related to Quasi-Newton (QN) algorithms (Dennis and Schnabel, 1983) and (Fletcher, 2000).

It is known that the (BB) method (Barzilai and Borwein, 1988) converge when this method is applied to solve linear systems of the form

$$Ax = b,$$

where A is symmetric and positive definite. For some finite difference discretizations of elliptic problems, one gets positive definite matrices that are almost symmetric. Practically, the BB method works for these matrices. However, the convergence of this method is not guaranteed theoretically.

For quadratics that the BB method has been shown to converge (Raydan, 1993) and its convergence is R-linear (Dai and Liao, 2002). For the case of $n = 2$, the method is R-superlinearly convergent (Barzilai and Borwein, 1988). For more details on the Barzilai and Borwein method see (Raydan, 1991) and (Dai and R. Fletcher, 2005). The BB algorithm for quadratic case is summarized in Algorithm 1.

Algorithm 1 Barzilai and Borwein (BB)

1. Given $x_0 \in \mathbb{R}^n$, choose arbitrary $\gamma_0 > 0$, for instance, $\gamma_k = \frac{(g_0, g_0)}{(g_0, Ag_0)}$, (or $\gamma_k = \frac{(Ag_0, g_0)}{(Ag_0, Ag_0)}$)
2. For $k=0, 1, 2, \dots$ (until convergence) do
3. $g_0 = Ax_0 - b$
4. $x_{k+1} = x_k - \gamma_k g_k$
5. $g_{k+1} = g_k - \gamma_k A g_k$
6. $\gamma_k = \frac{(g_{k-1}, g_{k-1})}{(g_{k-1}, Ag_{k-1})}$, (or $\gamma_k = \frac{(Ag_{k-1}, g_{k-1})}{(Ag_{k-1}, Ag_{k-1})}$)
7. End for

This paper presents the proof of convergence of the BB method when the Euclidian norm (l^2) of the unsymmetric part of a positive definite matrix is less than some value related to the smallest and the largest eigenvalues of the symmetric part of the given matrix. This means that the restriction of A be symmetric is removed, and required only that its symmetric part $(A + A^T)/2$ be positive definite.

Let us begin with some notations:

- 1- The Euclidean inner product of any two vectors $u, v \in R^n$ is defined by

$$(u, v) = u^T v,$$

and the induced Euclidean norm (or 2-norm) of u is

$$\|u\| = \sqrt{(u, u)}.$$

- 2- Let A be an $n \times n$ matrix and the associated matrix norm is given by

$$A := \text{Sup}_{\|u\|=1} \|Au\|.$$

It is known that A can be represented as

$$A = A_0 + A_1,$$

where A_0 and A_1 the symmetric and skew-symmetric of a matrix A , i.e.

$$A_0 = (A + A^T)/2, \quad A_1 = (A - A^T)/2.$$

- 4- Let A be positive definite, then the symmetric part A_0 is also positive definite. Hence A and A_0 are invertible and the eigenvalues of A_0 are all positive real numbers. If

$\lambda_{min}, \dots, \lambda_{max}$ are the eigenvalues of A_0 such that

$$0 < \lambda_{min} \leq \dots \leq \lambda_{max}$$

then $\lambda_{min}^{-1}, \dots, \lambda_{max}^{-1}$ are the eigenvalues of A_0^{-1} .

- 5- The condition number of A_0 is defined to be

$$\rho := \|A_0\| \|A_0^{-1}\|.$$

Since A_0 is symmetric,

$$\rho = \frac{\lambda_{max}}{\lambda_{min}} \geq 1.$$

- 6- The Rayleigh-quotient of any vector $u \in R^n$ with respect to A is defined by

$$r(u) = \frac{(u, Au)}{(u, u)}.$$

The lemmas that will be used frequently in the proof of the main theorem are follows:

Lemma 1. For any vector u ,

$$\lambda_{min} \|u\|^2 \leq (u, A_0 u) \leq \lambda_{max} \|u\|^2.$$

Proof. For a symmetric matrix positive definite matrix A_0 , the Rayleigh quotients are bounded by the smallest and the largest eigenvalues of the matrix. Thus, the following relation holds for any u

$$\lambda_{min} \leq \frac{(u, A_0 u)}{\|u\|^2} \leq \lambda_{max}.$$

Hence the proof is complete.

Lemma 2. For any vector u ,

$$\lambda_{max}^{-1} \|u\|^2 \leq (u, A_0^{-1} u) \leq \lambda_{min}^{-1} \|u\|^2.$$

Proof. Similar to proof Lemma 1.

Lemma 3. For any vector u ,

$$(u, Au) = (u, A_0 u) \quad \text{and} \quad (u, A_1 u) = 0$$

Proof. We have

$$(u, Au) = (u, A_0 u) + (u, A_1 u)$$

Since $u^T A_1 u$ is a real number,

$$u^T A_1 u = (u^T A_1 u)^T = u^T A_1^T u = -u^T A_1 u.$$

Therefore $(u, A_1 u) = 0$ and $(u, Au) = (u, A_0 u)$.

Lemma 4 (Cauchy-Schwarz inequality for a positive definite matrix).

If A_0 is a positive definite matrix, then for any u and v

$$|(u, A_0 v)| \leq \sqrt{(u, A_0 u)} \sqrt{(v, A_0 v)}.$$

2. Convergence of the BB Method

In this section, the proof of convergence of the BB method applied to the slightly unsymmetric linear system $Ax = b$ with the positive definite symmetric part is given.

Theorem 1. If

$$\|A_1\| < \lambda_{\min} \sqrt{\rho^{-1}} (-1 + \sqrt{1 + \rho^{-1}}),$$

then the BB method, defined by

$$g_{k+1} = g_k - \gamma_k A g_k \quad (2.1)$$

where

$$\gamma_k = \frac{(g_{k-1}, g_{k-1})}{(g_{k-1}, A g_{k-1})}, \quad (\text{or } \gamma_k = \frac{(A g_{k-1}, g_{k-1})}{(A g_{k-1}, A g_{k-1})}), \quad (2.2)$$

converges.

Proof. Using (2.1), (2.2), Lemma (3), Lemma (1) and since γ_k is the Rayleigh quotient of the symmetric positive definite matrix A_0 (i.e. $0 < \lambda_{\min} \leq \gamma_k \leq \lambda_{\max}$ for all k), we have

$$\begin{aligned} (g_k, A_0^{-1} g_k) &= (g_{k-1} - \gamma_k A g_{k-1}, A_0^{-1} g_{k-1} - \gamma_k A_0^{-1} A g_{k-1}) \\ &= (g_{k-1}, A_0^{-1} g_{k-1}) - 2\gamma_k (A g_{k-1}, A_0^{-1} g_{k-1}) + \gamma_k^2 (A g_{k-1}, A_0^{-1} A g_{k-1}) \\ &= (g_{k-1}, A_0^{-1} g_{k-1}) - 2\gamma_k [(g_{k-1}, g_{k-1}) + (A_1 g_{k-1}, A_0^{-1} g_{k-1})] \\ &\quad + \gamma_k^2 [(A g_{k-1}, g_{k-1}) + (A g_{k-1}, A_0^{-1} A_1 g_{k-1})] \\ &= (g_{k-1}, A_0^{-1} g_{k-1}) - \gamma_k (g_{k-1}, g_{k-1}) - 2\gamma_k (A_1 g_{k-1}, A_0^{-1} g_{k-1}) \\ &\quad + \gamma_k^2 (A g_{k-1}, A_0^{-1} A_1 g_{k-1}) \\ &= (g_k, A_0^{-1} g_k) - \gamma_k (g_k, g_k) - 2\gamma_k (A_1 g_k, A_0^{-1} g_k) \\ &\quad + \gamma_k^2 [(g_k, A_1 g_k) + (A_1 g_k, A_0^{-1} A_1 g_k)] \end{aligned} \quad (2.3)$$

$$\begin{aligned} &= (g_{k-1}, A_0^{-1} g_{k-1}) - \gamma_k (g_{k-1}, g_{k-1}) - 2\gamma_k (A_1 g_{k-1}, A_0^{-1} g_{k-1}) \\ &\quad + \gamma_k^2 (A_1 g_{k-1}, A_0^{-1} A_1 g_{k-1}). \end{aligned} \quad (2.4)$$

Assuming $\delta := \|A_1\|$ and $c_{k-1} := (g_{k-1}, A_0^{-1} g_{k-1})$ for any k , using Lemma 2 and since $\lambda_{\max}^{-1} \leq \gamma_k \leq \lambda_{\min}^{-1}$, we obtain

$$\gamma_k (g_{k-1}, g_{k-1}) \geq \lambda_{\max}^{-1} \lambda_{\min} c_{k-1} = \rho^{-1} c_{k-1}. \quad (2.5)$$

By using Lemma 2 twice, we have

$$\begin{aligned} |(A_1 g_{k-1}, A_0^{-1} A_1 g_{k-1})| &\leq \lambda_{\min}^{-1} \|A_1 g_{k-1}\|^2 \leq \lambda_{\min}^{-1} \delta^2 \|g_{k-1}\|^2 \\ &\leq \lambda_{\min}^{-1} \delta^2 \lambda_{\max} (g_{k-1}, A_0^{-1} g_{k-1}) \leq \rho \delta^2 c_{k-1}. \end{aligned} \quad (2.6)$$

Using Lemma 4, we get

$$|(A_1 g_{k-1}, A_0^{-1} g_{k-1})| \leq \sqrt{(A_1 g_{k-1}, A_0^{-1} A_1 g_{k-1})} c_{k-1} = \sqrt{\rho} \delta c_{k-1}. \quad (2.7)$$

From (2.3) to (2.7) and utilizing Lemma 3,

$$\begin{aligned} c_k &\leq c_{k-1} - \rho^{-1} c_{k-1} + 2\lambda_{\min}^{-1} \sqrt{\rho} \delta c_{k-1} + \lambda_{\min}^{-2} \rho \delta^2 c_{k-1} \\ &= c_{k-1} (1 - \rho^{-1} + 2\lambda_{\min}^{-1} \sqrt{\rho} \delta + \lambda_{\min}^{-2} \rho \delta^2). \end{aligned}$$

For the purpose of convergence, we need

$$1 - \rho^{-1} + 2\lambda_{\min}^{-1} \sqrt{\rho} \delta + \lambda_{\min}^{-2} \rho \delta^2 < 1,$$

$$2\lambda_{\min}^{-1} \sqrt{\rho} \delta + \lambda_{\min}^{-2} \rho \delta^2 - \rho^{-1} < 0,$$

multiplying by $\lambda_{\min}^2 \rho^{-1}$, we have

$$\delta^2 + 2\lambda_{\min} \sqrt{\rho^{-1}} \delta - \lambda_{\min}^2 \rho^{-2} < 0.$$

This satisfies when

$$\delta < -\lambda_{\min} \sqrt{\rho^{-1}} + \sqrt{\lambda_{\min}^2 \rho^{-1} + \lambda_{\min}^2 \rho^{-2}} = \lambda_{\min} \sqrt{\rho^{-1}} (-1 + \sqrt{1 + \rho^{-1}}).$$

This completes the proof. \square

3. Numerical Experiments

In this section, two numerical experiments are presented to show the rate of convergence of the Barzilai-Borwein (BB) algorithm for solving the linear system of equations $Ax = b$ where A is slightly unsymmetric positive definite matrix. They demonstrate that if the sufficient condition

$$\|A_1\| < \lambda_{min}\sqrt{\rho^{-1}}(-1 + \sqrt{1 + \rho^{-1}})$$

Satisfies, then the BB algorithm convergence. Simulations were run in MatLab 27.

Example 1. For the first experiment, a matrix A with size $d \times d$ can be taken as

$$\begin{bmatrix} 1.0000 & -0.0009 & -0.0001 & -0.0013 & -0.0007 & 0.0020 & 0.0021 \\ -0.0007 & 1.6667 & -0.0000 & -0.0022 & -0.0013 & 0.0012 & 0.0008 \\ 0.0013 & 0.0015 & 2.3333 & -0.0007 & -0.0029 & 0.0018 & 0.0003 \\ -0.0004 & -0.0014 & -0.0023 & 3.0000 & -0.0007 & -0.0019 & 0.0027 \\ 0.0000 & 0.0024 & -0.0014 & 0.0026 & 3.6667 & -0.0024 & 0.0026 \\ -0.0017 & 0.0011 & -0.0020 & 0.0007 & -0.0033 & 4.3333 & 0.001 \\ -0.0021 & 0.0021 & 0.0028 & -0.0027 & 0.0023 & 0.0020 & 5.0000 \end{bmatrix}$$

where $n = 7$, $A = [a_{ij}]$, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, n$ such that a_{ij} are equally spaced real numbers between 1 and 5 when $i = j$, and a_{ij} is a random number between -1 and 1 when $i \neq j$. The symmetric part of A (A_0) is symmetric positive definite matrix and its condition number $\rho = 5$ where $\lambda_{min} = 1$ and $\lambda_{max} = 5$. $\|A_1\| = 0.0041$. The value of $\lambda_{min}\sqrt{\rho^{-1}}(-1 + \sqrt{1 + \rho^{-1}}) = 0.0427$.

Figure 1 shows the rate of convergence $r_k = \frac{(g_{k+1}, g_{k+1})}{(g_k, g_k)}$ of the BB algorithm as a function of number of iteration $k=300$

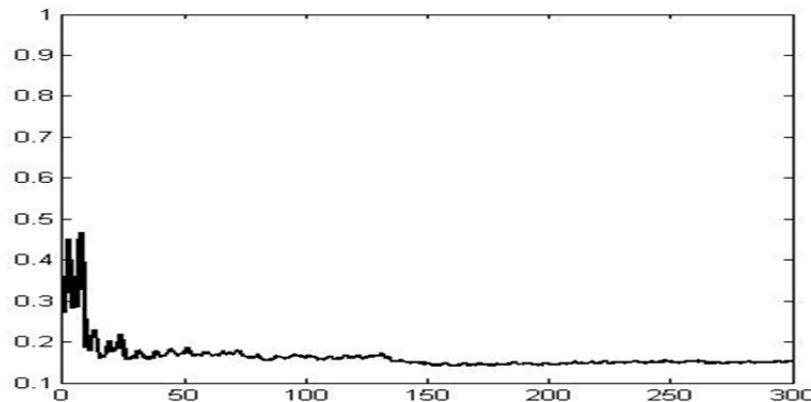


Figure 1: Rate r_k of convergence as a function of k for Example 1.

Example 2. For the second experiment, a matrix A with size $d \times d$ can be taken as

$$\begin{bmatrix} 11.0000 & 0.7621 & 0.7621 & 0.4057 & 0.0579 \\ 0.0000 & 11.0000 & 0.7919 & 0.9355 & 0.3529 \\ 0.0000 & 0.0000 & 11.0000 & 0.9169 & 0.8132 \\ 0.0000 & 0.0000 & 0.0000 & 11.0000 & 0.0099 \\ 0.0000 & 0.0000 & 0.0000 & 0.0000 & 11.0000 \end{bmatrix}$$

where $d = 5$, $A = [a_{ij}]$, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, n$ such that $a_{ij} = 11$ when $i = j$, and $a_{ij} = rand(d)$ when $i \neq j$. A_0 is symmetric positive definite matrix and its condition number $\rho = 1.1801$ where $\lambda_{min} = 10.3449$ and $\lambda_{max} = 12.2082$. $\|A_1\| = 1.0117$ and the value of $\lambda_{min}\sqrt{\rho^{-1}}(-1 + \sqrt{1 + \rho^{-1}}) = 3.4204$.

Figure 2 shows the rate of convergence $r_k = \frac{(g_{k+1}, g_{k+1})}{(g_k, g_k)}$ of the BB algorithm as a function of number of iteration $k = 100$.

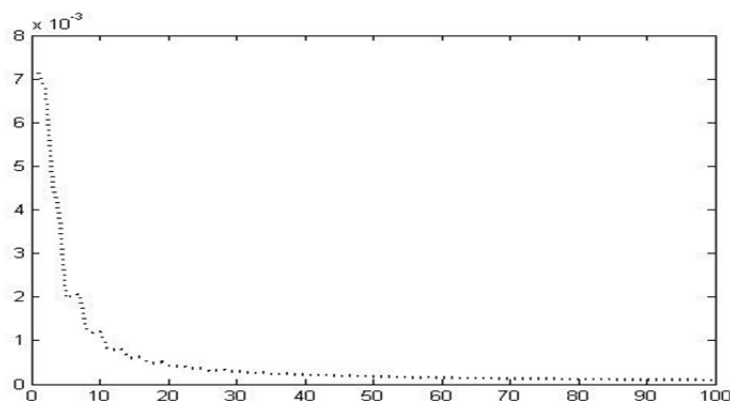


Figure 2: Rate r_k of convergence as a function of k for Example 2.

References

- Barzilai, J. and Borwein, J. M.: Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8, pp. 141–148, 1988.
- Dai, Y. H. and Fletcher, R.: On the asymptotic behaviour of some new gradient methods. *Mathematical Programming*, 103(3), pp541–559, 2005.
- Dai, Y. H. and Liao, L. Z.: R-linear convergence of the Barzilai and Borwein method for unconstrained optimization. *IMA Journal of Numerical Analysis*, 1, pp. 1–10, 2002.
- Dennis, J. E. and Schnabel, R. B.: *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, NJ, 1983.
- Fletcher, R.: *Practical Methods of Optimization*. Wiley, 2nd edition, 2000.
- Raydan, M. M.: On the Barzilai and Borwein choice of steplength for the gradient method. *IMA Journal of Numerical Analysis*, 13, pp. 321–326, 1993.
- Raydan, M. M.: *Convergence properties of the Barzilai and Borwein gradient method*. PhD thesis, Rice University, 1991.

پوختنه

پهیدا کرنا ریڼکه کا پلهی لاری Barzilai و Borwein کرنگیه کا زیده د بوارین زانستی جودادا ههیه جونکه هژمارتنا وی ب ساناهی و تواناتره . د قی فه کولینی دا مهرجین پیدقی هاتیه دیتن بو نیزیکرنا ریڼک Barzilai و Borwein ژ بو شیکارکرنا هاوکیشن جبری، د وی دومیدا وهختی ژمارین ریژکری نه هاوتایه بیئت تا وی رادی کو بهشک ژ وی نیف هاوتایه و موجهبه

الملخص

لاقت طريقة التدرج ل Barzilai و Borwein اهتمامات كثيرة في مختلف الحقول العلمية نظرا لبساطتها وكفائتها العددية. في هذا البحث تم ايجاد الشروط الكافي لتقارب طريقة Barzilai و Borwein لحل المعادلات الجبرية عندما تكون مصفوفة المعاملات غير متناظرة الى حد ما ويكون الجزء الشبه المتناظرة منها موجب