Journal of Computational Mathematics
Vol.37, No.6, 2019, 916–936.

http://www.global-sci.org/jcm
doi:10.4208/jcm.1911-m2019-0171

STABILIZED BARZILAI-BORWEIN METHOD*

Oleg Burdakov
Department of Mathematics, Linköping University, Linköping, Sweden
Email: oleg.burdakov@liu.se
Yuhong Dai
LSEC, ICMSEC, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
Email: dyh@lsec.cc.ac.cn
Na Huang[1)]
Department of Applied Mathematics, College of Science, China Agricultural University, Beijing 100083, China
Email: hna@cau.edu.cn

Abstract

The Barzilai-Borwein (BB) method is a popular and efficient tool for solving large-scale unconstrained optimization problems. Its search direction is the same as for the steepest descent (Cauchy) method, but its stepsize rule is different. Owing to this, it converges much faster than the Cauchy method. A feature of the BB method is that it may generate too long steps, which throw the iterates too far away from the solution. Moreover, it may not converge, even when the objective function is strongly convex. In this paper, a stabilization technique is introduced. It consists in bounding the distance between each pair of successive iterates, which often allows for decreasing the number of BB iterations. When the BB method does not converge, our simple modification of this method makes it convergent. For strongly convex functions with Lipschits gradients, we prove its global convergence, despite the fact that no line search is involved, and only gradient values are used. Since the number of stabilization steps is proved to be finite, the stabilized version inherits the fast local convergence of the BB method. The presented results of extensive numerical experiments show that our stabilization technique often allows the BB method to solve problems in a fewer iterations, or even to solve problems where the latter fails.

Mathematics subject classification: 65K05, 90C06, 90C30.
Key words: Unconstrained optimization, Spectral algorithms, Stabilization, Convergence analysis.

1. Introduction

In this paper, we consider spectral gradient methods for solving the unconstrained optimization problem

$$\min_{x \in R^n} f(x), \tag{1.1}$$

where $f : R^n \to R^1$ is a sufficiently smooth function. Its minimizer is denoted by x^*. Gradient-type iterative methods used for solving problem (1.1) have the form

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

* Received July 16, 2019 / Revised version received September 30, 2019 / Accepted November 11, 2019 / Published online November 21, 2019 /
1) Corresponding author

where $g_k = \nabla f(x_k)$ and $\alpha_k > 0$ is a stepsize. Methods of this type differ in the stepsize rules which they follow.

We focus here on the two choices of α_k proposed in 1988 by Barzilai and Borwein [1], usually refereed to as the BB method. The rationale behind these choices is related to viewing the gradient-type methods as quasi-Newton methods, where α_k in (1.2) is replaced by the matrix $D_k = \alpha_k I$. This matrix is served as an approximation of the inverse Hessian matrix. Following the quasi-Newton approach, the stepsize is calculated by forcing either D_k^{-1} (BB1 method) or D_k (BB2 method) to satisfy the secant equation in the least squares sense. The corresponding two problems are formulated as

$$\min_{D=\alpha I} \|D^{-1}s_{k-1} - y_{k-1}\| \quad \text{and} \quad \min_{D=\alpha I} \|s_{k-1} - Dy_{k-1}\|, \tag{1.3}$$

where $s_{k-1} = x_k - x_{k-1}$ and $y_{k-1} = g_k - g_{k-1}$. The solutions to these problems are

$$\alpha_k^{BB1} = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} \quad \text{and} \quad \alpha_k^{BB2} = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}, \tag{1.4}$$

respectively. Here and in what follows, $\|\cdot\|$ denotes the Euclidean vector norm and the induced matrix norm. Other norms used in this paper will be denoted in a different way.

Barzilai and Borwein [1] proved that their method converges R-superlinearly for two-dimensional strictly convex quadratics. Dai and Fletcher [7] analyzed the asymptotic behavior of BB-like methods, and they obtained R-superlinear convergence of the BB method for the three-dimensional case. Global convergence of the BB method for the n-dimensional case was established by Raydan [20] and further refined by Dai and Liao [10] for obtaining the R-linear rate. For nonquadratic functions, local convergence proof of the BB method with R-linear rate was, first, sketched in some detail by Liu and Dai [19], and then it was later rigorously proved by Dai et al. [9]. Extensive numerical experiments show that the two BB stepsize rules significantly improve the performance of gradient methods (see, e.g., [14,21]), both in quadratic and nonquadratic cases.

A variety of modifications and extensions have been developed, such as gradient methods with retards [15], alternate BB method [8], cyclic BB method [9], limited memory gradient method [4] etc. Several approaches were proposed for dealing with nonconvex objective functions, in which case the BB stepsize (1.4) may become negative. In our numerical experiments, we use the one proposed in [6]. The BB method has been extended to solving symmetric and nonsymmetric linear equations [6, 11]. Furthermore, by incorporating the nonmontone line search by Grippo et al. [17], Raydan [21] and Grippo et al. [18] developed the global BB method for general unconstrained optimization problems. Later, Birgin et al. [2] proposed the so-called spectral projected gradient method which extends Raydan's method to smooth convex constrained problems. For more works on BB-like methods, see [3,14,23] and references therein.

As it was observed by many authors, the BB method may generate too long steps, which throw the iterates too far away from the solution. In practice, it may not converge even for strongly convex functions (see, e.g., [14]). The purpose of this paper is to introduce a simple stabilization technique and to justify its efficiency both theoretically and practically. Our stabilization does not assume any objective function evaluations. It consists in uniformly bounding $\|s_k\|$, the distance between each pair of successive iterates. It should be emphasized that, if the BB method safely converges for a given function, then there is no necessity in stabilizing it. In such cases, the stabilization may increase the number of iterations. In other