Journal of Computational Mathematics
Vol.39, No.3, 2021, 375–391.

http://www.global-sci.org/jcm
doi:10.4208/jcm.2001-m2018-0205

TWO NOVEL GRADIENT METHODS WITH OPTIMAL STEP SIZES*

Harry Oviedo, Oscar Dalmau[1] and Rafael Herrera
Centro de Investigación en Matemáticas, CIMAT A.C. Guanajuato, Gto. Mexico
Email: harry.oviedo@cimat.mx, dalmau@cimat.mx, rherrera@cimat.mx

Abstract

In this work we introduce two new Barzilai and Borwein-like steps sizes for the classical gradient method for strictly convex quadratic optimization problems. The proposed step sizes employ second-order information in order to obtain faster gradient-type methods. Both step sizes are derived from two unconstrained optimization models that involve approximate information of the Hessian of the objective function. A convergence analysis of the proposed algorithm is provided. Some numerical experiments are performed in order to compare the efficiency and effectiveness of the proposed methods with similar methods in the literature. Experimentally, it is observed that our proposals accelerate the gradient method at nearly no extra computational cost, which makes our proposal a good alternative to solve large-scale problems.

Mathematics subject classification: 90C20, 90C25, 90C52, 65F10.
Key words: Gradient methods, Convex quadratic optimization, Hessian spectral properties, Steplength selection.

1. Introduction

In this paper, we consider the classical convex quadratic minimization problem,

$$\min_{x\in\mathbb{R}^n} f(x) = \frac{1}{2}x^\top Ax - x^\top b \tag{1.1}$$

where $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times n}$ is a symmetric and positive definite matrix.

One of the most studied algorithms to address unconstrained optimization problems is the *steepest descent method*, which was first introduced by Cauchy in 1847 [1]. This method is part of the so-called first-order methods, which only use information about the objective function and its derivative. It is well known that the steepest descent method is an ineffective method to solve (1.1) numerically, due to its slow convergence rate and zigzag pattern along the iterations. However, this method has a well-established convergence analysis that has been of great help to study other algorithms. In addition, the steepest descent method requires little memory per iteration. In spite of its drawbacks, in the 80's, a rebirth of the method arose due to the works of Iudin-Nemirowsky [2] and Nesterov [3] which proposed a technique of acceleration of the gradient method by using an extra-momentum step that can be used to address large-scale convex minimization problems. Additionally, in 1988 Barzilai and Borwein [4] introduced two new alternatives to select the step size in the *gradient method* for solving problem (1.1) that

* Received September 29, 2018 / Revised version received February 18, 2019 / Accepted January 19, 2020 / Published online November 4, 2020 /
1) Corresponding author

greatly accelerate the convergence rate of the method. In [5], Raydan developed an ingenious convergence study of the Barzilai-Borwein method. In addition, Dai et al. [6] demonstrated that this method has R-linear convergence rate.

Initially, some step sizes were proposed for the quadratic case (1.1) [4, 10–12, 16] and then extended to the non-linear case. An extension for the case of non-linear optimization of the gradient method with Barzilai-Borwein steps (BB-steps) size has been provided by Raydan et al. in [7]. Other extensions of the method are found in [8, 9]. From the seminal paper [4], many researchers have taken up the steepest descent study in order to design novel effective gradient-type methods by introducing new step sizes to address large-scale optimization problems efficiently. To date, several formulations have been proposed to select step size, which attempt to incorporate second-order information without increasing significatively the computational cost of the classical steepest descent method. For example, in [16] Yuan introduced an ingenious step size that was built by imposing finite termination for the two dimensional quadratic problem. In addition, Dai and Yuan in [10], Dai [11] proposed two step sizes that alternate the BB-steps and the exact step size of the classic steepest descent, in the odd and even iterations. Additionally, a new step size has been proposed very recently in [12], which uses the BFGS update formula to build a new approximate optimal step size.

In this paper we study two step sizes based on optimization models for the gradient methods. Our main contribution is to introduce and analyze two new choices of the step size which are obtained as solutions of such optimization models. Using these two-point step size we design an efficient gradient algorithm. In addition, we provide a convergence analysis of the proposed algorithm and some numerical experiments where we compare our algorithm with similar algorithms from the state of the art.

This paper is organized as follows. In the next section we briefly present the gradient method and some of its variants. Section 3 is dedicated to summarizing the step sizes proposed by Barzilai and Borwein. Our contribution of two new step sizes is introduced in Section 4. A convergence analysis is provided in Section 5. In Section 6 we present some numerical results to show the efficiency and effectiveness of the proposed step sizes. Finally, in Section 7 we present the general conclusions of this work.

2. Steepest Descent Method and Its Variants

The gradient method to solve (1.1) is an iterative method that uses the following recursive formula starting from a given point x_0:

$$x_{k+1} = x_k - \frac{1}{\alpha_k} g_k, \tag{2.1}$$

where $g_k = \nabla f(x_k) = Ax_k - b$, $\alpha_k > 0$ and $1/\alpha_k$ is known as the step-size. There are several alternatives for the step-size; the most popular one is given by choosing the step size of the k-th iteration (α_k) as the positive scalar that minimizes the objective function along the direction of the negative gradient, i.e.

$$\alpha_k = \arg\min_{\alpha>0} f(x_k - \frac{1}{\alpha} g_k). \tag{2.2}$$

It is not difficult to prove that the solution to (2.2) is given by

$$\alpha_k^{SD} = \frac{g_k^\top A g_k}{g_k^\top g_k}. \tag{2.3}$$