Applied Mathematics and Computation 000 (2017) 1–20

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

On the steplength selection in gradient methods for unconstrained optimization

Daniela di Serafino [a,*], Valeria Ruggiero [b], Gerardo Toraldo [c], Luca Zanni [d]

[a] Department of Mathematics and Physics, University of Campania Luigi Vanvitelli, viale A. Lincoln 5, Caserta I-81100, Italy
[b] Department of Mathematics and Computer Science, University of Ferrara, via Saragat 1, Ferrara I-44122, Italy
[c] Department of Mathematics and Applications, University of Naples Federico II, Via Cintia 21, Naples I-80126, Italy
[d] Department of Physics, Informatics and Mathematics, University of Modena and Reggio Emilia, via Campi 213/B, Modena I-41125, Italy

ARTICLE INFO

Article history:
Available online xxx

MSC:
65K05
90C20
90C30

Keywords:
Gradient methods
Steplength selection
Hessian spectral properties

ABSTRACT

The seminal paper by Barzilai and Borwein (1988) has given rise to an extensive investigation, leading to the development of effective gradient methods. Several steplength rules have been first designed for unconstrained quadratic problems and then extended to general nonlinear optimization problems. These rules share the common idea of attempting to capture, in an inexpensive way, some second-order information. However, the convergence theory of the gradient methods using the previous rules does not explain their effectiveness, and a full understanding of their practical behaviour is still missing. In this work we investigate the relationships between the steplengths of a variety of gradient methods and the spectrum of the Hessian of the objective function, providing insight into the computational effectiveness of the methods, for both quadratic and general unconstrained optimization problems. Our study also identifies basic principles for designing effective gradient methods.

1. Introduction

Many real life applications lead to nonlinear optimization problems whose very large size makes first-order methods the most suitable choice. Among first-order approaches, gradient methods have widely proved their effectiveness in solving challenging unconstrained and constrained problems arising in signal and image processing, compressive sensing, machine learning, optics, chemistry and other areas (see, e.g., [1–11] and the references therein).

These methods underwent a renaissance since the work by Barzilai and Borwein [12], which showed how a suitable choice of the steplength can significantly accelerate the classical Steepest Descent method [13,14]. Since then, several steplength rules have been designed in order to increase the efficiency of gradient methods, while preserving their simplicity and low memory requirement. Most of these rules have been first developed for the unconstrained convex quadratic problem [15–29], which is not only of practical importance in itself, but also provides a simple setting to design effective methods for more general problems. The extension of steplength selection strategies from convex quadratic to general nonlinear optimization has involved interesting theoretical issues, leading to the exploitation of line search strategies in order to guarantee convergence to stationary points [17,18,24,30–36].

* Corresponding author.
E-mail addresses: daniela.diserafino@unicampania.it (D. di Serafino), valeria.ruggiero@unife.it (V. Ruggiero), toraldo@unina.it (G. Toraldo), luca.zanni@unimore.it (L. Zanni).

http://dx.doi.org/10.1016/j.amc.2017.07.037
0096-3003/

The theoretical convergence results of gradient methods based on the previous steplength rules do not explain their effectiveness, and a full understanding of their practical behaviour is still missing. A feature shared by most of these methods consists in exploiting spectral properties of the Hessian of the objective function through (usually implicit) low cost approximations of expensive second-order information. This appears to be the main reason for their good behaviour (see, e.g., [21,24,26,33,37]); however, a deeper and more systematic analysis is needed.

The aim of this work is to investigate the relationships between the steplengths exploited by some well known gradient methods and the spectrum of the Hessian of the objective function, for convex quadratic and general problems of the form

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1}$$

where $f:\mathbb{R}^n \to \mathbb{R}$ is continuously differentiable. In this case, the gradient method iteration reads

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

where $g_k = \nabla f(x_k)$ and $\alpha_k > 0$ is the steplength.

We first consider the convex quadratic problem

$$\min_{x\in\mathbb{R}^n} f(x) := \frac{1}{2} x^T A x - b^T x, \tag{3}$$

where $A \in \mathbb{R}^{n\times n}$ is symmetric positive definite and $b \in \mathbb{R}^n$. It provides a simple framework for investigating the role of the eigenvalues of the Hessian matrix in the behaviour of gradient methods; furthermore, convergence results involving the spectrum of the Hessian are available in this case, which provide a sound basis for our analysis. We deal with a selection of approaches, representative of a wide class of gradient methods, as explained later in this paper. We consider the following methods: Barzilai–Borwein and Adaptive Barzilai–Borwein variants [20,22], Limited Memory Steepest Descent [24], Steepest Descent with Alignment and Steepest Descent with Constant (Yuan) steps [26,27]; we also consider methods such that the inverses of their steplengths follow predefined distributions obtained exploiting the Golden Arcsin rule [38] or the Chebyshev nodes [29].

In the second part of the paper, we deal with the general unconstrained problem, focusing on gradient methods whose steplengths are natural extensions of the rules developed for the convex quadratic case, combined with line search strategies forcing convergence. In particular, we investigate methods based on the Barzilai–Borwein, the $\text{ABB}_{\text{minmin}}$ Adaptive Barzilai–Borwein [22] and the Limited Memory Steepest Descent rules.

The main contribution of this paper is a careful and unifying analysis of a variety of steplength rules and their relationships with second-order information of the problem, aimed at better understanding the computational effectiveness of some state-of-the-art gradient methods. In particular, a deeper look at the basic principles used for “capturing” Hessian spectral properties provides useful guidelines for designing effective gradient approaches, not only for the quadratic case but also for general unconstrained minimization problems.

The paper is organized as follows. In Section 2, after some preliminary results on gradient methods applied to strictly convex quadratic problems, we discuss the relationships between the steplengths and the spectrum of the Hessian in the quadratic case, showing the results of a set of numerical experiments. This analysis is extended to the non-quadratic case in Section 3. Some conclusions are provided in Section 4.

2. Convex quadratic problems

We first consider the strictly convex quadratic problem (3), in order to highlight the strict relationship between the behaviour of gradient methods and the eigenvalues of the Hessian of the objective function. In particular, we show how some choices of the steplength exploit spectral properties of the Hessian matrix in order to achieve efficiency in the corresponding methods. We start by giving some preliminary results, which will be useful in our analysis.

2.1. Notation and preliminaries

Let $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ be the eigenvalues of the matrix A in (3), and $\{d_1, d_2, \ldots, d_n\}$ a set of associated orthonormal eigenvectors. The gradient g_k can be expressed as

$$g_k = \sum_{i=1}^{n} \mu_i^k d_i, \quad \mu_i^k \in \mathbb{R}, \tag{4}$$

where μ_i^k is called the ith eigencomponent of g_k. Henceforth, without loss of generality, we assume that

$$\lambda_1 > \lambda_2 > \cdots > \lambda_n > 0, \quad \mu_1^0 \neq 0, \quad \mu_n^0 \neq 0$$

(see, e.g., [27, Section 1] and [24, Section 2]).

Please cite this article as: D. di Serafino et al., On the steplength selection in gradient methods for unconstrained optimization, Applied Mathematics and Computation (2017), http://dx.doi.org/10.1016/j.amc.2017.07.037