Linear Algebra and its Applications 471 (2015) 544–568

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Adaptive matrix algebras in unconstrained minimization

S. Cipolla, C. Di Fiore *, F. Tudisco, P. Zellini

University of Rome "Tor Vergata", Department of Mathematics, Via della Ricerca Scientifica, 1 – 00133 Rome, Italy

A R T I C L E   I N F O

Article history:
Received 8 August 2014
Accepted 9 January 2015
Available online 30 January 2015
Submitted by R. Brualdi

MSC:
65K10
90C53
47J25

Keywords:
Unconstrained minimization
Quasi-Newton methods
Matrix algebras
Iterative procedures

A B S T R A C T

In this paper we study adaptive $\mathcal{L}^{(k)}$QN methods, involving special matrix algebras of low complexity, to solve general (non-structured) unconstrained minimization problems. These methods, which generalize the classical BFGS method, are based on an iterative formula which exploits, at each step, an *ad hoc* chosen matrix algebra $\mathcal{L}^{(k)}$. A global convergence result is obtained under suitable assumptions on f.

1. Introduction

Quasi-Newton methods for the unconstrained minimization of a function $f : \mathbb{R}^n \to \mathbb{R}$ are based on iterative schemes of the form $\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{d}_k$, where $\mathbf{d}_k$ is a descent direction in $\mathbf{x}_k$, i.e. $\nabla f(\mathbf{x}_k)^T \mathbf{d}_k < 0$, and λ_k is the steplength.

* Corresponding author.
E-mail addresses: cipolla@mat.uniroma2.it (S. Cipolla), difiore@mat.uniroma2.it (C. Di Fiore), tudisco@mat.uniroma2.it (F. Tudisco), zellini@mat.uniroma2.it (P. Zellini).

http://dx.doi.org/10.1016/j.laa.2015.01.010

Let us recall that *any* descent direction $\mathbf{d}_k$ for f in the current guess $\mathbf{x}_k$ solves the equation $A_k\mathbf{d}_k = -\mathbf{g}_k$ for some real symmetric positive definite (pd) matrix A_k approximating the Hessian of f in $\mathbf{x}_k$, where $\mathbf{g}_k$ is the first derivative vector $\nabla f(\mathbf{x}_k)$ (see [9]).

A *good* property that quasi-Newton methods should have, seems to be that A_{k+1} satisfies the equation $A_{k+1}\mathbf{s}_k = \mathbf{y}_k$ (Secant equation), where $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ and $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$. Quasi-Newton methods with such property will be referred to as Secant. Apparently, the secant equation is far to be a mere optional condition. In [12, p. 24] it is observed that the equality $A_{k+1}\mathbf{s}_k = \mathbf{y}_k$ mimics the fundamental property of the Hessian $\nabla^2 f(\mathbf{x}_{k+1})\mathbf{s}_k \approx \mathbf{y}_k$, whereas in [4, p. 54] the same equality "is central for the development of quasi-Newton methods, and therefore it has often been called the *quasi-Newton equation*". Also in [1, p. 223] the secant equation appears as a fundamental ingredient in the *definition* of quasi-Newton methods.

In [7,10,8,9,6] it was introduced a new class of algorithms, named $\mathcal{L}$QN, which includes methods of Secant type, in particular the well known BFGS method, and, at the same time, some methods which are not Secant but have relevant good properties (f.i. global convergence). The main purpose consisted in saving the second order information of the matrix B_k, produced by the BFGS method to approximate a full (not sparse) Hessian of f, in a form that allows to reduce the high ($O(n^2)$) computational cost per step of BFGS. More in detail, a substantial generalization of the BFGS scheme has been therein proposed by an updating Hessian approximation formula of the form

$$B_{k+1} = \Phi(\tilde{B}_k, \mathbf{s}_k, \mathbf{y}_k) \tag{1}$$

where $\tilde{B}_k$ is a suitable approximation of B_k and Φ is the BFGS-type rank-two correction of $\tilde{B}$:

$$\Phi(\tilde{B}, \mathbf{s}, \mathbf{y}) := \tilde{B} - \frac{1}{\mathbf{s}^T\tilde{B}\mathbf{s}}\tilde{B}\mathbf{s}\mathbf{s}^T\tilde{B} + \frac{1}{\mathbf{y}^T\mathbf{s}}\mathbf{y}\mathbf{y}^T.$$

The BFGS method is retrieved if $\tilde{B}_k = B_k$ for all k. Moreover, a suitable choice of $\tilde{B}_k$ yields the important class of $\mathcal{L}$QN methods, where the quasi-Newton matrix approximating the Hessian is defined also in terms of a matrix algebra $\mathcal{L}$. The matrices of this algebra $\mathcal{L}$ are simultaneously reduced to diagonal form by a unitary matrix U, i.e. $\mathcal{L} = \mathrm{sd}\, U = \{L = Ud(\mathbf{z})U^H\}$ where $d(\mathbf{z})$ denotes the diagonal matrix of the eigenvalues z_i of L. In fact, if $\tilde{B}_k$ is the best approximation $\mathcal{L}_{B_k}$ in $\mathcal{L}$ of B_k in Frobenius norm, then from (1) we obtain a simple single-array iteration to compute the eigenvalues of $\mathcal{L}_{B_{k+1}}$ from the eigenvalues of $\mathcal{L}_{B_k}$ [7]. At least two choices are possible for the *new* descent direction $\mathbf{d}_{k+1}$:

$$\mathbf{d}_{k+1} = -B_{k+1}^{-1}\mathbf{g}_{k+1} \quad \text{or} \quad \mathbf{d}_{k+1} = -\tilde{B}_{k+1}^{-1}\mathbf{g}_{k+1}.$$

The first choice yields a Secant (S) algorithm, because $B_{k+1}\mathbf{s}_k = \mathbf{y}_k$, whereas the second choice yields a Non-Secant (NS) procedure, as $\tilde{B}_{k+1}\mathbf{s}_k$ is in general different from $\mathbf{y}_k$.