Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER Computers and Mathematics with Applications 50 (2005) 1041–1050
An International Journal
computers & mathematics with applications
www.elsevier.com/locate/camwa

The Use of Alternation and Recurrences in Two-Step Quasi-Newton Methods

J. A. FORD
Department of Computer Science, University of Essex
Wivenhoe Park, Colchester, Essex, U.K.
fordj@essex.ac.uk

I. A. MOGHRABI
Computer Science Division, Faculty of Science
Beirut Arab University, Beirut, Lebanon
i_moghrabi@yahoo.com

Abstract—Of the multistep quasi-Newton methods introduced by the authors in [1], the most successful was the so-called fixed-point method using the existing Hessian approximation to compute, at each iteration, the parameters required in the interpolation. In order to avoid the burden of computing the additional matrix-vector products required by this approach, approximations based on the secant equation were proposed. In [2], a different approach to dealing with this difficulty was proposed, in which standard single-step quasi-Newton updates were alternated, on successive iterations, with two-step updates, so that approximations were no longer necessary.

Recent work has shown that the quantities required to compute the parameters referred to above may be computed exactly by means of a recurrence, so that the technique of alternation is no longer the only alternative if we wish to avoid approximations. In this paper, we describe the derivation of this recurrence. We present the results of a range of numerical experiments to compare and evaluate the three approaches of *approximation*, *alternation*, and *recurrence*. Finally, we show how the use of recurrences may be extended to multistep methods employing three or more steps.

Keywords—Unconstrained optimisation, Quasi-Newton methods, Multistep quasi-Newton methods.

1. INTRODUCTION

The problem we consider here is the unconstrained minimisation of a twice continuously differentiable objective function $f(\mathbf{x})$ (where $\mathbf{x} \in \mathbb{R}^n$). We denote the gradient and Hessian of f by $\mathbf{g}$ and G, respectively. Quasi-Newton methods for this problem imitate Newton's method, without requiring that the Hessian be available in explicit form. Instead, they compute approximations $\{B_i\}$ to the Hessian matrices at the various iterates $\{\mathbf{x}_i\}$ which are generated. In standard quasi-Newton methods, the new approximation B_{i+1} is required to satisfy the secant (or quasi-Newton) equation [3]

$$B_{i+1}\mathbf{s}_i = \mathbf{y}_i, \tag{1}$$

0898-1221/05/$ - see front matter Typeset by AMS-TEX
doi:10.1016/j.camwa.2005.08.006

where $\mathbf{s}_i$ is the step from $\mathbf{x}_i$ to $\mathbf{x}_{i+1}$ and $\mathbf{y}_i$ is the corresponding gradient difference

$$\mathbf{s}_i = \mathbf{x}_{i+1} - \mathbf{x}_i; \tag{2}$$

$$\mathbf{y}_i = \mathbf{g}(\mathbf{x}_{i+1}) - \mathbf{g}(\mathbf{x}_i). \tag{3}$$

Because such methods employ data from just one step in performing the update of B_i, we refer to them as *single-step* methods. In contrast, *two-step* quasi-Newton methods require that the approximation B_{i+1} satisfy a condition of the following form:

$$B_{i+1}(\mathbf{s}_i - \gamma_i \mathbf{s}_{i-1}) = \mathbf{y}_i - \gamma_i \mathbf{y}_{i-1} \tag{4}$$

or

$$B_{i+1}\mathbf{r}_i = \mathbf{w}_i, \text{ say.} \tag{5}$$

The derivation of (4) is described by Ford and Moghrabi [4,5]. Quadratic curves $\mathbf{x}(\tau)$ and $\mathbf{h}(\tau) \approx \mathbf{g}(\mathbf{x}(\tau))$ (where $\tau \in \mathbb{R}$) are constructed which interpolate, respectively, the three latest iterates $\mathbf{x}_{i-1}$, $\mathbf{x}_i$, and $\mathbf{x}_{i+1}$, and the three associated gradient evaluations (which are assumed to be available). The derivatives of these two curves (evaluated at $\tau = \tau_2$, where τ_j is the value of τ for which

$$\mathbf{x}(\tau_j) = \mathbf{x}_{i-1+j})$$

are then substituted into the relation (derived from the chain rule)

$$G(\mathbf{x}_{i+1})\mathbf{x}'(\tau_2) = \mathbf{g}'(\mathbf{x}(\tau_2)), \tag{6}$$

where primes denote differentiation with respect to τ. On making substitutions for the derivatives into (6) and removing a common scaling factor, we obtain a relation of the form (4) for

$$B_{i+1} \approx G(\mathbf{x}_{i+1})$$

to satisfy. Since equation (4),(5) has the same structural form as (1), it follows that B_{i+1} may then be obtained (for example) by use of an appropriately modified version of the BFGS formula [6–9]

$$B_{i+1} = B_i - \frac{B_i \mathbf{r}_i \mathbf{r}_i^\top B_i}{\mathbf{r}_i^\top B_i \mathbf{r}_i} + \frac{\mathbf{w}_i \mathbf{w}_i^\top}{\mathbf{w}_i^\top \mathbf{r}_i} \tag{7}$$

$$\triangleq BFGS(B_i, \mathbf{r}_i, \mathbf{w}_i). \tag{8}$$

The remainder of this paper is organised as follows: in Section 2, we give a description of two techniques for determining a suitable parametrisation of the interpolating curve $\mathbf{x}(\tau)$. Section 3 covers two strategies (namely, *approximation* and *alternation*) for reducing the computational effort involved in these interpolations for certain choices of a weighting matrix. In Section 4, we raise and answer (on the basis of numerical experiments) two questions about one of these strategies, while Section 5 develops an alternative strategy based on recurrences and Section 6 presents the results of further experiments to compare all three strategies.

2. DETERMINING THE PARAMETRISATION OF THE CURVE

The term γ_i in equation (4) is an expression depending on the three values τ_0, τ_1, and τ_2. It is therefore necessary to choose these three values with care, since the updating of the Hessian approximation (and, therefore, the numerical performance of such an algorithm) is determined by the value of γ_i (see equation (7)). Two successful approaches to the issue of defining suitable values for $\{\tau_k\}_{k=0}^m$ for general m-step multistep methods were developed by Ford and Moghrabi [1]. Distances between iterates $\mathbf{x}_j$ in $\mathbb{R}^n$ are measured by using a norm of the general form

$$\|\mathbf{z}\|_M \triangleq \{\mathbf{z}^\top M \mathbf{z}\}^{1/2}, \tag{9}$$