On the Generation of Updates for Quasi-Newton Methods

J. FLACHS

Communicated by R. A. Tapia

Abstract. We present a unified technique for updating approximations to Jacobian or Hessian matrices when any linear structure can be imposed. The updates are derived by variational means, where an operator-weighted Frobenius norm is used, and are finally expressed as solutions of linear equations and/or unconstrained extrema. A certain behavior of the solutions is discussed for certain perturbations of the operator and the constraints. Multiple secant relations are then considered. For the nonsparse case, an explicit family of updates is obtained including Broyden, DFP, and BFGS. For the case where some of the matrix elements are prescribed, explicit solutions are obtained if certain conditions are satisfied. When symmetry is assumed, we show, in addition, the connection with the DFP and BFGS updates.

Key Words. Quasi-Newton methods, updating formulas, Jacobian matrix, Hessian matrix, weighted Frobenius norm, sparsity, system of equations.

1. Introduction

In many cases, the solution of a system of nonlinear equations or an extremum of a function is found numerically by means of the so-called quasi-Newton methods. In most of these methods, the Jacobian (the Hessian in the case of an extremum) is reapproximated at each iteration by choosing the matrix nearest to the previous approximation from those matrices satisfying the secant relation and, perhaps, a particular known structure (for instance, symmetry with or without sparsity). For a comprehensive discussion, see Dennis and Schnabel (Ref. 1) and Dennis and Walker (Ref. 2).

1 This work was partially supported by a grant from Control Data.
2 Chief Researcher, National Research Institute for Mathematical Sciences, CSIR, Pretoria, South Africa.
In this paper, we propose and exploit a unified technique for finding updating formulae in which absolutely any linear structure of the current matrix, including the matricial secant relation (see below), can be imposed.

Bearing in mind the conditions of direct applicability, we detach the problem from its original context. The problem that has to be solved is the general one:

\[
\begin{align*}
\min \{ \| T_1^{-1}(X - X_0) \| _F \mid X \in \mathcal{F} \}, & \mathcal{F} \text{ is the solution set in } X \text{ of} \\
\min \{ \| T_2^{-1}(Y - Y_0) \| _F \mid V(X) + U(Y) = B, X \in \mathcal{L}_1 + \alpha, Y \in \mathcal{L}_2 \},
\end{align*}
\]

where \( \mathcal{L}_1 \subset \mathbb{R}^{m \times n} \) and \( \mathcal{L}_2 \subset \mathbb{R}^{k \times l} \) are subspaces of matrices; \( X_0, \alpha \in \mathbb{R}^{m \times n} \) and \( Y_0 \in \mathbb{R}^{k \times l} \) are matrices, while \( T_1, T_2, V, \) and \( U \) are linear operators taking values in finite-dimensional spaces; \( T_1 \) and \( T_2 \) are supposed nonsingular.

The notation \( \| \cdot \| _F \) denotes the Frobenius norm; ordinary capital letters (or Greek letters) are used for matrices, while script letters indicate sets. In our paper, the relation \( XP = Q \) is called the matricial secant relation, \( P \) and \( Q \) being matrices (rather than vectors, as in the usual secant relation).

In order to understand our problem (1) better, we now give a particular case. When \( \mathcal{L}_2 = \mathbb{R}^{m \times n} \), \( T_1 = T_2 = T \), and \( V = U \), the problem becomes: find the matrix closest to \( X_0 \) of all the matrices in \( \mathcal{L} + \alpha \) closest to \( \{ X \mid V(X) = B \} \), the norm being \( \| T^{-1}(\cdot) \| _F \). If we further suppose that \( V(X) = B \) is the usual secant relation (i.e., \( Xp = q \), where \( p \) and \( q \) are vectors) and that \( n = m \), we get the type of problems considered by Dennis and Schnabel (Ref. 1) (where \( T = I \), the identity, is considered) and Dennis and Walker (Ref. 2). Our results generalize those of Dennis and Schnabel, which inspired us. In Dennis and Walker's paper (Ref. 2), the update can eventually be expressed by means of a pseudo-inverse acting in the space of \( n \times n \) matrices, and orthogonal projections generated by a given inner product are involved. It is not shown how the update can be obtained or calculated. In contrast, we provide the equations that directly give the update and this we do for the more general case (1).

We consider the rather complex problem (1) since: (a) the equality constraint can express any linear structure as a matricial secant relation (i.e., \( XP = Q \), \( P \) and \( Q \) are matrices); (b) conditions involving higher derivatives or their estimates can be expressed; (c) updates of rectangular matrices can be obtained [see (b)]. In the dense case, the matricial secant relation is considered in the works of Gay, Schnabel, Salame, and Tewarson (Refs. 3-5). The \( W(\cdot)W \)-weighted Frobenius norm is of interest; however, either excessive computations are needed to obtain the update (see Proposition 5.3) or the algorithm using the update does not exhibit nice properties [see Dennis, Schnabel, Walker, Toint, Powell (Refs. 1, 2 and 6-8)]. This leads to the attempt to use a larger family of norms, which are of the form \( \| T^{-1}(\cdot) \| _F \), i.e., \( T^{-1} \)-weighted Frobenius norm.