Newton-type methods and quasi-Newton methods have proven to be very successful in solving dense unconstrained optimization problems. Recently there has been considerable interest in extending these methods to solving large problems when the Hessian matrix has a known a priori sparsity pattern. This paper treats sparse quasi-Newton methods in a uniform fashion and shows the effect of loss of positive-definiteness in generating updates. These sparse quasi-Newton methods coupled with a modified Cholesky factorization to take into account the loss of positive-definiteness when solving the linear systems associated with these methods were tested on a large set of problems. The overall conclusions are that these methods perform poorly in general—the Hessian matrix becomes indefinite even close to the solution and superlinear convergence is not observed in practice.


1. Introduction

The problem of concern in this paper is the unconstrained minimization of a twice-continuously differentiable function

\[ \min_{x \in \mathbb{R}^n} f(x). \]  

We shall consider the class of quasi-Newton methods applied to problem (1.1) when the Hessian matrix of the function \( f(x) \) has a known a priori sparsity pattern. The first quasi-Newton method was suggested by Davidon [1] and extended by Fletcher and Powell [5]. (For a comprehensive survey of quasi-Newton methods see [3].)

The idea behind the most popular quasi-Newton methods for (1.1) is to maintain a positive-definite symmetric matrix that approximates the Hessian matrix of \( f(x) \). If we let \( x_k \) denote the \( k \)th iterate, a quasi-Newton algorithm obtains a descent direction \( p_k \) by solving the system of equations

\[ B_k p_k = -g_k. \]  

* Research for this paper was performed at the Department of Operations Research, Stanford, CA 94305. The research was partially supported by the Department of Energy Contract AM03-76SF00326, PA# DE-AT03-76ER72018; Office of Naval Research Contract N00014-75-C-0267; National Science Foundation Grants MCS-7681259, MCS-7926009 and ECS-8012974.
where $B_k$ is an approximation to the Hessian matrix at iteration $k$ and $g_k$ is the gradient of $f$ at $x_k$. If $B_k$ is positive-definite, $p_k$ is guaranteed to be a descent direction. Once having obtained $p_k$, the new point $x_{k+1}$ is given by $x_k + \alpha_k p_k$, where $\alpha_k > 0$ is chosen so that $f(x_{k+1})$ is 'sufficiently' less than $f(x_k)$ (for a precise definition of this term, see, for example, [15]). If the new point $x_{k+1}$ does not satisfy the convergence criteria, a new approximation to the Hessian matrix $B_{k+1}$ is defined by

$$B_{k+1} = \varphi_k B_k + U_k,$$

where $\varphi_k$ is a scalar, and $U_k$ is a matrix chosen so that $B_{k+1}$ is symmetric, positive-definite and satisfies the quasi-Newton condition

$$B_{k+1}s_k = y_k,$$

with

$$s_k = x_{k+1} - x_k \text{ and } y_k = g_{k+1} - g_k.$$

Usually, $U_k$ is a matrix of low rank. The most popular update is the BFGS (Broyden-Fletcher-Goldfarb-Shanno) update (see, e.g., [3]), in which

$$\varphi_k = 1, \quad U_k = \frac{y_k y_k^T}{s_k y_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}.$$

To guarantee that the updated matrices are positive definite, we require that $s_k^T y_k > 0$ (which can be ensured by a suitable choice of $\alpha_k$).

Quasi-Newton methods have been very successful in solving unconstrained and constrained optimization problems of moderate size. The difficulty in applying these methods as described above to large problems is that a symmetric $n \times n$ matrix (or a factorization of a matrix) must be stored. However, many large problems have a sparse Hessian matrix whose sparsity pattern is known (or can be determined) a priori. In these cases it would be possible to maintain a suitably sparse approximation to the Hessian matrix. Note that, in general, the quasi-Newton methods described above generate dense approximations to the Hessian, regardless of the sparsity structure of the true Hessian. Much recent research has been directed towards generating Hessian approximations that have the same sparsity structure as the true Hessian (see [4, 14, 20, 24–27]). This paper treats various aspects of sparse versions of quasi-Newton methods in a uniform manner; and, reports results of extensive numerical testing of sparse quasi-Newton methods combined with a modified Cholesky factorization to provide a positive-definite matrix for computing the search direction.

Section 2 describes the notation to be used in the rest of the paper. Sparse quasi-Newton updates are described in Section 3. Section 4 describes positive-definiteness as related to these quasi-Newton methods. Storage and efficiency are considered in Section 5; and convergence of these methods is proved in Section 6. Finally, Section 7 discusses computational aspects of these methods and Section 8 summarizes the important results.