Modification of a Quasi-Newton Method for Nonlinear Equations with a Sparse Jacobian*

By L. K. Schubert

Abstract. For solving large systems of nonlinear equations by quasi-Newton methods it may often be preferable to store an approximation to the Jacobian rather than an approximation to the inverse Jacobian. The main reason is that when the Jacobian is sparse and the locations of the zeroes are known, the updating procedure can be made more efficient for the approximate Jacobian than for the approximate inverse Jacobian.

I. Introduction. In recent years a class of methods termed quasi-Newton have received considerable attention in the literature [1]–[3]. In one such method [1], [4] iterative approximations to the solution of the system of equations

$$f(x) = 0 \tag{1}$$

where f and x are n-vectors, are obtained by solving

$$G^{(k)}p^{(k)} = -f^{(k)} \tag{2}$$

and substituting $p^{(k)}$ in

$$x^{(k+1)} = x^{(k)} + t^{(k)}p^{(k)} . \tag{3}$$

The scalar $t^{(k)}$ is chosen to reduce some norm of f at each step, thus ensuring stability. The approximation $G^{(k)}$ to the Jacobian is revised after each step in accordance with

$$G^{(k+1)} = G^{(k)} + \frac{[f^{(k+1)} - (1 - t^{(k)})f^{(k)}]p^{(k)T}}{t^{(k)}p^{(k)T}p^{(k)}} . \tag{4}$$

This is the result of a primary condition requiring $G^{(k+1)}$ to predict the same changes in f in the direction $p^{(k)}$ that actually occurred at the $(k + 1)$th step (supplies n equations) and a secondary condition requiring $G^{(k+1)}$ to predict the same changes in f as $G^{(k)}$ in all directions orthogonal to $p^{(k)}$ (supplies the remaining $n^2 - n$ equations).

Broyden [1] described a class of methods containing the above method as a special case. However, he suggested the use of an approximation $H^{(k)}$ to the inverse Jacobian instead of $G^{(k)}$. He supplied an explicit updating algorithm for $H^{(k)}$ equivalent (in its simplest form) to (4). Other choices of the secondary con-

Received March 13, 1969.

AMS Subject Classifications. Primary 6550, 6565.

Key Words and Phrases. Quasi-Newton method, nonlinear simultaneous equations, approximation to Jacobian, nonlinear differential equations.

* The work presented in this note was supported by the National Research Council of Canada under Grant No. A2003 and by the Air Force Office of Scientific Research, United States Air Force, under AFOSR Grant No. 67-0672A.

dition on $G^{(k)}$ or $H^{(k)}$ are possible and this fact gives rise to the class of quasi-Newton methods.

When $H^{(k)}$ is used, the solution of the linear equations (2) is reduced to the matrix-vector multiplication

$$p^{(k)} = -H^{(k)} f^{(k)} . \tag{5}$$

This is certainly an advantage when all elements of the Jacobian are unknown. However, in many large systems of nonlinear equations, particularly the difference equations arising from nonlinear differential equations, most of the elements of the Jacobian are known to be zero and other elements may be known nonzero constants. If the known zeros are introduced into $G^{(k)}$, much less storage is required for $G^{(k)}$ than for the full matrix $H^{(k)}$. Also, if $G^{(k)}$ has a band structure, (2) need not be a great deal more time-consuming than (5). Moreover, when many of the elements of the Jacobian are known, the number of secondary conditions on $G^{(k+1)}$ can be greatly reduced, so that G converges more rapidly to the Jacobian. This requires a simple modification of (4) which will now be described.

II. Modification When the Jacobian is Sparse. The ith row $g_i^{(k)}$ of $G^{(k)}$ represents an approximation to the gradient of the ith function component f_i. When $n - r_i$ components of g_i are known constants, one first imposes the condition that these components shall remain unchanged in the Jacobian revision; the remaining choices have to be made on the basis of the remaining r_i coordinate directions.

Designate by $\hat{p}^{(k)}$ the column vector derived from $p^{(k)}$ by setting $p_j^{(k)}$ to zero whenever the corresponding element of g_i is a known constant. Note that $\hat{p}^{(k)}$ is dependent on i. Also let $\bar{g}_i$ be the row vector derived from g_i by setting its unknown elements to zero.

The known components of g_i account for a change $t^{(k)} \bar{g}_i p^{(k)}$ in f_i at the $(k + 1)$st step. The remainder of the change, $f_i^{(k+1)} - f_i^{(k)} - t^{(k)} \bar{g}_i p^{(k)}$, must then be attributed to the unknown components. Thus the primary condition on $g_i^{(k+1)}$, restricted to r_i-space, becomes

$$t^{(k)} g_i^{(k+1)} \hat{p}^{(k)} = f_i^{(k+1)} - f_i^{(k)} - t^{(k)} \bar{g}_i p^{(k)} , \qquad i = 1, 2, \cdots, n . \tag{6}$$

This is, in fact, identical to the usual primary condition $t^{(k)} G^{(k+1)} p^{(k)} = f^{(k+1)} - f^{(k)}$ because $g_i^{(k+1)} \hat{p}^{(k)} + \bar{g}_i p^{(k)} = g_i^{(k+1)} p^{(k)}$. The secondary condition is similarly obtained by restricting the usual secondary condition to the r_i-space corresponding to the unknown elements of g_i:

$$g_i^{(k+1)} \hat{q} = g_i^{(k)} \hat{q} , \qquad i = 1, 2, \cdots, n \tag{7}$$

where $\hat{q}$ satisfies $\hat{p}^{(k)T} \hat{q} = 0$. This does not reduce to the usual condition $G^{(k+1)} q = G^{(k)} q$.

It is easily verified that (6) and (7) are satisfied by the exact row-by-row analogue of (4), i.e.,

$$g_i^{(k+1)} = g_i^{(k)} + \frac{[f_i^{(k+1)} - (1 - t^{(k)}) f_i^{(k)}] \hat{p}^{(k)T}}{t^{(k)} \hat{p}^{(k)T} \hat{p}^{(k)}} , \qquad i = 1, 2, \cdots, n . \tag{8}$$

III. Example. A set of equations used by Broyden to test his methods is

$$\begin{aligned} f_1 &= -(3 + \alpha x_1)x_1 + 2x_2 - \beta\,, \\ f_i &= x_{i-1} - (3 + \alpha x_i)x_i + 2x_{i+1} - \beta\,, \qquad i = 2, 3, \cdots, n-1\,, \\ f_n &= x_{n-1} - (3 + \alpha x_n)x_n - \beta\,. \end{aligned} \tag{9}$$

These equations are also suitable for illustrating the present variant, if the zero entries in the Jacobian are regarded as known. The parameter values chosen were $\alpha = -.5$; $\beta = 1$; $n = 5, 10, 20$; $x_i^{(0)} = -1$ for all i. Both the unit matrix and a difference approximation based on a differencing interval of .001 were used for $G^{(0)}$. Broyden's mean convergence rate

$$R = \frac{1}{m} \ln \frac{N_1}{N_m}\,, \tag{10}$$

where N_1 and N_m are the initial and final Euclidean norms of f, was computed in each case. m has been redefined as the total number of function *component* evaluations divided by n. In this way m reflects the fact that one can take advantage of the Jacobian's sparseness in computing $G^{(0)}$ by differencing.

Results for the present method ("modified Jacobian revision"), Broyden's 1/fsr method ("basic Jacobian revision"), and the constant matrix method ("no Jacobian revision") are shown in Tables I — III.

TABLE I. $n = 5$

Nature of $G^{(0)}$	Method	N_1	N_m	m	R
Difference	Mod. Jac. rev.	1.803	9.592×10^{-7}	8	1.901
Approximation	Basic Jac. rev.	1.803	9.657×10^{-8}	9	1.947
to Jacobian	No Jac. rev.	1.803	2.149×10^{-7}	11	1.504
	Mod. Jac. rev.	1.803	3.272×10^{-7}	20	0.776
Unit	Basic Jac. rev.	1.803	7.262×10^{-7}	23	0.640
Matrix	No Jac. rev.	1.803	5.920×10^{-7}	73	0.205

TABLE II. $n = 10$

Nature of $G^{(0)}$	Method	N_1	N_m	m	R
Difference	Mod. Jac. rev.	2.121	1.408×10^{-7}	9	1.878
Approximation	Basic Jac. rev.	2.121	2.098×10^{-7}	11	1.493
to Jacobian	No Jac. rev.	2.121	5.404×10^{-7}	15	1.026
	Mod. Jac. rev.	2.121	1.707×10^{-7}	26	0.628
Unit	Basic Jac. rev.	2.121	4.391×10^{-7}	61	0.252
Matrix	No Jac. rev.	2.121	8.363×10^{-7}	88	0.168

The results indicate that modified Jacobian revision becomes increasingly desirable as n is increased, particularly if the initial approximation to the Jacobian is poor.

The modification may also be useful when the Jacobian is full, but most of the

TABLE III. $n = 20$

Nature of $G^{(0)}$	Method	N_1	N_m	m	R
Difference	Mod. Jac. rev.	2.646	3.130×10^{-7}	9	1.792
Approximation	Basic Jac. rev.	2.646	3.846×10^{-7}	12	1.323
to Jacobian	No Jac. rev.	2.646	3.473×10^{-7}	19	0.838
	Mod. Jac. rev.	2.646	3.402×10^{-7}	25	0.635
Unit	Basic Jac. rev.	2.646	9.850×10^{-7}	118	0.125
Matrix	No Jac. rev.	2.646	9.222×10^{-7}	97	0.153

entries are easily computed constants. In this case, however, storage space is not economized and the solution of (2) may be time-consuming.

Institute for Aerospace Studies
University of Toronto
Toronto 5, Ontario, Canada

1. C. G. BROYDEN, "A class of methods for solving nonlinear simultaneous equations," *Math. Comp.*, v. 19, 1965, pp. 577–593. MR **33** #6825.

2. E. M. ROSEN, "A Review of Quasi-Newton Methods in Nonlinear Equation Solving and Unconstrained Optimization," *Proc. Twenty-first Nat. Conf. ACM*, Thompson, Washington, D. C., 1966, pp. 37–41.

3. F. J. ZELEZNIK, "Quasi-Newton methods for nonlinear equations," *J. Assoc. Comput. Mach.*, v. 15, 1968, pp. 265–271.

4. J. G. P. BARNES, "An algorithm for solving nonlinear equations based on the secant method," *Comput. J.*, v. 8, 1965, pp. 66–72. MR **31** #5330.