Computers & Structures Vol 13, pp 73–81
Printed in Great Britain

0045-7949/81/010073-09$02 00/0

COMPUTATIONAL STRATEGIES FOR THE SOLUTION OF LARGE NONLINEAR PROBLEMS VIA QUASI-NEWTON METHODS

M. GERADIN†, S. IDELSOHN‡ and M. HOGGE §
Aerospace Laboratory of the University of Liège, Rue E. Solvay, 21, B-4000 Liege, Belgium

(*Received* 9 *May* 1980)

Abstract—The usefulness of quasi-Newton methods for the solution of nonlinear systems of equations is demonstrated. After a review of the Newton iterative method, several quasi-Newton updates are presented and tested. Special attention is devoted to the solution of large sparse systems of equations such as those issued from spatial discretization of continua by finite elements.

The numerical examples presented comprise static and dynamic analyses of geometrical, material and combined nonlinear structural problems and a model fluid flow problem with different levels of nonlinearity. All the results are assorted with a complete discussion of the different methods used, of the convergence rates and of the associated computer costs.

From the present studies, it can be concluded that computational costs for the solution of large nonlinear systems of equations can be reduced drastically by using convenient quasi-Newton updates or by adequate combined Newton/quasi-Newton strategies.

The best known method for solving large systems of nonlinear equations iteratively is Newton's method, sometimes modified so as to improve its computational efficiency. Davidon, for the minimization problem, and Broyden, for systems of equations, introduced in the early sixties new methods which, although iterative in nature, were quite unlike any other one in use at the time [1]. This new class of algorithms has been called by the names quasi-Newton, variable metric, secant, update or modification methods, the basic idea being to replace the costly evaluation of the effective Jacobian or Hessian matrix involved by some economically obtained approximation.

In recent years there has been a proliferation of quasi-Newton methods applicable to the unconstrained minimization problem. The same is not true for solving nonlinear equations: according to [1], the only quasi-Newton method that has been seriously used to solve nonlinear equations is the one proposed by Broyden. In the context of nonlinear structural and continuum analysis using the finite element method, the application of quasi-Newton methods for the solution of the associated systems of equations has been suggested for the first time by Strang and Mathies [2]. Since then a growing amount of literature has developed on the subject through various nonlinear finite element applications [3–8, 10]. At first sight, quasi-Newton methods seem to be particularly attractive to dynamic analysis where the unknown increments are necessarily kept small in order to achieve a sufficient accuracy in the time-marching procedure [3, 4, 7]. In this paper it will be shown that various quasi-Newton updates are also of interest for static nonlinear problems, either of structural or continuum nature and that important savings can be obtained on the total cost of such problems too. It will also be demonstrated how, in the context of nonlinear analysis using the finite element method, advantage can be taken of the sparse pattern of the structural matrices to achieve an optimum implementation of the method.

The remaining of the paper is divided into five sections: in the second one, we recall the basic Newton method for the solution of systems of nonlinear equations and the composition of such systems issued from finite element structural and fluid problems. In Section 3, the most common quasi-Newton updating formulas are described, including rank-one and rank-two updates. Approximations to the inverse Jacobian are presented together with the concept of line search that can be associated with the iterative procedure. A brief outline of stability and convergence properties of Newton and quasi-Newton procedures is given. Section 4 deals with the practical implementation of the updating method in relation with sparse finite element systems of equations. Coupling between Newton and quasi-Newton methods is proposed for highly nonlinear problems and a shifting strategy is presented and tested. Several numerical applications are described in Section 5 where nonlinear structural and fluid flow problems with different level of nonlinearity are analyzed by Newton method and various quasi-Newton updates. The final section draws the conclusions of the analysis and present research directions that should be explored in the future.

2. NEWTON METHODS

Consider the problem of finding a solution to the system of equations

$$\mathbf{r}(\mathbf{q}) = \mathbf{0} \tag{1}$$

where $\mathbf{r}$ and $\mathbf{q}$ are n-dimensional vectors.

Newton's method of solution can be derived by assuming that we have an approximation $\bar{\mathbf{q}}$ to $\mathbf{q}$, and

‡Visiting scientist from CONICET; Professor, Univ. Nacional, Rosario, Rep. Argentina.
§Senior Assistant.

that in the neighbourhood of $\tilde{\mathbf{q}}$ the linear mapping

$$\mathbf{r}_L(\mathbf{q}) = \mathbf{r}(\tilde{\mathbf{q}}) + \frac{\partial \mathbf{r}(\tilde{\mathbf{q}})}{\partial \mathbf{q}} (\mathbf{q} - \tilde{\mathbf{q}}) \tag{2}$$

is a good approximation to $\mathbf{r}(\mathbf{q})$. A presumably better approximation to $\mathbf{q}$ can then be obtained by equating (2) to zero.

Thus, Newton's method takes an initial approximation $\mathbf{q}_0$ to $\mathbf{q}$, and attempts to improve it iteratively by

$$\mathbf{q}_{k+1} = \mathbf{q}_k - \mathbf{S}_k^{-1} \mathbf{r}_k, \quad k = 0, 1, \ldots \tag{3}$$

taking $\mathbf{r}_k = \mathbf{r}(\mathbf{q}_k)$ and with the definition of the Jacobian matrix

$$\mathbf{S}_k = \mathbf{S}(\mathbf{q}_k) = \left(\frac{\partial \mathbf{r}}{\partial \mathbf{q}} \right)_{\mathbf{q}_k} \tag{4}$$

The finite element discretization of static nonlinear structural problems leads to systems of type (1) with

$$\mathbf{r}(\mathbf{q}) = \mathbf{K}(\mathbf{q})\mathbf{q} - \mathbf{g} = \mathbf{0} \tag{5}$$

where $\mathbf{q}$ is the vector of the unknown displacements and $\mathbf{g}$ the vector of the applied nodal loads.

Nonlinearities arise in general from material behavior or adaptation of the geometry; they are implicitly contained in the internal forces $\mathbf{K}(\mathbf{q})\mathbf{q}$ which result from the spatial integration of the internal stresses σ

$$\mathbf{K}(\mathbf{q})\mathbf{q} = \int_V \mathbf{B}^T \sigma \, dV \tag{6}$$

where $\mathbf{K}(\mathbf{q})$ is the structural stiffness matrix. The Jacobian matrix (4) is in this case the tangent stiffness matrix

$$\mathbf{S}(\mathbf{q}) = \mathbf{K}^t(\mathbf{q}) = \frac{\partial}{\partial \mathbf{q}} [\mathbf{K}(\mathbf{q})\mathbf{q}] \tag{7}$$

plus a contribution of the external forces $\partial \mathbf{g}/\partial \mathbf{q}$ if these forces are dependent upon geometry changes: this term is generally omitted to preserve the symmetry of the Jacobian matrix.

In nonlinear structural dynamics, the effective loads in (5) are the difference between externally applied loads and inertia forces, so that the spatially discretized systems read

$$\mathbf{r}(\mathbf{q}) = \mathbf{K}(\mathbf{q})\mathbf{q}(t) + \mathbf{M}\ddot{\mathbf{q}}(t) - \mathbf{g}(t) = \mathbf{0}. \tag{8}$$

The Jacobian matrix of Newton's method is thus not only a function of the tangent stiffness matrix $\mathbf{K}^t$ but also of the temporal integration scheme used in the response. If such schemes are limited to those contained in Newmark's formula:

$$\begin{aligned} \dot{\mathbf{q}}_{i+1} &= \dot{\mathbf{q}}_i + (1-\gamma)h\ddot{\mathbf{q}}_i + \gamma h \ddot{\mathbf{q}}_{i+1} \\ \mathbf{q}_{i+1} &= \mathbf{q}_i + h\dot{\mathbf{q}}_i + (\tfrac{1}{2} - \beta)h^2 \ddot{\mathbf{q}}_i + \beta h^2 \ddot{\mathbf{q}}_{i+1} \end{aligned} \tag{9}$$

where the subscript i denotes the ith time-step, h the time-step size and β, γ the Newmark's parameters, the Jacobian matrix becomes

$$\mathbf{S}(\mathbf{q}) = \mathbf{K}^t(\mathbf{q}) + \frac{1}{\beta h^2} \mathbf{M} + \frac{\partial \mathbf{g}}{\partial \mathbf{q}}. \tag{10}$$

The last term appears only for geometry-dependent external forces and again is usually omitted for symmetry purposes.

In viscous incompressible fluid flow problems [8], the system of discretized nonlinear equations of motion reads

$$\mathbf{r}(\mathbf{q}) = [\mathbf{K} + \mathbf{C}(\mathbf{q})]\mathbf{q} - \mathbf{g} = \mathbf{0} \tag{11}$$

where $\mathbf{K}$ and $\mathbf{C}(\mathbf{q})$ are the diffusive and convective matrices, $\mathbf{q}$ is the vector of unknown nodal velocities and pressures and $\mathbf{g}$ is the vector representing "virtual work" equivalent body forces and surface tractions. Note that only $\mathbf{K}$ is symmetrical and unknown-independent, so that the Jacobian matrix

$$\mathbf{S}(\mathbf{q}) = \mathbf{K} + \frac{\partial}{\partial \mathbf{q}} [\mathbf{C}(\mathbf{q})\mathbf{q}] \tag{12}$$

is always unsymmetrical.

3. QUASI-NEWTON METHODS

3.1 *Direct updates*

The major expense in Newton's method is the calculation of the Jacobian $\mathbf{S}(\mathbf{q}_k)$ and its inversion.

In contrast, quasi-Newton methods consist in deriving an approximation $\mathbf{G}$ to the Jacobian by evaluating $\mathbf{r}(\mathbf{q})$ at two successive points $\tilde{\mathbf{q}}$ and $\mathbf{q}$. Indeed if we expand $\mathbf{r}$ around $\mathbf{q}$ by Taylor's theorem

$$\mathbf{S}(\mathbf{q})\mathbf{d} = \mathbf{r}(\mathbf{q}) - \mathbf{r}(\tilde{\mathbf{q}}) + \Delta \mathbf{r} \tag{13}$$

where $\mathbf{d} = \mathbf{q} - \tilde{\mathbf{q}}$ and $\Delta \mathbf{r} \to \mathbf{0}$ as $\mathbf{q} \to \tilde{\mathbf{q}}$. When neglecting the term $\Delta \mathbf{r}$ in eqn (13), we obtain

$$\mathbf{G}(\mathbf{q})\mathbf{d} = \mathbf{r}(\mathbf{q}) - \mathbf{r}(\tilde{\mathbf{q}}) = \mathbf{y} \tag{14}$$

which is called the quasi-Newton equation. It is exact if $\mathbf{r}$ derives from a quadratic functional and nearly exact in a sufficiently small neighbourhood of the solution if that functional is not quadratic but strictly convex. Therefore it is desirable that any matrix candidate to $\mathbf{G}$ satisfies eqn (14). It is also desirable that the approximation $\mathbf{G}$ to $\mathbf{S}$ be easily computable from $\mathbf{G}(\tilde{\mathbf{q}})$, $\mathbf{y}$ and $\mathbf{d}$ by adding to $\mathbf{G}(\tilde{\mathbf{q}})$ a correction matrix which depends upon the above quantities while satisfying eqn (14). The simplest among such relations is the single-rank update

$$\mathbf{G} = \tilde{\mathbf{G}} + \frac{[\mathbf{y} - \tilde{\mathbf{G}}\mathbf{d}]\mathbf{u}^T}{\mathbf{u}^T\mathbf{d}} \tag{15}$$

where $\mathbf{u}$ is an arbitrary vector such that $\mathbf{u}^T\mathbf{d} \neq 0$.

Quasi-Newton iteration consists thus, given initial arbitrary $\mathbf{q}_0$ and $\mathbf{G}_0$, to calculate a new direction by eqn (3) and next, to generate a new matrix $\mathbf{G}_{k+1}$ by eqn (15), i.e.

(1) $\mathbf{d}_k = -\mathbf{G}_k^{-1}\mathbf{r}_k$ (16)

(2) Compute $\mathbf{y}_k, \mathbf{u}_k \quad k = 0, 1, \ldots$

(3) $\mathbf{G}_{k+1} = \mathbf{G}_k + \dfrac{(\mathbf{y}_k - \mathbf{G}_k\mathbf{d}_k)\mathbf{u}_k^T}{\mathbf{u}_k^T\mathbf{d}_k}$ (17)

Several rank-one updates are possible, Obviously it is highly desirable that $\mathbf{u}$ depends only on $\mathbf{d}$, $\mathbf{y}$ and $\tilde{\mathbf{G}}$. Broyden proposes $\mathbf{u} = \mathbf{d}$, so that

$$\mathbf{G}_B = \tilde{\mathbf{G}} + \frac{[\mathbf{y} - \tilde{\mathbf{G}}\mathbf{d}]\mathbf{d}^T}{\mathbf{d}^T\mathbf{d}}. \tag{18}$$

It has been shown that in this way $\mathbf{G}$ is the "closest" to $\mathbf{S}$ when measuring the distance by the Frobenius norm [1]. Note that Broyden's update is unsymmetric and hence does not preserve the eventual symmetry of $\tilde{\mathbf{G}}$.

For symmetric systems of equations, Davidon suggests to use the direction $\mathbf{u} = \mathbf{y} - \tilde{\mathbf{G}}\mathbf{d}$. The new corrective matrix becomes

$$\mathbf{G}_D = \tilde{\mathbf{G}} + \frac{(\mathbf{y} - \tilde{\mathbf{G}}\mathbf{d})(\mathbf{y} - \tilde{\mathbf{G}}\mathbf{d})^T}{(\mathbf{y} - \tilde{\mathbf{G}}\mathbf{d})^T\mathbf{d}} \tag{19}$$

which insures the symmetry of the successive approximation matrices. It is known to dispense with accurate line searches [12] but there is no guarantee that $\mathbf{G}_D$ is positive definite even if $\tilde{\mathbf{G}}$ exhibits this property.

Rank-two formulas are often proposed, for instance the Powell symmetric Broyden update (PSB), the Brodie update, etc... Several of them, in addition to preserving symmetry, have the property of safeguarding positive definite matrices. Among them, the most widely used are the Davidon–Fletcher–Powell update (DFP):

$$\mathbf{G}_{\mathrm{DPF}}=\left(\mathbf{I}-\frac{\mathbf{y}\mathbf{d}^T}{\mathbf{y}^T\mathbf{d}}\right)\tilde{\mathbf{G}}\left(\mathbf{I}-\frac{\mathbf{d}\mathbf{y}^T}{\mathbf{y}^T\mathbf{d}}\right)+\frac{\mathbf{y}\mathbf{y}^T}{\mathbf{y}^T\mathbf{d}} \tag{20}$$

and the Broyden–Fletcher–Goldfard–Shanno formula (BFGS):

$$\mathbf{G}_{\mathrm{BFGS}}=\tilde{\mathbf{G}}+\frac{\mathbf{y}\mathbf{y}^T}{\mathbf{y}^T\mathbf{d}}-\frac{\tilde{\mathbf{G}}\mathbf{y}\mathbf{y}^T\tilde{\mathbf{G}}}{\mathbf{d}^T\tilde{\mathbf{G}}\mathbf{d}}. \tag{21}$$

Both formulas satisfy the quasi-Newton equation (14). In the same manner as for eqn (17), the iterative procedure is obtained by setting in eqns (20) and (21) $\mathbf{G}=\mathbf{G}_{k+1}$, $\tilde{\mathbf{G}}=\mathbf{G}_k$, $\mathbf{y}=\mathbf{y}_k$, $\mathbf{d}=\mathbf{d}_k$.

3.2 *Inverse updates*

To solve the linear problem (16) at least expense, it is convenient to obtain directly from (15) the new approximation to the inverse Jacobian. This is possible using the property that [12]:

$$(\mathbf{A}\quad \alpha\mathbf{u}\mathbf{v}^T)^{-1}=\mathbf{A}^{-1}-\beta\mathbf{x}\mathbf{z}^T \tag{22}$$

with $\mathbf{x}=\mathbf{A}^{-1}\mathbf{u}$, $\mathbf{z}=\mathbf{A}^{-T}\mathbf{v}$ and $\beta=\alpha(1+\alpha\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u})^{-1}$. Thus, the general rank-one update (15) becomes

$$\mathbf{G}^{-1}=\tilde{\mathbf{G}}^{-1}+\frac{(\mathbf{d}-\tilde{\mathbf{G}}^{-1}\mathbf{y})\mathbf{v}^T}{\mathbf{v}^T\mathbf{y}} \tag{23}$$

for an arbitrary vector $\mathbf{v}$, with $\mathbf{v}^T\mathbf{y}\neq 0$. Broyden's update is obtained when $\mathbf{v}=\tilde{\mathbf{G}}^{-T}\mathbf{d}$, and Davidon's symmetric update when $\mathbf{v}=\mathbf{d}-\tilde{\mathbf{G}}^{-1}\mathbf{y}$, i.e.

$$\mathbf{G}_B^{-1}=\tilde{\mathbf{G}}^{-1}+\frac{(\mathbf{d}-\tilde{\mathbf{G}}^{-1}\mathbf{y})\mathbf{d}^T\tilde{\mathbf{G}}^{-1}}{\mathbf{d}^T\tilde{\mathbf{G}}^{-1}\mathbf{y}} \tag{23'}$$

$$\mathbf{G}_D^{-1}=\tilde{\mathbf{G}}^{-1}+\frac{(\mathbf{d}-\tilde{\mathbf{G}}^{-1}\mathbf{y})(\mathbf{d}-\tilde{\mathbf{G}}^{-1}\mathbf{y})^T}{(\mathbf{d}-\tilde{\mathbf{G}}^{-1}\mathbf{y})^T\mathbf{y}} \tag{23''}$$

All the rank-two updates may also be transformed in the same manner to obtain directly the inverse matrix $\mathbf{G}^{-1}$, yielding

$$\mathbf{G}_{\mathrm{DPF}}^{-1}=\tilde{\mathbf{G}}^{-1}+\frac{\mathbf{d}\mathbf{d}^T}{\mathbf{d}^T\mathbf{y}}-\frac{\tilde{\mathbf{G}}^{-1}\mathbf{y}\mathbf{y}^T\tilde{\mathbf{G}}^{-1}}{\mathbf{y}^T\tilde{\mathbf{G}}^{-1}\mathbf{y}} \tag{24}$$

and

$$\mathbf{G}_{\mathrm{BFGS}}^{-1}=\left(\mathbf{I}-\frac{\mathbf{d}\mathbf{y}^T}{\mathbf{y}^T\mathbf{d}}\right)\tilde{\mathbf{G}}^{-1}\left(\mathbf{I}-\frac{\mathbf{y}\mathbf{d}^T}{\mathbf{y}^T\mathbf{d}}\right)+\frac{\mathbf{d}\mathbf{d}^T}{\mathbf{y}^T\mathbf{d}}. \tag{25}$$

It is useful to note that DFP and BFGS updates are related by the transformation

$$\mathbf{d}\rightarrow\mathbf{y};\quad \mathbf{G}\rightarrow\mathbf{G}^{-1}$$

(see eqns 20–25 and 21–24); these updates are called "dual" or "complementary" updates [1].

3.3 *Line search*

In order to improve the convergence rate, an optimal step length σ_k in the direction determined by eqn (16) can be evaluated such as to cancel the projection of the residual vector in that direction, i.e.

$$\delta=\mathbf{d}_k^T\mathbf{r}(\mathbf{q}_k+\sigma_k\mathbf{d}_k)=0 \tag{26}$$

and then

$$\mathbf{q}_{k+1}=\mathbf{q}_k+\sigma_k\mathbf{d}_k. \tag{27}$$

This is an expensive operation since it may involve numerous evaluations of the residual vector to achieve great accuracy. One may expect, however, that the more accurate the line search is, the better is the chance of achieving convergence in a minimum number of iterations.

In Ref. [3], the authors report that satisfactory rate of convergence is obtained without line search when

$$|\mathbf{d}_k^T\mathbf{r}(\mathbf{q}_k+\mathbf{d}_k)|\leqslant\eta|\mathbf{d}_k^T\mathbf{r}(\mathbf{q}_k)|\quad\text{with}\quad\eta=0.5 \tag{28}$$

This has been confirmed by the numerical experiments described in the present paper. When eqn (28) is not satisfied, successive linear interpolations may be performed in order to determine the optimal length σ_k such that

$$|\mathbf{d}_k^T\mathbf{r}(\mathbf{q}_k+\sigma_k\mathbf{d}_k)|\leqslant\eta|\mathbf{d}_k^T\mathbf{r}(\mathbf{q}_k)|. \tag{29}$$

Strang [13] reports that the choice $\eta=0.9$ should be a good compromise for accuracy vs cost of the line search, especially for fluid problems.

3.4 *Stability and convergence of quasi-Newton methods* [1, 12]

Under the assumption that $\mathbf{r}$ is continuously differentiable in an open convex set C pertaining to R^n and that there is a solution $\mathbf{q}^*$ to eqn (1) for which $\mathbf{S}(\mathbf{q}^*)$ is nonsingular, then Newton algorithm is known to possess a domain of attraction A, which is an open set containing $\mathbf{q}^*$ such that for any $\mathbf{q}_0\in A$ the Newton iterates are well-defined, remain in A and converge to $\mathbf{q}^*$. This implies that if Newton iterates pertain to A, they will remain in A and insures in some sense the stability of the iterative procedure.

Moreover, there exists a sequence $\{\alpha_k\}$ which converges to zero and such that

$$\|\mathbf{q}_{k+1}-\mathbf{q}^*\|\leqslant\alpha_k\|\mathbf{q}_k-\mathbf{q}^*\|,\quad k=0,\ldots \tag{30}$$

where $\|\cdot\|$ stands for the L_2 vector norm $\|x\|=(\Sigma_i x_i^2)^{1/2}$ or the consistent matrix norm. This result is known as superlinear convergence. This is more than linear convergence for which with $\alpha\in(0,1)$

$$\|\mathbf{q}_{k+1}-\mathbf{q}^*\|\leqslant\alpha\|\mathbf{q}_k-\mathbf{q}^*\|\quad k\geqslant k_0 \tag{31}$$

and guarantees only that the error will eventually be decreased by the factor $\alpha<1$. If in addition $\mathbf{r}$ satisfies a Lipshitz condition at $\mathbf{q}^*$, i.e. if there is a constant β such that

$$\|\mathbf{S}(\mathbf{q})-\mathbf{S}(\mathbf{q}^*)\|\leqslant\beta\|\mathbf{q}-\mathbf{q}^*\|,\quad \mathbf{q}\in C \tag{32}$$

then second order or quadratic convergence is obtained, i.e. there is a constant γ such that

$$\|\mathbf{q}_{k+1}-\mathbf{q}^*\|\leqslant\gamma\|\mathbf{q}_k-\mathbf{q}^*\|^2,\quad k=0,\ldots \tag{33}$$

which is a well-known property of Newton method seldom obtained in practice due to requirement (32).

Any quasi-Newton iteration generated by eqn (16):

$$\mathbf{q}_{k+1}=\mathbf{q}_k-\mathbf{G}_k^{-1}\mathbf{r}_k\quad k=0,1,\ldots$$

will be locally convergent at $\mathbf{q}^*$, i.e. $\{\mathbf{q}_k\}$ is well-defined and converges to $\mathbf{q}^*$, if there is an $\varepsilon>0$ and a $\delta>0$ such

that whenever $\mathbf{q}_0 \in A$ and $\mathbf{G}_0 \in A_M$ (A_M is the set of the various Jacobian approximations which might be used in the iterative process) they satisfy

$$\begin{aligned} &\|\mathbf{q}_0 - \mathbf{q}^*\| < \varepsilon \\ &\|\mathbf{G}_0 - \mathbf{S}(\mathbf{q}^*)\| < \delta. \end{aligned} \tag{34}$$

Now such a sequence converges superlinearly to $\mathbf{q}^*$ if and only if

$$\lim_{k\to\infty} \frac{\|[\mathbf{G}_k - \mathbf{S}(\mathbf{q}^*)](\mathbf{q}_{k+1} - \mathbf{q}_k)\|}{\|\mathbf{q}_{k+1} - \mathbf{q}_k\|} = 0. \tag{35}$$

An equivalent but more geometric formulation of this condition is that it requires $\mathbf{d}_k$ in the iterative method to asymptotically approach the Newton correction

$$\mathbf{d}_k^N = -\mathbf{S}_k^{-1}\mathbf{r}_k$$

in both magnitude and direction. This follows from the fact that

$$\mathbf{d}_k - \mathbf{d}_k^N = \mathbf{d}_k + \mathbf{S}_k^{-1}\mathbf{r}_k = \mathbf{S}_k^{-1}[\mathbf{S}_k - \mathbf{G}_k]\mathbf{d}_k$$

and thus eqn (35) is equivalent with

$$\lim_{k\to\infty} \frac{\|\mathbf{d}_k - \mathbf{d}_k^N\|}{\|\mathbf{d}_K\|} = 0. \tag{36}$$

This characteristic of local and superlinear convergence is shared by Broyden's method eqn (15) and its modification by Schubert for sparse Jacobians (see next section). Convergence of the other direct updates or of the inverse updates has only been proved in the frame of unconstrained minimization, in which case $\mathbf{r}$ is the gradient vector (and $\mathbf{S}$ the Hessian matrix) of an objective function [1, 12].

4. COMPUTATIONAL IMPLEMENTATION OF QUASI-NEWTON UPDATES

The natural way of performing quasi-Newton corrections to the Jacobian matrix is in the form implied by the updates of Section 3, i.e. by adding a correction matrix to the previous approximation or by implementing a correction in product form [2].

Inspection of the procedure in a finite element context, where most of the elements of the Jacobian matrix $\mathbf{S}$ are known to be zero owing to the topology of the discretization mesh and where a frontal solution technique is used with substructuring to perform block elimination, reveals that careful attention has to be devoted to the correction procedure in order to preserve the sparse pattern of the true Jacobian. Schubert [9] has proposed a variant of Broyden's unsymmetric update in which $\mathbf{G}_{k+1}$ is forced to have the same sparsity as $\mathbf{S}$. Such a technique has been developed for symmetric correction in [7] since we expect an optimal correction procedure for symmetric systems using symmetric updates.

The procedure is however rather heavy to handle and another way of performing the quasi-Newton update [2, 7] consists of applying the correction on the direction of search $\mathbf{d}$ instead of modifying the matrix $\mathbf{G}$ itself. In fact, using the inverse update as described by eqn (23), at the kth iteration, $\mathbf{G}^{-1}$ can be written as

$$\mathbf{G}_k^{-1} = \mathbf{G}_0^{-1} + \sum_{i=0}^{k} \beta_i \mathbf{v}_i \mathbf{v}_i^T. \tag{37}$$

For instance, for Davidon's update, eqn (23'), we have $\mathbf{v}_i = \sigma_i \mathbf{d}_i - \mathbf{G}_i^{-1}\mathbf{y}_i$ and $\beta_i = [\sigma_i \mathbf{d}_i - \mathbf{G}_i^{-1}\mathbf{y}_i)^T \mathbf{y}_i]^{-1}$. If at each iteration the correction vector $\mathbf{v}_i$ and coefficient β_i are stored on auxiliary memory, the kth direction can be obtained from (16) as

$$\mathbf{d}_k = -(\mathbf{G}_0^{-1} + \sum_{i=0}^{k} \beta_i \mathbf{v}_i \mathbf{v}_i^T)\mathbf{r}(\mathbf{q}_k). \tag{38}$$

The new correction vector for Davidon's update is then

$$\mathbf{v}_k = \sigma_k \mathbf{d}_k - \mathbf{G}_0^{-1}\mathbf{y}_k - \sum_{i=0}^{k} \beta_i \mathbf{v}_i \mathbf{v}_i^T \mathbf{y}_k. \tag{39}$$

This procedure is also applied by Crisfield using only one correction vector at each iteration [5].

Computational efficiency of this updating technique stems from the fact that, if an initial sparse Jacobian $\mathbf{G}_0$ is given, it may be triangularized and stored only once. The successive products $\mathbf{G}_0^{-1}\mathbf{r}(\mathbf{q}_k) = \mathbf{d}_k^0$ needed in eqns (38) and (39) may easily be performed solving the triangularized system of equation

$$\mathbf{G}_0 \mathbf{d}_k^0 = \mathbf{r}(\mathbf{q}_k).$$

In this manner only the nonzero elements of $\mathbf{G}_0$ after Gauss elimination, the vectors $\mathbf{v}_i$ and the coefficients β_i have to be stored. When the number of correction vectors becomes too large [from our experience, say around 10 without exceeding this limit since convergence will not be reached later on], the algorithm may be restarted with the initial matrix $\mathbf{G}_0^{-1}$.

In practice a new problem should be attacked first with the quasi-Newton iteration procedure. If strong nonlinearities are present and require Newton method, this latter technique should be used for k iterations until the convergence test ε be reasonably approached (say $\|\mathbf{r}_k\| < 10^2\,\varepsilon$); then the iterative scheme should be shifted to the quasi-Newton update for the end of the solution procedure. This changing strategy is illustrated in the next section for fluid problems and requires obviously the simultaneous implementation of the two algorithms into the associated computer program. Such an implementation is symbolized on the flow chart of Fig. 1.

A last observation is about the theoretical profit that one can expect between Newton and quasi-Newton iteration. In Newton method, the computation and triangularization of S_k requires $0(n^3)$ arithmetic operations. In quasi-Newton method, for every iteration from the second, this expense is reduced to $0(n^2)$.

5. NUMERICAL APPLICATIONS

5.1 *Clamped spherical cap*

The first example considered is the nonlinear structural analysis of a clamped spherical cap submitted to a sudden pressure loading, and where geometric and material nonlinearities are simultaneously present. Its geometric and material properties are summarized on Fig. 2. This is a classical example taken from [10].

The structure is modelled with 8 axisymmetric cubic shell elements [14]. The resulting finite element model numbers 72 degrees of freedom. Only 3 Gauss points are used to integrate the constitutive law over the thickness: this relatively crude integration rule may be foreseen to generate oscillations in the numerical solution when plasticity develops.

Static analysis. The structure was first tested statically with the pressure load described on Fig. 2.

The iterative procedure is stopped when

$$\|\mathbf{r}_k\|/(\|\mathbf{g}\| + \|\mathbf{g}_{\text{int}}\|) \leq 10^{-4}$$

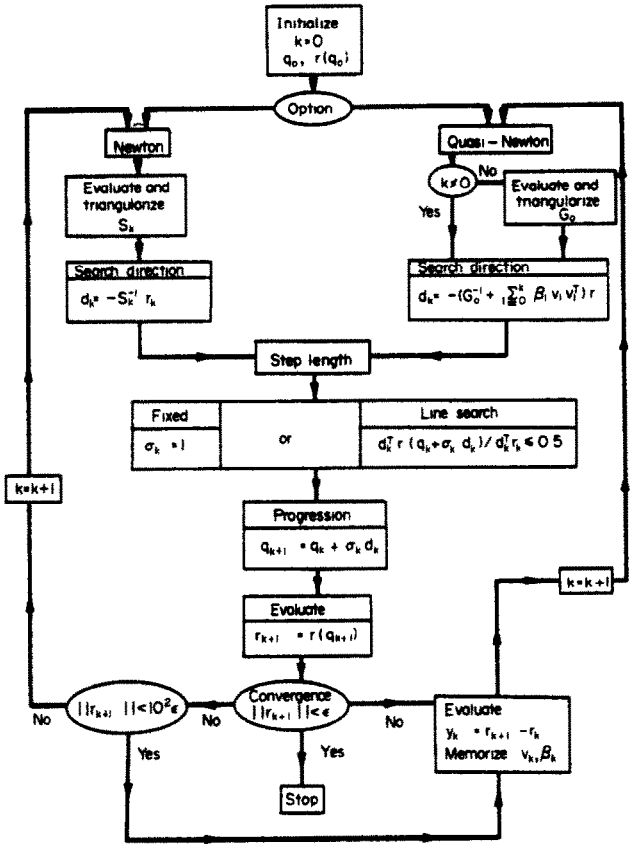


Fig. 1. Flow chart for Newton and quasi-Newton iterative procedures.

where $\mathbf{g}_{int}$ stands for $\mathbf{K}(\mathbf{q})\mathbf{q}$ in eqn (5).

The purpose of this analysis is the comparison in computer times necessary to obtain the solutions with Newton and quasi-Newton methods.

In this problem, the only external existing loads are the pressure loads. These loads introduce an unsymmetrical contribution to the Jacobian matrix which is neglected in the present analysis (see Section 2). Therefore, it seems natural that an unsymmetrical quasi-Newton update would give the best results.

Table 1 summarizes the performances obtained when using a Newton technique and quasi-Newton iterations, with the symmetrical Davidon's update and the unsymmetrical Broyden's update respectively.

In opposition to what was expected Davidon's updates give the best efficiency with a gain of 35% with respect to the standard Newton method. In fact, in this problem the geometrical nonlinearities are mild and thus the successive Jacobian matrices are nearly symmetrical. It is the reason why a symmetrical update is the most efficient.

Dynamic analysis. For the nonlinear response to step loading, time integration is performed with Newmark's scheme ($\beta = 1/4$, $\gamma = 1/2$) and a relatively large time step Δt of $1.5\ 10^{-5}$ sec has been adopted. Equilibrium iteration is now stopped within each time step n when

$$\|\mathbf{r}_k^n\| / (\|\mathbf{g}(t_n)\| + \|\mathbf{g}_{int}^n\|) \leq 10^{-3}.$$

Figure 3 displays the time history of the axial displacement at the apex of the cap for the following material and geometrical behaviors:

—linear elastic,

—elastic–plastic material, geometrically linear,

—material and geometrical nonlinearities simultaneously present.

Very little difference is observed in the numerical results with different methods of solution. For this example also, the only interest of the comparison lies in computer times and numbers of iterations to obtain the solution.

To solve this problem, the comparison has been made between Newton iterations and the quasi-Newton method using successively the Davidon and BFGS updates. The performances obtained to integrate the first 17 steps have been summarized in Table 2 for the combined nonlinear response.

The Newton solution corresponds to a strategy in which the stiffness is reevaluated at iterations 1, 2, 5 and 8 of each time step.

Quasi-Newton iterations have been performed with and without line search. Davidon's update has been tested using the vectorial correction (starting from $\mathbf{K}_0$ at each time step). The best results were obtained without line search.

The last two columns correspond to the BFGS updates with substructure correction (starting from the tangent stiffness matrix at each time step) [7]. One observes a significant increase in the number of iterations when the process is not restarted at each time step, due to the fact that the number of updates on $\mathbf{G}_0$ becomes excessive.

In spite of the small size of this problem (involving only 72 d.o.f.), the difference of computer costs between the reevaluation of stiffness (with Gauss elimination)

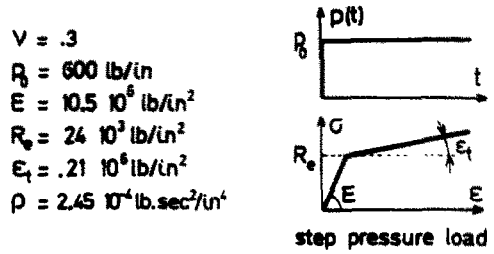
Table 1. Spherical cap static analysis. Efficiency of Newton and quasi-Newton iterations

		NEWTON	QUASI-NEWTON	
			G_D update	G_B update
1	Total number of Jacobian evaluations	4	1	1
2	Total number of residual evaluations	5	7	9
3	C.P.U.* Time per iteration	5.34	5.34/2.	5.34/2.2
4	Total number of iterations	5	6	8
5	Total C.P.U. Time*	32.	20.6	25.9

* IBM 370/158

Table 2. Spherical cap dynamic analysis. Efficiency of Newton and quasi-Newton iterations

	NEWTON	QUASI-NEWTON			
		G_D update	G_D update	G_{BFGS} update	G_{BFGS} update
Number of iteration per step	3.35	5.0	4.88	3.29	6.0
Total number of Jacobian evaluations	40	1	1	17	1
Total number of residual evaluations	57	102	111	73	135
Total number of Line Search	-	-	11	-	16
C.P.U. time per iteration	5.40	2.75	2.97	4.83	3.6
Total number of iterations	57	85	83	56	102
Total C.P.U. time	308.	233.6	247.	271.	368.



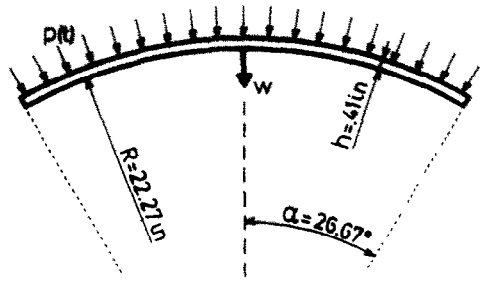


Fig. 2. Spherical cap submitted to step pressure loading.

and the calculation of the residual vector is yet significant. Quasi-Newton iteration is the most efficient procedure.

5.2 *Two-dimensional fluid flow*

In fluid flow problems the relation between the viscosity and the density of the fluid plays an important role in the nonlinear character of the solution. Thus similar problems with differents Reynolds number become nearly linear or largely nonlinear for lower or higher Reynolds number. This interacts with the convergence properties of the solution and is an easy way of testing the differents methods proposed here. The second example deals thus with the computation of the velocity profiles of two-dimensional fluid flow between two parallel walls.

A 4×4 isoparametric finite element mesh is used yielding a total of 129 d.o.f. Each element possesses a quadratic velocity field and a linear pressure field [8]. The boundary condition are represented on Fig. 4.

Table 3 shows the efficiency of the different methods used for $Re = 10$. In all cases the starting solution $\mathbf{q}_0$ corresponds to the Stokes solution, i.e. the solution of eqn

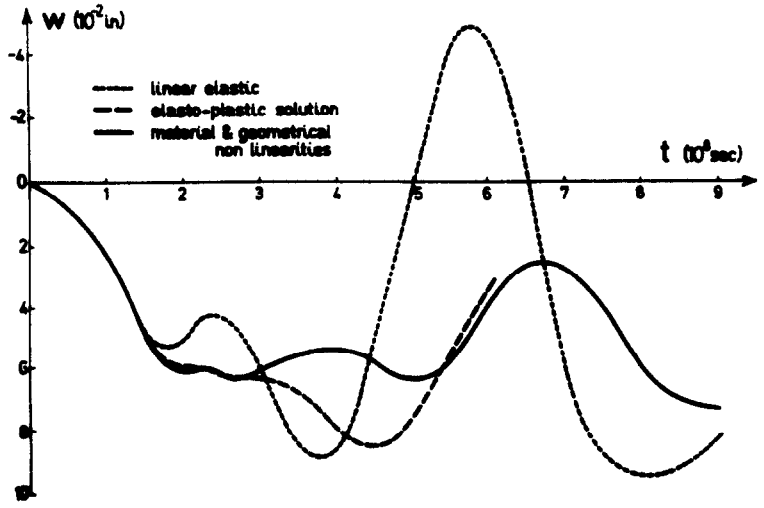


Fig. 3. Spherical cap, displacement W at apex node.

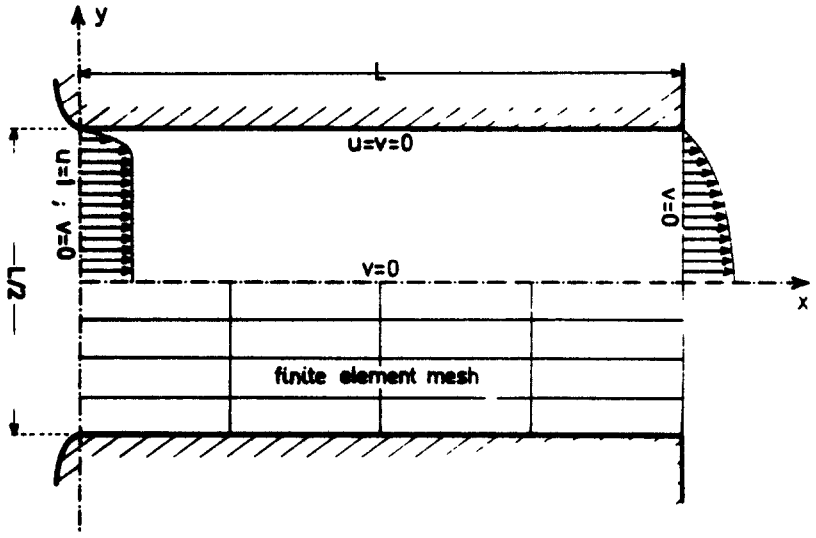


Fig. 4. 2 D-Flow between parallel walls.

Table 3. Fluid flow problem (Re = 10). Efficiency of Newton and quasi-Newton iterations

	NEWTON	QUASI-NEWTON	
		G_D update	G_B update
Total number of Jacobian evaluations	4	1	1
Total number of residual evaluations	4	4	3
C.P.U. Time per iteration	15.1	15.1/3.3	15.1/3.4
Number of iterations	4	3	2
Total C.P.U. time	68.8	35.2	31.4

(11) with $\mathbf{C}=\mathbf{0}$, and the iteration is stopped when

$$\|\mathbf{r}_k\|/(\|\mathbf{g}\|+\|\mathbf{g}_{int}\|) \quad 10^{-3}$$

where $\mathbf{g}$ includes the reactions to the imposed velocity field and $\mathbf{g} = [\mathbf{K}+\mathbf{C}(\mathbf{q})]\mathbf{q}$ (see Section 2).

Up to Reynolds number 10, quasi-Newton updates can be performed from a single Jacobian evaluation. They reveal to be quite competitive in comparison to standard Newton iterations. It is remarkable to note that for the present problem the need for an approxim-

Table 4. Fluid flow problem (Re = 100.). Efficiency of Newton and quasi-Newton iterations

	NEWTON	QUASI-NEWTON	
		G_D update	G_B update
Total number of Jacobian evaluations	10	6	6
Total number of residual evaluations	10	12	11
C.P.U. Time per iteration	14.	14./2.8	14./3.8
Number of iterations	10	11	10
Total C.P.U. Time	154.	111.	139.

ate unsymmetrical Jacobian matrix makes Broyden's update the most adequate.

For Reynolds number higher than 10, the problem becomes strongly nonlinear and a quasi-Newton method using directly the linear matrix **K** as initial matrix does not converge. Nevertheless, another strategy was successfully tested: during the first iterations the Jacobian matrix was evaluated and then, the processus is continued with quasi-Newton updates according to the procedure outlined in the flow chart of Fig. 1.

Table 4 shows the comparison of procedures for Re = 100. The above procedure is applied as follows: the first 6 iterations are done evaluating the Jacobian matrix, then the remaining 5 for the Davidon's update, or the remaining 4 for the Broyden's update, are done by quasi-Newton corrections of the last Jacobian matrix. This is the only way to keep quasi-Newton methods very competitive.

Unsymmetrical Broyden's updates exhibit a better convergence rate than Davidon's ones but are more expensive since the former require the solution of two linear systems of equations against only one for the latter [see eqns (23′)–(23″)]. In the present case the quasi-Newton procedure was started, in accordance with the flow diagram (Fig. 1), when

$$\|\mathbf{r}_k\|/(\|\mathbf{g}\| + \|\mathbf{g}_{int}\|) \leqslant 10^{-1}.$$

Similar conclusions can be drawn from other fluid flow applications [8].

CONCLUSIONS

The adequacy of various updating methods to solve nonlinear systems of equations of finite element structural and continuum mechanics applications has been demonstrated, and their implementation for sparse systems has been discussed.

Quasi-Newton methods converge almost always in a larger number of steps than an "optimal" Newton strategy. The former become thus competitive only when the cost of Jacobian evaluation is significantly larger than that of the residual vector calculation. This superiority of quasi-Newton methods is thus increased with the number of unknowns in a problem.

Conversely, it is observed that strong nonlinearities lead to large number of quasi-Newton updates which in turn can lead to an ill-conditioned iteration matrix. It is thus advised to restart periodically the iteration procedure either using the initial Jacobian or by calculating the effective one in the actual stage of the response.

As a corollary, the so-called vectorial correction is well adapted since it allows for an easy restart of the updating procedure from the initial Jacobian.

The line search does not introduce a significant improvement in the convergence of quasi-Newton methods for the problems at hand, and should be performed only in exceptional cases.

Rank-two corrections do not yield an important improvement of the convergence rates. Hence, Davidon and Broyden rank-one corrections should be preferred due to their lower cost. In problems where Jacobians are definitely unsymmetrical, Broyden's formula should be preferred despite the need for a double linear system solution.

Future research and numerical tests should be devoted to optimal coupling between Newton and quasi-Newton strategies to reach always the minimal cost. Safeguarding methods described for these methods in the context of unconstrained minimization [12] should be explored to ascertain stability and convergence properties even in cases when the solution does no longer correspond to a minimum of a functional.

REFERENCES

1. J. E. Dennis and Jorge J. More, Quasi-Newton methods, motivation and theory. *SIAM Rev.* **19**(1), 46–89 (1977).
2. H. Mathies and G. Strang, The solution of nonlinear finite element applications. *Int. J. Num. Meth. Engng* **14**, 1613–1626 (1979).
3. K. J. Bathe and A. Cimento, Some practical procedure for the solution of nonlinear finite element applications. *Comp. Meth. Appl. Mech. Engng* **22**, 59–85 (1980).
4. K. J. Bathe and V. Sonnad, On effective implicit time integration in analysis of fluid-structure problems. *Int. J. Num. Meth. Engng* **15**, 943–948 (1980).
5. M. A. Crisfield, Iterative solution procedures for linear and nonlinear structural analysis Transport and Road Research Laboratory—TRRL Lab. Rep. 900, ISSN 0305-1293. Crowthorne, Berkshire (1979).
6. M. A. Crisfield, A faster modified Newton–Raphson

iteration. *Comp. Meth. Appl. Mech. Engng* **20**, 267–278 (1978).
7. M. Geradin, S. Idelsohn and M. Hogge, Nonlinear structural dynamics via Newton and quasi-Newton methods. *Nucl. Engng Des.* **54**, 339–348 (1980).
8. P. Beckers and S. Idelsohn, A comforming finite element for the analysis of viscous incompressible fluid flow. 3rd Int. Conf. Finite Elements in Water Resources, Univ. Mississippi, 19–23 May 1980.
9. L. K. Schubert, Modification of a quasi-Newton method for nonlinear equations with a sparse Jacobian. *Math. Comp.* **24**, 27–30 (1970).
10. M. P. Kamat and N. F. Knight, Nonlinear transient analysis via energy minimization. *AIAA J.*, **17**(9), 968–969 (1979).
11. S. Nagarajan and P. Popov, *Comput. Structures* **4**, 1117–1134 (1974).
12. M. A. Wolfe, *Numerical Methods for Uncontrained Optimization*. Van Nostrand Reinhold, Wokingham (1978).
13. G. Strang, Private communication (January 1980).
14. M. Geradin *et al.*, Module d'analyse dynamique non-linéaire NLDYN. L.T.A.S. Report VF-40, Aerospace Lab., Universite de Liège (1979).