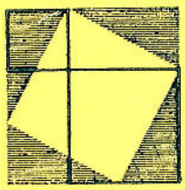
ITI-16

INFORME TECNICO INTERNO

Nº. 16

INSTITUTO DE MATEMATICA DE BAHIA BLANCA

INMABB (UNS - CONICET)



UNIVERSIDAD NACIONAL DEL SUR

Avda. ALEM 1253 - 8000 BAHIA BLANCA

República Argentina

INFORME TECNICO INTERNO N°16.

SCALE INVARIANT QUASI-NEWTON METHODS FOR THE SOLUTION OF NONLINEAR EQUATIONS

J.R.PALOSCHI* and J.D.PERKINS**

Dept. of Chemical Engineering and Chemical Technoloy

IMPERIAL COLLEGE - London - SW7 2AZ

* Presently at: Depto.de Matematica - Universidad del Sur
8000 - Bahia Blanca - Argentina

** Presently at: University of Sydney - Chem.Eng.Dept
Sydney - Australia

Accepted in Computers & Chem.Engineering (Pergamon Press)

1988

SCOPE:

The numerical solution of systems of nonlinear equations is the final stage in the computer simulation of steady state processes.

Many methods are available, based on Newton's method, obtained by attempting to emulate different characteristics of Jacobian approximations. None of the available methods achieve a very important property of Newton's method,its invariance to function or variable scaling.

In this paper, we propose scale invariant methods (under scaling with diagonal matrices for the variables).The new methods are very similar to the well known Broyden's method [1965] and exhibit similar superlinear convergence properties. These are shown by extending the convergence results of Broyden et al [1973].

SUMMARY

Sufficient conditions for a rank-one Quasi-Newton method for solving systems of algebraic nonlinear equations being scale invariant are derived. Methods which are invariant under changes in the scale of the equations with a general matrix and with diagonal matrices for the variables are proposed. The convergence results of Broyden,Dennis and More [1973] are extended in order to show local superlinear convergence of the new scale invariant methods. Extensive numerical tests of the proposed methods are performed showing the superiority of the new methods especially on "badly scaled" problems.

1. INTRODUCTION

We are concerned with the problem of finding a vector x_* (which will be assumed to exist) such that for a given function $f:R^n \rightarrow R^n$, x_* is a solution of

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

Almost all methods available for solving this problem are usually iterative. They generate a sequence of vectors $\{x_k\}$ in R^n which provided the method succeeds, will converge to x_*.

If $F'(x)$ is the Jacobian of the function f evaluated at x then for a linear function f we can see that for some approximation x_k to the solution, if $F'(x_k)$ is nonsingular, defining

$$p^N = -F'(x_k)^{-1} f(x_k) \tag{2}$$

allows us to find x_* from x_k as

$$x_* = x_k + p^N \tag{3}$$

If f is not a linear function but x_k is sufficiently near x_* we can expect that defining

$$x_{k+1} = x_k + p^N \tag{4}$$

will provide a better approximation to x_*, i.e. $\|x_{k+1}-x_*\| < \|x_k-x_*\|$. This is the basis for Newton's method and thus p^N is called the Newton step. A very important feature of Newton's method is its invariance to function or variable scaling. Given nonsingular matrices S_x and S_f define

$$\underline{x} = S_x\, x \tag{5}$$

$$\underline{f}(x) = S_f\, f(x). \tag{6}$$

If $\{x_k\}$ is the sequence of vectors generated by Newton's method for the unscaled problem then if $\{\underline{x}_k\}$ denotes the sequence generated for the problem scaled using (5) and (6) we can see that

$$\underline{x}_k = S_x\, x_k \text{ , for all k.} \tag{7}$$

Since the evaluation of the jacobian will often be expensive Newton like methods have been developed which do not need to evaluate $F'(x)$. If B_k is an approximation to $F'(x_k)$ then if we define

$$p_k = -B_k^{-1} f(x_k) \tag{8}$$

we will have an approximation to p^N and thus we can predict a new approximation x_{k+1} as

$$x_{k+1} = x_k + p_k. \tag{9}$$

We will denote $f(x_k)$ by f_k and then define

$$y_k = f_{k+1} - f_k \tag{10}$$

and if we assume as before that f is a linear function then we

can easily verify that

$$F'(x_{k+1})\ p_k = y_k$$

and it will be natural to ask for each approximation B_{k+1} to the Jacobian to satisfy

$$B_{k+1}\ p_k = y_k. \qquad (11)$$

This relation between B_{k+1} , p_k and y_k is called the secant relation and is the basis for a very important class of methods for solving systems of nonlinear equations called Quasi-Newton methods. Each method in this class generates a sequence of vectors $\{x_k\}$ and a sequence of matrices $\{B_k\}$ satisfying (8),(9) and (11). This class of methods will not in general generate a sequence of vectors $\{x_k\}$ invariant to changes of scale of the form (5) and (6) i.e., (7) will not in general be satisfied.

We will obtain in section 2 sufficient conditions for a Quasi-Newton method being invariant under function or variable scaling. We will propose in particular a method satisfying the conditions for scale invariance for any matrix S_f and diagonal matrices S_x.

An extension to the convergence theory of Broyden,Dennis and More [1973] will be proposed in section 3. This will in particular allow us to show superlinear convergence for the proposed scale invariant methods.

In section 4 we will present a description of our implementation of Quasi-Newton methods and finally in section 5 we will present some numerical results.

2. SCALE INVARIANT QUASI-NEWTON METHODS

The relation between the Jacobian of f evaluated in the new scale (considering changes of scale of the form (5) and (6)) and the Jacobian in the original scale is given by

$$\bar{F}'(\bar{x}) = S_f\ F'(x)\ S_x^{-1} \qquad (12)$$

The following lemma will show that this relation is the basis for a sufficient condition for a Quasi-Newton method being scale invariant.

Lemma 2.1:

Given the changes of scale (5) and (6) and sequence of nonsingular matrices $\{B_k\}$ satisfying the following relation

$$\bar{B}_k = S_f\ B_k\ S_x^{-1} \qquad (13)$$

then the Quasi Newton method defined by (8),(9) and (11) satisfies (7) i.e. ,is scale invariant.

Proof:

Since x_0 satisfies (7) it will be sufficient to show that if x_k satisfies (7) then it also does x_{k+1}.From (8) and (13) we can show that $\bar{p}_k = S_x\ p_k$ and then from (9) follows that x_{k+1} satisfies (7).

A very important class of Quasi-Newton methods is obtained with the following rank-one update formula

$$B_{k+1} = B_k + a_k\ b_k^T. \qquad (14)$$

Since (11) must be satisfied it follows

$$a_k = (y_k - B_k p_k)$$

and $b_k p_k = 1$.

We then have a family of methods which can be characterised with a sequence $\{v_k\}$ in R^n using the update formula

$$B_{k+1} = B_k + \frac{(y_k - B_k p_k) v_k^T}{v_k^T p_k} . \qquad (15)$$

By using lemma 2.1 we can now establish sufficient conditions for a rank-one Quasi-Newton method of the form (15) being scale invariant.

Lemma 2.2:

Consider a rank-one Quasi-Newton method generating sequences $\{x_k\}$ and $\{B_k\}$ such that (8),(9) and (15) are satisfied and that for any change of scale of the form (5) and (6) B_0 is such that

$$\underline{B}_0 = S_1 B_0 S_1^{-1} .$$

Then if $\{v_k\}$ is such that

$$\underline{v}_k^T = v_k^T S_k^{-1} , \text{ for all } k \qquad (16)$$

$\{x_k\}$ satisfies (7), i.e. the method is scale invariant.

Proof:

Since

$$\underline{B}_{k+1} = \underline{B}_k + \frac{(\underline{y}_k - \underline{B}_k \underline{p}_k) \underline{v}_k^T}{\underline{v}_k^T \underline{p}_k}$$

we can prove using induction that $\{B_k\}$ satisfies (13). From the hypothesis B_0 satisfies (13) hence we should show that provided B_k satisfies (13) then it also does B_{k+1}. This follows by using (16) and the hypothesis. Now by using lemma 2.1 we can show that the method is scale invariant.

We will now show the existence of methods satisfying lemma 2.2 when we restrict the scaling matrix S_k to be diagonal.The following notation will be used for a general vector v

$[v]_i$ denotes the i-th component of v

and for a general matrix A

$[A]_{i,j}$ denotes the element in row i, column j.

Definition: Define the pseudoinverse a^+ of a real number a as

$$a^+ = \begin{cases} 0 & \text{if } a=0 \\ 1/a & \text{if } a\neq 0 \end{cases}$$

Given the sequence $\{x_k\}$ define the sequence $\{v_k\}$ such that for each i, $1 <= i <= n$

$$[v_k]_i = [p_k]_i ([x_k]_i)^2 \qquad (17)$$

Lemma 2.3:

The sequence $\{v_k\}$ defined by (17) satisfies (16).

Proof:

$$[\underline{y}_k]_i=[p_k]_i\ ([\underline{x}_k]_i)^2=[S_k\ p_k]_i\ [(S_k\ x_k]_i)^2.$$

Since S_k is restricted to be diagonal then

$$[\underline{y}_k]_i = [S_k]_{ii}\ [p_k]_i\ ([S_k]_{ii}\ [x_k]_i)^2$$

and since S_k is nonsingular it follows

$$[\underline{y}_k]_i = [v_k]_i/[S_k]_{ii}$$

hence (16) is satisfied .

3. CONVERGENCE RESULTS

In this section we will extend a convergence theorem of Broyden,Dennis and More [1973] and apply the result in order to show local superlinear convergence for the scale invariant method proposed in the previous section .

Theorem 3.1: (Broyden,Dennis and More [1973],Theorem 4.3,pp.235)

Let $f:R^n \longrightarrow R^n$ be differentiable in the open,convex set D and assume that for some x_* in D, $f(x_*)=0$ and $F'(x_*)$ is nonsingular and for some $b>0$

$$\|F'(x)-F'(x_*)\| < b\ \|x-x_*\| \qquad (18)$$

for all x in D.

Assume there exist $u_1>0$ and a nonsingular symmetric matrix M such that

$$\|M\ v_k-M^{-1}\ p_k\|\le u_1\|M^{-1}\ p_k\|\max(\|x_{k+1}-x_*\|,\|x_k-x_*\|) \qquad (19)$$

for all x_k in D and B_k such that $\|B_k-F'(x_*)\|<d$ for some $d>0$, where p_k is defined by (8) and x_{k+1} by (9). Then the method defined by (8),(9) and (15) is well defined in a neighborhood of x_*, and the sequence $\{x_k\}$ is locally and superlinearly convergent to x_* .

If in (15) we replace v_k by p_k Broyden's [1965] method is obtained. This method is scale invariant for changes of the form (5) (that is, scaling the function), but if a change of scale of the form (6) is applied, a different iteration sequence will be obtained (unless S_k is orthogonal). The new update proposed in section 2 also retains the invariance with respect to transformations of the function which Broyden's method posseses.

We will consider choices of sequences $\{v_k\}$ such that there exist a sequence of matrices $\{V_k\}$ and

$$\{v_k\} = \{V_k\ p_k\}. \qquad (20)$$

With $\{v_k\}$ defined in this way we can establish an equivalent condition like (16) for $\{V_k\}$ as

$$\underline{V}_k = (S_k^{-1})^T\ V_k\ S_k^{-1} . \qquad (21)$$

We will now extend theorem 3.1 for the case when (v_k) is of the form (20) by proposing a relation similar to (19).

Theorem 3.2:

Let $f:R^n \rightarrow R^n$ satisfy the conditions of theorem 3.1 and let (V_k) be a sequence of matrices such that for a fixed symmetric positive definite matrix V_* the following relation is satisfied for all k

$$||V_k - V_*|| \leq u \max\{||x_{k+1}-x_*||, ||x_k - x_*||\} \qquad (22)$$

for some real positive u. Then the method defined by (8),(9),(15),(20) and (21) is locally and superlinearly convergent.

Proof:
Since the method is scale invariant it will be sufficient to show convergence for a particular scale. Define the change of scale

$$\underline{x}_k = V_*^{1/2} x_k .$$

We can see that

$$||\underline{y}_k - \underline{p}_k|| = ||V_*^{-1/2} v_k - V_*^{1/2} p_k|| = ||V_*^{-1/2} (V_k - V_*) p_k||$$

hence

$$||V_*^{-1/2} v_k - V_*^{1/2} p_k|| \leq (||V_*^{-1/2}||)^2 ||V_k - V_*|| \; ||V_*^{1/2} p_k|| .$$

Now defining $u_1 = u \, ||V_*^{-1/2}||^2$

it follows

$$||V_*^{-1/2} v_k - V_*^{1/2} p_k|| \leq$$

$$u_1 \max\{||x_{k+1}-x_*||, ||x_k - x_*||\} \; ||V_*^{1/2} p_k|| .$$

We now apply theorem 3.1 identifying M as $V_*^{-1/2}$.

We will now show that (v_k) as defined by (17) satisfies the conditions of theorem 3.2, i.e. generates a scale invariant rank-one Quasi-Newton method with local superlinear convergence. The solution vector x_* will be required to have no zero components.

Lemma 3.3:

If x_* is such that it has no zero components then the sequence (v_k) defined by (17) satisfies condition (22).

Proof:
The matrix V_k in this case is diagonal, its elements are

$$[V_k]_{ii} = ([x_k]_i)^2$$

Define V_* as the diagonal matrix defined by

$$[V_*]_{ii} = ([x_*]_i)^2 .$$

V_* is symmetric and positive definite since x_* has no zero components. Since x_k must be in an open set containing x_*, we can assume it to be sufficiently small such that x_k

has also all its components away from the origin. Then we can put

$$[V_k]_{ii}-[V_*]_{ii} = ([x_*]_i^2-[x_k]_i^2)([x_*]_i^{\dagger}[x_k]_i^{\dagger})^2$$

$$= ([x_*]_i-[x_k]_i)([x_*]_i+[x_k]_i)([x_*]_i^{\dagger}[x_k]_i^{\dagger})^2$$

and we can show there exists a constant g_i such that

$$([x_*]_i+[x_k]_i)^2([x_*]_i^{\dagger}[x_k]_i^{\dagger})^4 < g_i^2$$

hence by using the Frobenius norm for matrices and since V_* and V_k are diagonal

$$||V_k-V_*||^2 = \sum_{i=1}^{n} ([V_k]_{ii}-[V_*]_{ii})^2$$

$$<= \sum_{i=1}^{n} ([x_*]_i-[x_k]_i)^2 g_i^2).$$

If we define $g^2 = \max_{1<=i<=n} (g_i^2)$ then it follows

$$||V_k-V_*|| <= g ||x_k-x_*||$$

and (22) is immediate.

4. IMPLEMENTATION OF QUASI NEWTON METHODS

We will present in this section our implementation of general Quasi-Newton methods generated with the updating formula (15). In particular, Broyden's [1965] "good" method will be a particular case ($v_k=p_k$) as will the scale invariant method proposed in section 2.

4.1 The initial approximation B_o.

B_o is obtained using a finite difference approximation to the Jacobian of f evaluated at x_o. For this purpose a perturbation step d is used defined as

$$[d]_i = \begin{cases} 1.E-8 & \text{if } |[x_o]_i| <= 1.E-6 \\ 0.01\,[x_o]_i & \text{if } |[x_o]_i| > 1.E-6 \end{cases}$$

B_o obtained in this way will satisfy (13), for k=0, provided

$$|[x_o]_i| >= 1.E-6 \ , \ 1 <= i <= n \ .$$

4.2 Obtaining the Quasi-Newton step

We must solve the linear system $B_k\, p_k = -f_k$ at each iteration. This have been solved in the past by updating the inverse of B_k using the formula of Sherman and Morrison [1949] (Broyden [1965]). More recently, a QR factorization of B_k has been used (More and others [1980]), which is claimed to be more stable but with the disadvantage of using $n^2/2$ additional

storage. We have used for our implementation an LU factorization of B_k (also used by Chen and Stadtherr [1981]). This does not require additional storage and our numerical experience suggests that it is reliable.

The initial matrix B_0 is factorized into L_0 and U_0, L_0 being lower triangular and U_0 unit upper triangular. If we find that B_0 is singular (or numerically singular), which results in a small diagonal element in L_0, we simply replace the small element with one which does not make B_0 singular. On each iteration L_k and U_k must be updated according to (15) and for this we use a modified version of Bennett's [1965] algorithm which can be found in Paloschi and Perkins [1986]. The modification has been specially designed to avoid L_{k+1} being singular.

More and Cosnard [1979] haver proposed to control the norm of the step at each iteration to satisfy the following relation:

$$||p_k|| <= d_k \qquad (23)$$

where

$$d_0 = \max \{10, 10\ ||x_0||\}$$

$$d_k = \max \{d_{k-1}, 10\ ||x_k||\} \ , \ k>0 \ .$$

To achieve this, instead of (8) the following is used

$$p_k = \text{lamda}\ B_k^{-1}\ f(x_k) \qquad (24)$$

where $0 < \text{lamda} <= 1$ is chosen to satisfy (23) .

This simple control, as stated, is scale dependent and we have modified it in order to achieve scale invariance. We have used

$$|[p_k]_i| <= [d_k]_i \ , \ 1<=i<=n \qquad (25)$$

where

$$[d_k]_i = \begin{cases} 50 & \text{if } [x_k]_i = 0 \\ 50\ |[x_k]_i| & \text{if } [x_k]_i \neq 0 \end{cases}$$

The constant lamda is then chosen in (24) satisfying (25).

4.3 Reinitialising the approximation to the Jacobian

We have obtained a considerable improvement in the performance of the code by reinitialising B_k using finite differences when good progress is not being made. We consider good progress is not being made if after 10+n consecutive iterations the norm of the function has not been reduced at least by a factor of 0.9 last reinitialisation the norm of the function has been reduced at least once by a factor 0.9 . The point used for the reinitialisation is the one with the minimum norm found during the iteration process.

4.4 Other implementation details

A problem is considered to be solved when

$$|[f(x_k)]_i| <= 1.E-7 \ , \ 1<=i<=n \ .$$

A simple control on the norm of $f(x_k)$ is used such that

$$||f(x_k)|| <= 100\ ||f(x_o)||$$

The constant lamda in (24) is used to reduce the step if necessary to achieve it. This control will make the code fail to converge to a solution of (1) if an initial point close to a nonzero minimum of $||f(x)||$ is specified. The use of this stepping rule causes loss of scale invariance

5. NUMERICAL RESULTS

Hiebert [1982] did a comprehensive comparison of different codes available for solving systems of nonlinear equations. We will use the same set of examples which was proposed by More,Garbow and Hillstrom [1978] and was also used by Chen and Stadtherr [1981]. In the Hiebert report an implementation due to More and others [1983] of the hybrid Powell's method was found to be the best on the general set of problems. We will use the same code for our comparison.

5.1 Methods to be compared

In addition to the scale invariant method proposed in section 2 many more methods can be found. The following definitions will also give sequences $\{v_k\}$ satisfying the scale invariance property (16).

$$[v_k]_i = [x_{k+1}]_i \qquad (26)$$

$$[v_k]_i = [p_k]_i\ ([p_o]_i)^2 \qquad (27)$$

$$[v_k]_i = [p_k]_i\ ([x_k-x_o]_i)^2 \qquad (28)$$

The methods given by (27) and (28) are also scale invariant under affine changes of the more general form

$$\bar{x} = S_k\ (x-b)$$

for a constant vector b.

The method given by (27) will be local superlinearly convergent provided the vector p_o has all its components away from the origin while the one given by (28) will require (x_*-x_o) having all its components away from the origin in order to have superlinear local convergence.

We will compare these four methods with the classical Broyden's "good" method ($v_k=p_k$ in (15)) and the hybrid code of More and others [1980]. The methods will be numbered:

Method 1	v_k by (26)
Method 2	v_k by (17)
Method 3	v_k by (27)
Method 4	v_k by (28)
Method 5	$v_k=p_k$ (Broyden's "good" method)
Method 6	hybrid method

Methods 1 to 5 are implemented using the same code (only differing in the way v_k is obtained) in order to minimize the influence of external factors (such as precision in intermediate calculations) on the bahaviour of the methods.

All computations were performed on a CDC CYBER 174 in single precision.

5.2 The problems

5.2.1 The basic set of problems

The basic set of examples consist of 18 different problems, each one having a standard initial point.
The list of problems is:

A. Rosenbrock's function, n=2
B. Powell's singular function, n=4
C. Powell's badly scaled function, n=2
D. Wood's function, n=4
E. Hellical valley function, n=3
F. Watson's function
G. Chebyquad function
H. Brown's almost linear function
I. Discrete boundary value problem
J. Discrete integral equation function
K. Trigonometric function
L. Variable dimensioned function
M. Broyden's tridiagonal function
N. Broyden's banded function

They are collected in the MINPACK test routines VECFCN and INITP, problems F to N are of variable dimension.

5.2.2 The selected sets of problems

Two sets of problems have been constructed. The first one, which will be called the "general set", consists of 54 problems taken from the basic set, 21 having as initial point x_o (the standard one), 18 with $20x_o$ and 15 with $100x_o$. In Table 1 we summarize the problems in this set. This set of problems was used in the Hiebert report and also in the work of Chen and Stadtherr.

The second set will be taken as a subset of the general set and thus will be called the "general subset". It consists of the 16 problems listed in Table 2 all having the standard initial point. All the problems of this last set were selected such that all methods considered converge from the standard initial point.

5.3 Results

For the purpose of testing the behaviour of methods under different scaling conditions a diagonal matrix $S_{m,n}$ is defined as:

$$\log_{10}[S_{m,n}]_{ii} = m((2i-n-1)/(n-1)),\ 1<=i<=n\ . \qquad (29)$$

To compare the methods regarding its efficiency a number c_j is defined for each method on each problem as:

$$c_j = \begin{cases} 0 & \text{if the method failed to converge} \\ n_o/n_j & \text{if the method converged} \end{cases}$$

where j indicates the method, n_j is the number of function evaluations used for method j and n_o is the number of function evaluations used for the most efficient method of all on this particular problem.

The first results are presented in Table 3. They have been obtained using the general set in the same way as in the Hiebert's report. The problems are used in their original form plus the following two cases: for testing a set with variables badly scaled we used (as did Hiebert)

$$f(\underline{x}) = f(S_{s,n}\, x)$$

and for function badly scaled

$$\underline{f}(x) = S_{s,n}\, f(x)$$

where $S_{s,n}$ is defined by (29). In all we have then 162 problems.

The second set of results is presented in Table 4. For this we have used the general subset and to test the behaviour under a gradual deterioration in the scaling we have used

$$f(\underline{x}) = f(S_{m,n}\, x)$$

for m=0,4,8,12,and 16, which gives then a set of 80 problems.

The storage requirements for the methods is:

methods 1-5 $\quad n^2 + 8n$

method 6 $\quad n^2 + n^2/2 + 10n$

For the statistics related to the efficiency measure c_j, only those cases where convergence has been achieved are considered (i.e., failing cases do not degrade efficiency). By this means we hope to study efficiency and robustness independtly of each other.

From the results in Table 3 we can see that the hybrid code is clearly superior on the unscaled set of problems but on badly scaled sets its performance is poor whereas bad scaling does not cause the performance of the other methods to deteriorate. Thus overall the hybrid method gives worse results than the other methods. This finding coincides with the results reported by Chen and Stadtherr [1981].

While it can be seen in Table 3 that there is not too much difference in the behaviour of the scale invariant methods and Broyden's "good" method on the general set, we can see in Table 4 the effect of scale invariance on problems which are very badly scaled. Since the amount of work per iteration is the same (as is the storage needed for the implementation) for all the Quasi Newton methods it is clearly an advantage to use method 3 with its scale invariance and superlinear convergence properties instead of the classic method of Broyden.

6. CONCLUSIONS

We have proposed scale invariant Quasi-Newton methods which are competitive with the best codes available. Numerical results show the advantage of using them as an alternative to the classic Broyden's "good" method.

The theory of convergence of Broyden,Dennis and More [1973] has been extended such that, in particular, it can be used to show superlinear convergence for the methods proposed here.

REFERENCES

BENNETT J.M. - "Triangular factors of modified matrices"
Numerische Mathematik - 7 - [1965],217-224

BROYDEN C.G. - "A class of methods for solving nonlinear simultaneous equations"
Math.of Comp. - 19 - [1965],577-293

BROYDEN C.G.,DENNIS J.E. and MORE J.J. - "On the local and superlinear convergence of Quasi-Newton methods"
Journal.of the IMA - 12 - [1973],223-245

CHEN H.A.,STADTHERR M.A. - "A modification of Powell's dogleg method for solving systems of nonlinear equations"
Comp.& Chem.Eng. - 5 -[1981],143-150

HIEBERT K.L. - "An evaluation of mathematical software that solves systems of nonlinear equations"
ACM TOMS - 8 - [1982]

MORE J.J.,COSNARD M.Y. - "Numerical solution of nonlinear equations"
ACM TOMS - 5 - [1979],64-85

MORE J.J.,GARBOW B.S. and HILLSTROM K.E. - "Testing unconstrained optimization software"
Argonne Nat.Lab - Report ANL-AMD-TM-324 [1978]

MORE J.J.,GARBOW B.S. and HILLSTROM K.E. - "User's guide for MINPACK 1"
Argonne Nat.Lab - Report ANL-80-74 [1980]

PALOSCHI J.R.,PERKINS J.D. - The updating of LU factors in Quasi-Newton methods"
Comp. & Chem.Eng. - 10 - [1986],241

SHERMAN J.,MORRISON W.J. - "Adjustment of an inverse matrix corresponding to changes in elements of a given column or a given row of the original matrix"
Ann.Math.Statist. - 20 - [1949],624.

Prob	Orig prob	Dim	Prob	Orig prob	Dim	Prob	Orig prob	Dim
1	A	2	19	L	10	37	L	10
2	B	4	20	M	10	38	M	10
3	C	2	21	N	10	39	N	10
4	D	4	22	A	2	40	A	2
5	E	3	23	B	4	41	B	4
6	F	6	24	C	2	42	D	4
7	F	9	25	D	4	43	E	3
8	G	5	26	E	3	44	G	5
9	G	6	27	F	6	45	G	6
10	G	7	28	F	9	46	G	7
11	G	9	29	G	5	47	H	10
12	H	10	30	G	6	48	I	10
13	H	30	31	G	7	49	J	2
14	H	40	32	H	10	50	J	10
15	I	10	33	I	10	51	K	10
16	J	2	34	J	2	52	L	10
17	J	10	35	J	10	53	M	10
18	K	10	36	K	10	54	N	10

Table 1: List of problems and dimensions for the general set

Prob	Orig.prob	Dim
1	A	2
2	B	4
3	C	2
4	F	6
5	F	9
6	G	5
7	G	6
8	G	7
9	H	10
10	H	30
11	I	10
12	J	2
13	J	10
14	L	10
15	M	10
16	N	10

Table 2: List of problems and dimensions for the general subset

		method					
		1	2	3	4	5	6
fails	unscaled	14	12	13	14	12	5
	var.scaled	12	14	13	13	13	21
	func.scaled	18	18	14	14	15	22
	TOTAL	44	44	40	41	40	48
averages of c_j	unscaled	0.75	0.76	0.81	0.77	0.80	0.87
	var.scaled	0.74	0.78	0.76	0.76	0.75	0.88
	func.scaled	0.70	0.84	0.84	0.80	0.81	0.87
	TOTAL	0.73	0.79	0.80	0.78	0.79	0.88

Table 3: Summary of results for the general set of problems

		method				
	m	1	2	3	4	5
fails	0	0	0	0	0	0
	4	0	0	0	0	0
	8	3	0	0	1	3
	12	9	8	5	5	10
	16	14	14	11	12	14
	TOTAL	26	22	16	18	27
averages of c_j	0	0.82	0.84	0.91	0.84	0.89
	4	0.79	0.82	0.92	0.87	0.85
	8	0.76	0.74	0.84	0.88	0.85
	12	0.40	0.49	0.82	0.89	0.45
	16	0.44	0.77	0.48	0.90	0.97
	TOTAL	0.73	0.76	0.85	0.87	0.82

Table 4: Summary of results for the general subset

NOTATION USED

$\underline{x}$ denotes vector x in the new scale given by equation (5)

$\underline{f}$ denotes vector f in the new scale given by equation (6)

$\underline{F}'$ denotes the Jacobian F' in the new scale given by equations (5) and (6)

$[x]_i$ denotes the pseudoinverse of the i-th component of x

$\|x\|$ denotes a vector norm

$\|A\|$ denotes a matrix norm consistent with the previous vector norm