

## **MODIFICATIONS OF THE CONTINUATION METHOD FOR THE SOLUTION OF SYSTEMS OF NONLINEAR EQUATIONS**

**G.R. LINDFIELD and D.C. SIMPSON**

Computer Centre  
The University of Aston in Birmingham  
15 Coleshill Street  
Birmingham B4 7PA, England

(Received September 20, 1978)

ABSTRACT. Modifications are proposed to the Davidenko-Broyden algorithm for the solution of a system of nonlinear equations. The aim of the modifications is to reduce the overall number of function evaluations by limiting the number of function evaluations for any one subproblem. To do this alterations are made to the strategy used in determining the subproblems to be solved. The modifications are compared with other methods for a wide range of test problems, and are shown to significantly reduce the number of function evaluations for the difficult problems. For the easier problems the modified method is equivalent to the Davidenko-Broyden algorithm.

KEY WORDS AND PHRASES. *Davidenko-Broyden algorithm, Nonlinear equations, Numerical Solutions.*

AMS (MOS) SUBJECT CLASSIFICATION (1970) CODES. *65H10, 65D15.*

1. INTRODUCTION

This paper considers the solution of a system of  $n$  nonlinear equations with  $n$  unknowns  $\underline{x} = (x_1, x_2, \dots, x_n)^T$ , i.e. the system

$$\underline{f}(\underline{x}) = 0$$

An essential requirement of numerical techniques for solving such a system is the existence of an initial approximation to the solution. In general the closer the initial approximation is to the solution the easier this solution is to obtain.

Davidenko [3] approached the problem of providing a good approximation to the solution of the above system by transforming the original problem into a sequence of subproblems. Although the solution of each of the subproblems involves the solution of a nonlinear equation system the aim of the Davidenko method is to provide a sufficiently good approximation to each to ensure the overall difficulty of the problem is reduced.

Davidenko proposes a specific embedding or continuation which has the form

$$\underline{g}(\underline{x}, \theta) = \underline{f}(\underline{x}) - \theta \underline{f}_0 \quad \text{where } \underline{f}_0 = \underline{f}(\underline{x}_0)$$

$\theta$  is a scalar parameter and  $\underline{x}_0$  the initial approximation to the solution. The variation of the parameter  $\theta$  provides the subproblems. In particular when  $\theta = 0$

$$\underline{g}(\underline{x}, 0) = \underline{f}(\underline{x})$$

gives the original system and when  $\theta = 1$  then  $\underline{x}_0$  is a solution of

$$\underline{g}(\underline{x}, 1) = 0$$

this provides the initial subproblem. We can now specify a sequence of subproblems by taking values of  $\theta$

$$1 = \theta_0 > \theta_1 > \theta_2 > \dots > \theta_m = 0$$

then the subproblems will be

$$\underline{g}(\underline{x}, \theta_i) = 0 \quad ; \quad i = 1, 2, 3, \dots, m$$

The  $\theta_i$  should be chosen sufficiently close to one another so that the solution of one subproblem provides a good initial approximation for the next.

The major problem with this method is to make an effective choice of the  $\theta_i$  values which reduces the difficulty of the problem. Broyden [2] has studied the combination of his quasi Newton method [1] with the method of Davidenko, the quasi Newton method being used as the subproblem solver. In this paper Broyden suggested a specific method for choosing the  $\theta_i$  values. Broyden chose  $\theta_0 = 1$ ,  $\theta_1 = .99$  and  $\theta_2 = .98$  and fitted a quadratic approximation through the solution points of the latest three subproblems. Broyden obtains an estimate of the  $\theta$  value required for the next subproblem by extrapolating for  $\theta$  using the current quadratic approximation, the extent of the extrapolation being limited according to the local nonlinearity of the problem.

Broyden's [2] numerical studies showed that a significant improvement was obtained for harder problems by using this modified Davidenko technique with the quasi Newton method. However examination of the Davidenko-Broyden algorithm shows that for a number of problems a comparatively large proportion of the function evaluations arise from the solution of one or two subproblems indicating a poor choice of  $\theta$ . The next section proposes modifications to the Davidenko-Broyden algorithm with the aim of reducing the number of function evaluations.

## 2. MODIFICATION OF THE CONTINUATION METHOD

The aim of the modification is to prevent the number of function evaluations for solving any subproblem becoming excessive. The Davidenko-Broyden continuation method, as discussed in the previous section, solves the sequence of subproblems

$$\underline{g}(\underline{x}, \theta_i) = 0 \quad i = 1, \dots, m \quad (2.1)$$

If in solving equation (2.1) for some  $\theta_{r+1}$  the number of function

evaluations reaches a value MAXK without convergence to the solution, then it suggests that the step from  $\theta_r$  to  $\theta_{r+1}$  is too large and the solution of

$$\underline{g}(\underline{x}, \theta_{r+1}) = 0 \quad (2.2)$$

should be terminated. We must now choose a revised  $\theta$  for which convergence may be obtained within MAXK iterations without wasting the effort expended in part solving (2.2).

The strategy proposed here is to compute a new value,  $\theta^*$ , from  $\underline{x}^*$ , the solution vector at the point of terminating the solution of (2.2). The sub-problem  $\underline{g}(\underline{x}, \theta^*) = 0$  is then solved,  $\underline{x}^*$  being used as the initial approximation.

If  $\underline{x}^*$  is the solution vector of (2.2) after MAXK function evaluations, then as this is not the exact solution we can write

$$\underline{g}(\underline{x}^*, \theta) = \underline{\epsilon} \quad (2.3)$$

i.e. 
$$\underline{f}(\underline{x}^*) - \theta \underline{f}_0 = \underline{\epsilon}$$

We wish to calculate a  $\theta$  such that the error vector  $\underline{\epsilon}$  is as small as possible. That is, choose  $\theta$  so as to minimise

$$\begin{aligned} S &= \sum_{i=1}^n || \underline{\epsilon} ||^2 \\ &= \sum_{i=1}^n (f_i(\underline{x}^*) - \theta f_i(\underline{x}_0))^2 \end{aligned}$$

Differentiating

$$\begin{aligned} \frac{\partial S}{\partial \theta} &= -2 \sum_{i=1}^n f_i(\underline{x}_0) (f_i(\underline{x}^*) - \theta f_i(\underline{x}_0)) \\ &= -2(\underline{f}_0^T \cdot \underline{f}^* - \theta \underline{f}_0^T \cdot \underline{f}_0) \end{aligned}$$

The second derivative  $\partial^2 S / \partial \theta^2 = 2 \underline{f}_0^T \cdot \underline{f}_0 > 0$ , so that  $S$  is a minimum when  $\theta$  has the value  $\theta^* = (\underline{f}_0^T \cdot \underline{f}^*) / (\underline{f}_0^T \cdot \underline{f}_0)$ . Or, by writing  $\underline{f}^* = \underline{g}^* + \theta_{r+1} \underline{f}_0$ , where  $\underline{g}^* = \underline{g}(\underline{x}^*, \theta_{r+1})$ ,

$$\theta^* = \theta_{r+1} + (\underline{f}_0^T \cdot \underline{g}^*) / (\underline{f}_0^T \cdot \underline{f}_0) \quad (2.4)$$

The  $\theta^*$  computed from equation (2.4) may lie outside the range  $[\theta_{r+1}, \theta_r]$ . If this occurs then  $\theta^*$  is set equal to  $\theta_{r+1}$ , i.e. we continue with the solution of (2.2).  $\theta^*$  is also set equal to  $\theta_{r+1}$  if the following condition on the norm of  $\underline{f}(\underline{x}^*)$  is not satisfied.

$$\|\underline{f}(\underline{x}^*)\| < \|\underline{f}(\underline{x}_r)\| \quad (2.5)$$

If the condition (2.5) is satisfied then prior to the solving of the subproblem  $\underline{g}(\underline{x}, \theta^*) = 0$  a new initial approximation to the inverse Jacobian used by Broyden quasi Newton method is computed. In other words the computed value of  $\theta^*$  is only used if the solution vector  $\underline{x}^*$  is a better approximation to the solution of the nonlinear equations  $\underline{f}(\underline{x}) = 0$  than  $\underline{x}_r$ , the solution of the previous subproblem. If condition (2.5) is not satisfied or  $\theta^*$  is out of range then we continue with the solution of (2.2) with  $\underline{x}^*$  as the starting vector for another MAXK function evaluation.

Once a solution of  $\underline{g}(\underline{x}, \theta^*) = 0$  has been found we proceed in the normal way and compute the next  $\theta$  from  $\theta_{r-1}$ ,  $\theta_r$  and  $\theta^*$  and the solutions of the corresponding subproblems.

In solving the text problems in the next section the value for MAXK is chosen to be 25. This is an arbitrary choice and there may be some other optimal choice for this parameter, probably problem dependent. However Swift and Lindfield[6] performed experiments with different MAXK values for single nonlinear equations and found that within a range the value of MAXK was not critical.

### 3. RESULTS AND DISCUSSION

In this section the proposed modification is compared with the unmodified algorithm of Broyden[1] and the continuation method of Broyden[2].

Also considered is the algorithm of Powell[5]. This is a hybrid method using approximations to the Jacobian and inverse Jacobian to estimate a Newton step and

a step along the direction of steepest descent. The actual correction to the solution is a linear combination of these two steps. This algorithm is included as it is currently in the Mark 5 release of the Numerical Algorithms Group (NAG) Library.

The test equations are now presented. Test problems 1) - 3) are those of Broyden[2], problems 2) and 3) being particularly difficult. The algebraic equations 4) and 5) again prove difficult to solve; problem 6) is the test example appearing in the NAG specification of the Powell algorithm[5].

1) BROYDEN[2] EXAMPLE 3

$$f_i = \sum_{j=1}^6 \cot(x_j \beta_i) \quad j \neq i \quad i = 1, \dots, 6$$

$$10^2 \beta_i = 2.249, 2.166, 2.083, 2.000, 1.918, 1.835$$

$$\text{initial values } x_i = 75.0, i = 1, \dots, 6$$

2) BROYDEN[2] EXAMPLE 6

$$f_1 = 2(x_1 - 1) - 400x_1(x_2 - x_1^2)$$

$$f_2 = 200(x_2 - x_1^2)$$

$$\text{initial values } x_1 = -1.2, x_2 = 1.0$$

3) BROYDEN[2] EXAMPLE 7

$$\text{Equations as in 2), initial values } x_1 = -1.0,$$

$$x_2 = 1.0$$

4)  $f_1 = x_1 + x_2 + 2x_3 + 2$

$$f_2 = x_1x_2 + x_3^2 + x_4^2 + 2x_3(x_1 + x_2)$$

$$f_3 = 2x_1x_2x_3 + (x_1 + x_2)(x_3^2 + x_4^2) - 1$$

$$f_4 = x_1x_2(x_3^2 + x_4^2) + 2$$

initial values  $x_1 = 7.5$ ,  $x_2 = 11.0$ ,  $x_3 = 5.0$ ,  $x_4 = 7.0$

$$5) \quad f_1 = -13 + x_1 + ((5-x_2)x_2 - 2)x_2$$

$$f_2 = -29 + x_1 + ((1+x_2)x_2 - 14)x_2$$

initial values  $x_1 = 15.0$ ,  $x_2 = -2.0$

$$6) \quad f_1 = x_1(1 - x_2^2) - 2.25$$

$$f_2 = x_1(1 - x_2^3) - 2.625$$

$$7) \quad f_1 = -400x_1(x_2 - x_1^2) - 2(1 - x_1)$$

$$f_2 = 200(x_2 - x_1^2) + 20.2(x_2 - 1) + 19.8(x_4 - 1)$$

$$f_3 = -360x_3(x_4 - x_3^2) + 2(1 - x_3)$$

$$f_4 = 180(x_4 - x_3^2) + 20.2(x_4 - 1) + 19.8(x_2 - 1)$$

initial values  $x_1=0.5$ ,  $x_2=1.5$ ,  $x_3=1.5$ ,  $x_4=0.5$

8) Equations as in 7)

initial values  $x_1=-1.2$ ,  $x_2=1.0$ ,  $x_3=1.0$ ,  $x_4= 2.0$

The results of testing the various algorithms on these problems are presented in Table I. For each method the figures given are the total number of times the equations are evaluated in solving a problem to the specified accuracy. In all cases no more than 500 function evaluations were permitted. The convergence criterion used is that  $\|f(\underline{x})\| < 10^{-9}$ , or for the subproblems in the continuation algorithms  $\|g(\underline{x}, \theta)\| < 10^{-9}$ .

The abbreviations used for the different methods presented in Table I are that Broyden's unmodified algorithm given in section 2, with the parameter  $t = 1$ , is denoted by B. The Davidenko-Broyden method is denoted by DB. This method

also uses  $t = 1$  and the results correspond to the method of Broyden[2] where an initial approximation to the inverse Jacobian is computed before the start of each subproblem. The NAG routine of Powell is denoted by P and the proposed modified Davidenko-Broyden algorithm by MDB.

Comparison of the first two columns in Table I shows that the unmodified version of Broyden's algorithm and the NAG routine of Powell are very similar in their performance. Powell's algorithm being superior for problems 6), 7) and 8), but inferior for problem 1). However, out of the eight test problems the NAG routine fails to produce solutions in five of them. The failures marked F in column P are not due to exceeding the maximum number of function evaluations but due to some other failure in the NAG routine.

The Davidenko-Broyden algorithm produces solutions to all but one of the problems. It fails for problem 7) but is very satisfactory for problem 8), which is the same problem but with different starting values. The DB algorithm tends to take more function evaluations than the Broyden algorithm for the easier problems, but is much better for the difficult problems.

The modified Davidenko-Broyden algorithm gives converged results for all the test problems. For the difficult problems 2), 4), 7) and 8), where the DB algorithm takes a large number of function evaluations, the modified method significantly reduces the number of function evaluations. For the easier problems it performs the same as the DB algorithm.

An examination of the sequence of  $\theta$  values used by the Davidenko-Broyden algorithm in solving the four difficult problems, shows that only three subproblems are solved for  $\theta$  values of 0.99, 0.98 and 0.0. In all cases the majority of function evaluations arise from the subproblem with  $\theta = 0$ , as shown in Table II.



The modified method however, stops the solution of the subproblem with  $\theta = 0$  after MAXK function evaluations and calculates a  $\theta^*$  close to zero and solves the subproblems  $\underline{g}(\underline{x}, \theta^*) = 0$  as is shown in Table III for problems 4) and 8).

The value of 30 function evaluations appearing after the first attempt to solve the subproblem with  $\theta = 0$ , comprises 25 evaluations (MAXK) from the solving of the subproblem and  $(n+1)$  function evaluations from the computation of the new initial approximation to the inverse Jacobian before solving the subproblem.

For problem 4) the  $\theta^*$  computed after the termination of the solution of  $\underline{g}(\underline{x}, 0) = 0$  is out of range and so the solution of the subproblem with  $\theta = 0$  is continued after computing a new initial approximation to the inverse Jacobian. This is sufficient for convergence to be attained after another 13 function evaluations.

It can be seen therefore that the modifications to the Davidenko-Broyden algorithm have the desired effect of spreading out the function evaluations among the subproblems and in reducing the total number of function evaluations.

TABLE I. Number of function evaluations for solutions of the test problems.

problem	P	B	DB	MDB
1	F	20	27	27
2	>500	>500	232	220
3	>500	>500	32	32
4	F	>500	197	58
5	F	>500	73	73
6	30	72	32	32
7	251	>500	>500	296
8	90	158	90	64

TABLE II. Function evaluations for subproblems of Davidenko-Broyden algorithm.

problem	$\theta=0.99$	$\theta=0.98$	$\theta=0.0$	total
2	5	4	223	232
4	9	6	182	197
7	7	4	>500	>500
8	8	4	78	90

TABLE III. Sequence of subproblems for modified Davidenko-Broyden algorithm.

problem 8							
$\theta$	0.99	0.98	0.0	$0.4 \times 10^{-4}$	0.0	total	
f.e.	8	4	30	15	7	64	
problem 4							
$\theta$	0.99	0.98	0.0	0.0	total		
f.e.	9	6	30	13	58		

#### REFERENCES

1. Broyden, C. G. A class of methods for solving nonlinear equations. Math. Comp. 19 (1965), 577-593.
2. Broyden, C. G. A new method of solving nonlinear simultaneous equations. Computer J. 12 (1969), 94-99.
3. Davidenko, D. F. On a new method of numerical solution of systems of nonlinear equations. Dokl. Akad. Nauk SSSR. 88 (1953), 601-602.
4. Rabinowitz, P. Numerical Methods for Nonlinear Equations, Gordon and Breach, London, 1970.
5. Numerical Algorithms Group Library Manual Mark 5, (1977).
6. Swift, A. and Lindfield, G. R. Comparison of a continuation method with Brent's method for the numerical solution of a single nonlinear equation. Computer J. 21 (1978), 359-362.