

## 1 SUMMARY

To calculate the solution of a system of non-linear equations when an initial approximation is given, i.e. to find  $x_1, x_2, \dots, x_N$  such that

$$f_i(x_1, x_2, \dots, x_N) = 0 \quad i = 1, 2, \dots, N.$$

The user must supply an initial estimate of  $x_1 \dots x_N$  and a subroutine CALFUN to calculate each  $f_i$ .

**ATTRIBUTES** — **Version:** 1.0.0. **Types:** NS11A; NS11AD. **Calls:** MB01C/CD. **Original date:** June 1993. **Origin:** M. J. D. Powell, Harwell, modified by M. J. Hopper.

## 2 HOW TO USE THE PACKAGE

### 2.1 The argument list and calling sequence

*The single precision version*

```
CALL NS11A (N, X, F, A, STEP, STPMAX, ACC, MAXFUN, IPRINT, W, IW)
```

*The double precision version*

```
CALL NS11AD (N, X, F, A, STEP, STPMAX, ACC, MAXFUN, IPRINT, W, IW)
```

The calling program must set the following:  $N$ ,  $X$ ,  $STEP$ ,  $STPMAX$ ,  $ACC$ , and the following are arrays which must appear in the calling program's DIMENSION statement:  $X$ ,  $F$ ,  $A$ ,  $W$ ,  $IW$ .

$N$  is an INTEGER variable set to the number of equations as in Section 1.

$X$  is a REAL (DOUBLE PRECISION in the D version) array, of minimum length  $N$ , which on entry to NS11A/AD must be set to an approximation to the solution. On a normal return  $X(J)$  will contain the  $J$ th component of the solution  $x_j$ .

$STEP$  is a REAL (DOUBLE PRECISION in the D version) variable set by the user which enables the subroutine to calculate estimates of the partial derivatives  $\frac{\partial f_i}{\partial x_j}$  numerically. The estimate calculated is the difference approximation

$$\frac{\partial f_i}{\partial x_j} \approx [f_i(x_1, x_2, \dots, x_j + STEP, \dots, x_N) - f_i(x_1, x_2, \dots, x_j, \dots, x_N)] / STEP$$

The user must choose  $STEP$  so that this estimate is reasonably accurate for each  $f_i$ , therefore the variables  $x_1, x_2, \dots, x_N$  must not be very different in magnitude (see §5).

$STPMAX$  is a REAL (DOUBLE PRECISION in the D version) variable set by the user to a generous estimate of the 'distance' between the initial approximation and the required solution [see §5 for 'distance']. If NS11A/AD predicts steps that are much larger than  $STPMAX$  an error return is made and a diagnostic is printed. This condition is often due to the fact that there is no solution of the equations close to the user's initial estimate.

**NOTE**  $STPMAX \tau STEP$

$ACC$  is a REAL (DOUBLE PRECISION in the D version) variable set by the user to the accuracy required. A normal return is made when

$$\sum_{k=1}^N [f_k(x_1, x_2, \dots, x_N)]^2 \leq ACC$$

MAXFUN is an INTEGER variable set to the maximum number of calls of CALFUN allowed, acting as a ‘safety valve’. An error return is made if the solution has not been found to the required accuracy within the number of calls. [A moderate choice of MAXFUN is  $10*N$ , but use a larger value if this is feasible.]

IPRINT is an INTEGER variable which determines printing, other than diagnostics, as follows:

IPRINT = 0: No printing

IPRINT  $\geq$  0: Values of  $x_1, x_2, \dots, x_N, f_1, f_2, \dots, f_N$  are printed at every IPRINT calls of CALFUN

IPRINT  $\leq$  0: These same items are printed at every |IPRINT| iterations. The final solution values are always printed for IPRINT $\neq$ 0 so to get final printing only set IPRINT=MAXFUN.

See §7 for error diagnostics.

The following arguments are set by NS11A/AD

- F is a REAL (DOUBLE PRECISION in the D version) array, of minimum length N, which will be set to the values of the  $f_j$  and the returned values of the variables  $x_1, x_2, \dots, x_N$ .
- A is a REAL (DOUBLE PRECISION in the D version) 2-dimensional array, of dimensions N\*N, set by the subroutine to an estimate of the inverse of the Jacobian matrix, i.e. if J has elements  $\frac{\partial f_i}{\partial x_j}$   $i, j = 1, 2, \dots, N$  then the array A contains an estimate of J-1. This matrix is useful for estimating the precision of the solution.
- W is a REAL (DOUBLE PRECISION in the D version) array used as working space and must be of length at least  $N(2N+5)$ .
- IW is an INTEGER array used as working space and must be of length at least N.

### 3. USER SUPPLIED SUBROUTINE CALFUN

```
SUBROUTINE CALFUN(N, X, F)
```

```
  DIMENSION X(1), F(1)
```

This subroutine must calculate each  $f_j(x_1, x_2, \dots, x_N)$  for  $j=1, 2, \dots, N$  when  $x_i, i=1, 2, \dots, N$  are given in the array X. i.e. it must set

$$F(K) = f_j(X(1), X(2), \dots, X(N)) \quad j = 1, 2, \dots, N.$$

### 4. OTHER SUBROUTINES

Library subroutine MB01C is called.

### 5. CHOICE OF SCALING

The functions  $f_i$  should be scaled so that all are similar in magnitude. The same applies to the variables  $x_i$ . The reasons for this scaling is partly because the accuracy depends on  $\sum_k f_k^2$ , and partly because the ‘distance’ between two estimates of the solution,  $(Y_1, Y_2, \dots, Y_N)$  and  $(Z_1, Z_2, \dots, Z_N)$  say, is taken to be

$$\sqrt{\sum_{k=1}^N (Y_k - Z_k)^2}$$

If these recommendations are not followed the routine may waste computing time.

## 6. METHOD

The method uses the ideas of Newton-Raphson and Steepest Descent, coupled with Broyden's method for improving Jacobian matrices. See A.E.R.E. Report R-5947, M.J.D. Powell "A FORTRAN subroutine for solving systems of non-linear algebraic equations".

## 7. ERROR RETURNS

The routine prints one of the following diagnostics in cases of failure and the error parameter IERR in COMMON, see §8, is set to the values indicated.

'ERROR RETURN FROM NS11A BECAUSE CALLS OF CALFUN FAILED TO IMPROVE THE RESIDUALS' (IERR=1)

'ERROR RETURN FROM NS11A BECAUSE F(X) FAILED TO DECREASE USING A NEW JACOBIAN' (IERR=2)

'ERROR RETURN FROM NS11A BECAUSE THERE HAVE BEEN MAXF CALLS OF CALFUN' (IERR=3)

'ERROR RETURN FROM NS11A BECAUSE A NEARBY STATIONARY POINT OF F(X) IS PREDICTED' (IERR=4)

For a description of the conditions causing these diagnostics see the reference mentioned in §6.

## 8. LABELLED COMMON

The subroutine uses a labelled COMMON block:

*The single precision version*

```
COMMON/NS11B/LP,LPD,IERR
```

*The double precision version*

```
COMMON/NS11BD/LP,LPD,IERR
```

LP is an INTEGER and may be set to the Fortran stream number to be used to output the results controlled by IPRINT. To suppress printing set LP=0. The default value is 6.

LPD is an INTEGER and may be set to the Fortran stream number to be used to output diagnostic printing. It can be set to zero to suppress printing and is defaulted to 6.

IERR is an INTEGER used by NS11 as an error code. On a successful return it will be set to zero. Other values indicate an error and these are given in §7.