

1 SUMMARY

This subroutine allows the user **to assess which parameters are well determined after a nonlinear least-squares fit** of a model to data has been found. Before the subroutine is called, the data $\beta_i, i=1, \dots, m$ must have been fitted to the model $m(\mathbf{x}, \alpha_i)$ by finding the values \mathbf{x}^* of the n parameters \mathbf{x} for which the sum of the squares of the residuals $r_i = m(\mathbf{x}, \alpha_i) - \beta_i$ is smallest. The sensitivity of these parameters to perturbations in the model and data is then determined – this information is present in the Jacobian matrix of the residuals at \mathbf{x}^* . The Jacobian matrix is estimated using finite differences and the essential information extracted from the Jacobian by forming its singular value decomposition (SVD). Any singular values smaller than a user specified tolerance will be regarded as zero and correspond to certain parameters being poorly determined by the model and data. The dependence of perturbations to the well-determined parameters upon arbitrary perturbations to the remaining parameters is calculated and the variance/covariance matrix for the well determined parameters is obtained.

ATTRIBUTES — **Version:** 1.0.0. **Types:** SV02A, SV02AD. **Remark:** Supersedes SV01. **Calls:** EB10, _DOT. **Original date:** December 1987. **Origin:** N.I.M.Gould and F.Law, Harwell.

2 HOW TO USE THE PACKAGE

During execution of the subroutine, control is passed back to the user in order to obtain additional information. This “reverse communication” is controlled via the parameter KNTRL described below. An example of this mode of operation is given in §4.

2.1 The argument list and calling sequence

The single precision version

```
CALL SV02A(NPARM, M, PARMS, F, SUMSQS, DROP, NGOOD,  
*         KNTRL, IWK, LIWK, WK, LWK, INFORM)
```

The double precision version

```
CALL SV02AD(NPARM, M, PARMS, F, SUMSQS, DROP, NGOOD,  
*         KNTRL, IWK, LIWK, WK, LWK, INFORM)
```

NPARM is an INTEGER variable which must be set to n , the number of parameters fitted to the data by the model, and is not altered by the subroutine. **Restriction:** NPARM \geq 0.

M is an INTEGER variable which must be set to m , the number of data values used in fitting the model to the data, and is not altered by the subroutine. **Restriction:** M \geq NPARM.

PARMS is a REAL (DOUBLE PRECISION in the D version) array of length NPARM. It must be set on input to \mathbf{x}^* , the values of the parameters \mathbf{x} which give the best (least-squares) fit of the model to the data. On an intermediate exit (KNTRL=1), PARMS will have been perturbed in order to allow a finite-difference approximation to the Jacobian matrix to be formed. However, on a final exit (KNTRL=2), PARMS will be reset to its initial (KNTRL=0) value.

F is a REAL (DOUBLE PRECISION in the D version) array of length M. On input, the i -th element of F must be set to the i -th residual r_i evaluated at the point \mathbf{x}^* . In addition, if an exit is made from SV02 for which the parameter KNTRL has the value 1, the user must re-enter SV02 after evaluating the residuals at the point \mathbf{x} given in PARMS and placing the calculated values in F. On a final exit (KNTRL=2), F will contain the residuals at the point originally input in PARMS.

SUMSQS is a REAL (DOUBLE PRECISION in the D version) variable which will be set to the sum of squares of the

residuals evaluated at the point \mathbf{x}^* originally input in PARMs.

DROP is a REAL (DOUBLE PRECISION in the D version) variable which must be set to a tolerance below which singular values of the Jacobian matrix will be considered to be zero. As a rough rule, DROP should be set to the reciprocal of the precision required for the components of the parameters, \mathbf{x}^* . For instance, if each component of \mathbf{x}^* is required to be accurate to $\pm 10^2$, DROP should be set to 10^{-2} .

NGOOD is an INTEGER variable which gives the number of singular values which are larger than DROP, and consequently the number of parameters considered to be well determined by the data.

KNTRL is an INTEGER variable which is used to control the action of the subroutine and to request additional information from the user. On initial input KNTRL must have the value 0. If KNTRL has the value 1 on output, the user must re-enter SV02 after evaluating the residuals at the point given in PARMs and placing the calculated values in F. On a final exit KNTRL will have the value 2.

IWK is an INTEGER workspace array of length LIWK. On final output (KNTRL=2), the first NGOOD entries of IWK are the indices of the parameters which have been assessed to have been well determined by the model and data. The next NPARM-N entries contain the indices of the remaining parameters.

LIWK is an INTEGER variable which gives the actual length of the array IWK. LIWK must be at least $2 * \text{NPARM}$.

WK is a REAL (DOUBLE PRECISION in the D version) workspace array of length LIWK. On final output (KNTRL=2), WK contains the values of a number of the matrices and vectors pertinent to the analysis performed within SV02. See Section 2.2 for details.

LWK is an INTEGER variable which gives the actual length of the array WK. LWK must be at least $3 * \text{NPARM} + 2 * M + 2 * M * \text{NPARM} + 2 * \text{NPARM} * \text{NPARM}$.

INFORM is an INTEGER variable whose value on final exit (KNTRL=2) is used to indicate whether the subroutine has terminated successfully (INFORM=0) or not. If INFORM=-2, insufficient workspace has been provided and the user should check the parameters LIWK and LWK. Other nonzero values of INFORM should not occur, but indicate a failure in the SVD subroutine EB10A/AD. Please report if such a value of INFORM arises.

2.2 The array WK.

On final output (KNTRL=2), WK contains the values of a number of the matrices and vectors pertinent to the analysis performed within SV02 as summarized in Figure 1 and defined in section 4.

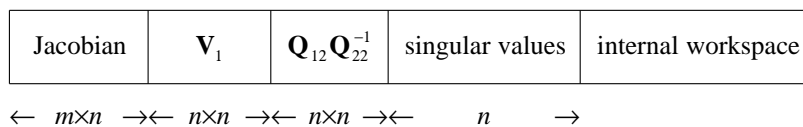


Figure 1. Partitioning of the workspace array WK

The estimated Jacobian matrix of the model at \mathbf{x}^* has its (i, j) -th entry in position $i + M * (j - 1)$ of WK and the (i, j) -th element of the NGOOD by NGOOD variance/covariance matrix of the well determined parameters occurs in position $\text{NPARM} * M + i + M * (j - 1)$ of WK. Next, the (i, j) -th entry of the $\text{NGOOD} * (\text{NPARMs} - \text{NGOOD})$ matrix $\mathbf{Q}_{12} \mathbf{Q}_{22}^{-1}$, which describes how perturbations of the well determined parameters depend upon arbitrary perturbations of the ill-determined parameters, occurs at position $\text{NPARM} * M + \text{NPARM} * \text{NPARM} + i + M * (j - 1)$ of WK. Finally, the i -th singular value of the Jacobian matrix is given in position $\text{NPARM} * M + 2 * \text{NPARM} * \text{NPARM} + i$ of WK. These terms are all described in section 4.

2.3 Common blocks

A single common block SV02B/BD is used by SV02A/AD.

The single precision version

```
COMMON/SV02B/ LP,MP,DSTEP
```

The double precision version

```
COMMON/SV02BD/ LP,MP,DSTEP
```

The following parameters may be altered by the user if required:

LP is an INTEGER variable which specifies whether output is required or not. If LP is less than or equal to 0, no output is given; otherwise output from the subroutine is written to Fortran output channel LP. A default value of 6 is supplied via the BLOCK DATA subprogram SV02C/CD.

MP is an INTEGER variable which specifies whether error messages are required or not. If MP is less than or equal to 0, no error messages will be printed; otherwise error messages from the subroutine will be written to Fortran output channel MP. A default value of 6 is supplied via the BLOCK DATA subprogram SV02C/CD.

DSTEP is a REAL (DOUBLE PRECISION in the D version) variable which is used to calculate the finite-difference approximations to the Jacobian matrix \mathbf{A} at \mathbf{x}^* . The approximation

$$A_{ij} = \frac{\partial r_i}{\partial x_j} = \frac{r_i(\mathbf{x}^* + \text{DSTEP } \mathbf{e}_j) - r_i(\mathbf{x}^*)}{\text{DSTEP}}$$

is used, where \mathbf{e}_j is the j -th column of the identity matrix. For well scaled residuals, a value of 10^{-3} (10^{-5} for the DOUBLE PRECISION version) should suffice and is supplied as a default via the BLOCK DATA subprogram SV02C/CD.

3 GENERAL INFORMATION

Use of common: See §2.4.

Other routines called directly: calls EB10 and _DOT.

Input/output: Output is under the control of argument LP.

Restrictions: $0 \leq n \leq m$.

4 METHOD

The method used is an implementation of that described in detail by A. R. Curtis in the paper "Analysis of covariance after nonlinear least-squares fitting", I. M. A. Journal of Numerical Analysis" **6** (1986), pp. 453-461. Further details and examples of the use of SV02 in the context of the Harwell code FACSIMILE/CHEKMAT are given by A. R. Curtis and W. P. Sweetenham in "FACSIMILE/CHEKMAT user's manual", AERE report R-12805 (1987), Harwell Laboratory, Oxfordshire, England.

The algorithm works in three stages. Firstly, the Jacobian matrix of the residuals is estimated by forming finite-differences around the best least-squares fit, \mathbf{x}^* of the model to the data. The user is required to supply values for the residuals at a number of small perturbations around \mathbf{x}^* . Secondly, a singular value decomposition of the Jacobian matrix is obtained. That is, the $m \times n$ ($m \geq n$) Jacobian matrix \mathbf{A} is factorized as

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{Q}^T,$$

where \mathbf{P} is $m \times n$ with orthonormal columns, \mathbf{D} is $n \times n$ and diagonal and \mathbf{Q} is $n \times n$ and orthogonal. The diagonal entries of \mathbf{D} are known as the singular values and small singular values indicate that certain of the parameter values \mathbf{x}^* are likely to be poorly determined. Finally, having decided which singular values are too small, and consequently how

many of the parameters are potentially badly determined, a decision as to which parameters are badly determined is made. If the parameters are reordered so that the well determined ones precede the badly determined ones (the order is given in the array `IWK` in the subroutine) and the matrices \mathbf{D} and \mathbf{Q} are partitioned as

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{pmatrix} \begin{array}{l} \text{- well determined parameters -} \\ \text{- badly determined parameters -} \end{array} \quad \text{and} \quad \mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix}$$

the variance/covariance matrix of the well-determined parameters can be expressed as

$$\mathbf{V}_1 = (\mathbf{Q}_{11} - \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1} \mathbf{Q}_{21})^T \mathbf{D}_1^{-2} (\mathbf{Q}_{11} - \mathbf{Q}_{12} \mathbf{Q}_{22}^{-1} \mathbf{Q}_{21})$$

and perturbations $\delta \mathbf{x}_2$ in the badly determined parameters will induce perturbations $\mathbf{Q}_{12} \mathbf{Q}_{22}^{-1} \delta \mathbf{x}_2$ in the well determined ones. Both of the matrices \mathbf{V}_1 and $\mathbf{Q}_{12} \mathbf{Q}_{22}^{-1}$ are calculated and, optionally, printed by `SV02`.

5 EXAMPLE OF USE

As a simple example, we suppose that we wish to fit the model

$$m(\mathbf{x}, \boldsymbol{\alpha}) = \alpha x_1 x_3 + x_2 e^{\alpha x_3}$$

to the data β for $(\alpha, \beta) = (0.2, 10.0), (0.4, 9.0), (0.6, 8.0), (0.8, 7.0), (1.0, 6.0)$ and that we find that the best least-squares fit is given by $\mathbf{x}^* = (-13.875814, 8.7827963, 0.39689345)$ (using, for instance, the Harwell Subroutine Library code `VA05`). The following piece of code can then be used to assess the parameter values obtained – we ask for the parameters to be correct to within 2.0.

```

      INTEGER          I, M, NPARM, INFORM, KNTRL, LIWK, LWK, NGOOD
      REAL            FLOAT
      DOUBLE PRECISION DROP, SUMSQS, R
      INTEGER          IWK( 6 )
      DOUBLE PRECISION PARMS( 3 ), F( 5 ), WK( 67 ),
*                   BETA( 5 ), DBLE, DEXP
      DATA           BETA / 1.0D+1, 9.0D+0, 8.0D+0, 7.0D+0, 6.0D+0 /
      M                = 5
      NPARM            = 3
      LIWK             = 6
      LWK              = 67
      KNTRL            = 0
      DROP             = 5.0D-1
      PARMS( 1 )      = - 1.3875814D+1
      PARMS( 2 )      =  8.7827963D+0
      PARMS( 3 )      =  3.9689345D-1
C
C   START OF THE MAIN LOOP.
C
100 CONTINUE
C
C   CALCULATE THE RESIDUALS.
C
      DO 110 I = 1, 5
         R = 2.0D-1 * DBLE( FLOAT( I ) )
         F( I ) = PARMS( 2 ) * DEXP ( R * PARMS( 3 ) )
*           + R * PARMS( 3 ) * PARMS( 1 ) - BETA( I )
110 CONTINUE
      CALL SV02AD( NPARM, M, PARMS, F, SUMSQS, DROP, NGOOD, KNTRL,
*               IWK, LIWK, WK, LWK, INFORM )
      IF ( KNTRL .EQ. 1 ) GO TO 100
      WRITE( 6, 2000 ) INFORM
      STOP
2000 FORMAT( /, ' EXIT FROM SV02, INFORM = ', I3 )

```

END

This produces the following output.

POST OPTIMAL ANALYSIS FOR NON-LINEAR LEAST SQUARES FITTING

THE METHOD AND NOTATION USED IS DESCRIBED IN :
ANALYSIS OF COVARIANCE AFTER NON-LINEAR LEAST SQUARES FITTING
A.R. CURTIS, IMA J. NUM. ANAL. 6 (1986) PP.453-461

THE DROP TOLERANCE IS 5.0000E-01

SINGULAR VALUES OF DEPENDENCE MATRIX
4.0566E+00 6.1618E-01 1.6709E-01

PERMUTE SEQUENCE OF PARAMETERS FROM

1	2	3
TO		
3	2	1

DATA DO NOT DETERMINE THE FOLLOWING 1 PARAMETERS,
SO DEFINE U(I) = PARAMETER(I) - VALUE(I) AS INDEPENDENT NUMBERS

I	PARAMETER NUMBER	VALUE(I)
3	1	-1.3876E+01

THE WELL DETERMINED PARAMETERS VARY AS Q12 * Q22(INVERSE) * U2
(SEE PAPER EQUATION (5.2))
THE MATRIX Q12 * Q22(INVERSE) IS

PARAMETER NUMBER	ENTRIES
3	-1.0628E-01
2	-2.9669E-01

COVARIANCE MATRIX OF WELL DETERMINED PARAMETERS (SEE PAPER, SECTION 6)

ROW 1	1.4520E+00	1.3910E+00
ROW 2	1.3910E+00	1.4537E+00

EXIT FROM SV02, INFORM = 0

If, instead, we ask for the parameters to be correct to within 10.0 by changing DROP to 1.0D-1 in the above program, we obtain the following output.

POST OPTIMAL ANALYSIS FOR NON-LINEAR LEAST SQUARES FITTING

THE METHOD AND NOTATION USED IS DESCRIBED IN :
ANALYSIS OF COVARIANCE AFTER NON-LINEAR LEAST SQUARES FITTING
A.R. CURTIS, IMA J. NUM. ANAL. 6 (1986) PP.453-461

THE DROP TOLERANCE IS 1.0000E-01

SINGULAR VALUES OF DEPENDENCE MATRIX
4.0566E+00 6.1618E-01 1.6709E-01

ALL PARAMETERS WELL DETERMINED BY DATA

COVARIANCE MATRIX OF WELL DETERMINED PARAMETERS (SEE PAPER, SECTION 6)

ROW 1	3.2774E+01	-9.1967E+00	-2.9675E+00
ROW 2	-9.1967E+00	4.0260E+00	2.2154E+00
ROW 3	-2.9675E+00	2.2154E+00	1.7125E+00

EXIT FROM SV02, INFORM = 0