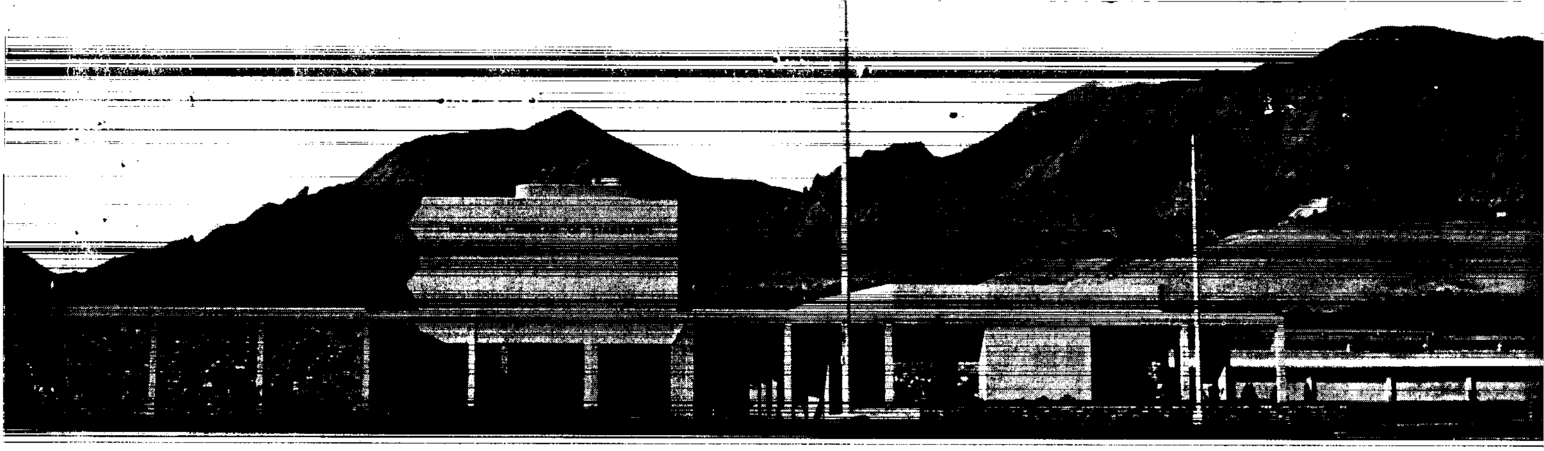

NBS REPORT

9258

FORTRAN PACKAGE FOR LEAST SQUARES CURVE FITTING AND THE SOLUTION OF SIMULTANEOUS EQUATIONS

J. G. Hust and R. D. McCarty

GPO PRICE $

CFSTI PRICE(S) $

Hard copy (HC) 1.00

Microfiche (MF) .50

ff 653 July 65

NBS

FACILITY FORM 602

N67 12160 (ACCESSION NUMBER)

(THRU) 1

16 (PAGES)

(CODE)

CR-88100 (NASA CR OR TMX OR AD NUMBER)

08 (CATEGORY)

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS
BOULDER LABORATORIES
Boulder, Colorado

THE NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards is a principal focal point in the Federal Government for assuring maximum application of the physical and engineering sciences to the advancement of technology in industry and commerce. Its responsibilities include development and maintenance of the national standards of measurement, and the provisions of means for making measurements consistent with those standards; determination of physical constants and properties of materials; development of methods for testing materials, mechanisms, and structures, and making such tests as may be necessary, particularly for government agencies; cooperation in the establishment of standard practices for incorporation in codes and specifications; advisory service to government agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; assistance to industry, business, and consumers in the development and acceptance of commercial standards and simplified trade practice recommendations; administration of programs in cooperation with United States business groups and standards organizations for the development of international standards of practice; and maintenance of a clearinghouse for the collection and dissemination of scientific, technical, and engineering information. The scope of the Bureau's activities is suggested in the following listing of its four Institutes and their organizational units.

Institute for Basic Standards. Applied Mathematics. Electricity. Metrology. Mechanics. Heat. Atomic Physics. Physical Chemistry. Laboratory Astrophysics.* Radiation Physics. Radio Standards Laboratory:* Radio Standards Physics; Radio Standards Engineering. Office of Standard Reference Data.

Institute for Materials Research. Analytical Chemistry. Polymers. Metallurgy. Inorganic Materials. Reactor Radiations. Cryogenics.* Materials Evaluation Laboratory. Office of Standard Reference Materials.

Institute for Applied Technology. Building Research. Information Technology. Performance Test Development. Electronic Instrumentation. Textile and Apparel Technology Center. Technical Analysis. Office of Weights and Measures. Office of Engineering Standards. Office of Invention and Innovation. Office of Technical Resources. Clearinghouse for Federal Scientific and Technical Information.**

Central Radio Propagation Laboratory.* Ionospheric Telecommunications. Tropospheric Telecommunications. Space Environment Forecasting. Aeronomy.

* Located at Boulder, Colorado 80301.
** Located at 5285 Port Royal Road, Springfield, Virginia 22171.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT | **NBS REPORT**

31502-40-3150422 October 10, 1966 9258

FORTRAN PACKAGE FOR LEAST SQUARES CURVE FITTING

AND THE SOLUTION OF SIMULTANEOUS EQUATIONS

J. G. Hust and R. D. McCarty
Cryogenics Division
Institute for Materials Research
Boulder Laboratories
Boulder, Colorado

NBS

U.S. DEPARTMENT OF COMMERCE

NATIONAL BUREAU OF STANDARDS

CONTENTS

FORTRAN PACKAGE FOR LEAST SQUARES CURVE FITTING AND THE SOLUTION OF SIMULTANEOUS EQUATIONS

J. G. Hust and R. D. McCarty

FORTRAN subroutines are described for the solution of a least squares problem. Three subroutines are listed which: (1) establish the normal equations, (2) constrain the normal equation, and (3) perform matrix inversion by the Gauss-Jordan method. Subroutine (3) may also be used for the simultaneous solution of any set of linear equations.
Author

1. INTRODUCTION

Many mathematical techniques are presented in the literature for representing data. Of these methods, the most popular seems to be the method of least squares, not necessarily because it is the best, but because it is relatively simple and also because of the ease with which statistical information can be extracted from the data. The solution of the least squares problem, although mathematically unique, can be programmed in many ways. This report describes how this problem has been handled in the Evaluation Unit of the Cryogenic Data Center.

The package consists of three double precision subroutines. The first, DSUMUP, is used to establish the normal equation; the second, CONLSQ, is used to apply constraints if desired; the third, DGJFIT, is a Gauss-Jordan with pivotal condensation matrix inversion routine which computes and prints out statistical information along with the parameters. The third routine can also be used to solve any set of n linear equations for n linear coefficients. The computer codes for these subroutines are listed in the Appendix.

2. NORMAL EQUATIONS

We will consider only linear equations of the form

$$y = \sum_{i=1}^{m} a_i f_i \tag{1}$$

where a_i are the parameters and f_i and y are functions of the measured variables. Since often it is impossible to attach any significance to "dependent" and "independent" variables, this terminology is not used here. If an equation is non-linear in the parameters, it must be linearized before this package can be used.

Upon applying the least squares condition, i.e., minimize the weighted sum of the squared deviations of y, the normal equations (2) are obtained:

$$[F]\{a\} = \{B\} \tag{2}$$

where $[F]$ is a square array and $\{B\}$ is a column array of values dependent upon the f_i and y evaluated at each data point. This is described in more detail by Hust and McCarty [1].

The subroutine DSUMUP establishes the arrays $[F]$ and $\{B\}$ in a summation fashion, i.e., DSUMUP is called for each data point and the appropriate additive change in $[F]$ and $\{B\}$ is accomplished. Since $[F]$ and $\{B\}$ depend upon the data in an accumulative fashion, it is important that $[F]$ and $\{B\}$ are zero at the outset. This can be accomplished by calling DSUMUP with the argument NCOF equal to zero. The argument list of DSUMUP is as follows: C is an array generated by DSUMUP, the first NCOF columns represent $[F]$ and the next column represents $\{B\}$. G is an array representing f_i $(i = 1, 2 \ldots m)$ evaluated at the current data point. NCOF is the number of coefficients and thus, the order of $[F]$. Y is y_{exp} at the current data point. YY, WTSUM, and NPTSUM are generated by DSUMUP. YY is a quantity used to determine the sum of squares, WTSUM is the sum of the weights, and NPTSUM is the sum of points, i.e., the current number of points. WT is the weight assigned to the current point.

3. CONSTRAINTS

After having called DSUMUP for all data points, the constraints may be applied. It has been shown [1] that the constrained normal equations are given by

$$\left[\begin{array}{c|c} [F] & [g]^T \\ \hline [g] & 0 \end{array}\right] \left\{\begin{array}{c} \{a\} \\ \hline \{\lambda\} \end{array}\right\} = \left\{\begin{array}{c} \{B\} \\ \hline \{c\} \end{array}\right\} \qquad (3)$$

where $[g]\,\{a\} = \{c\}$ represents the constraints and $\{\lambda\}$ represents the Lagrangian multipliers.

The subroutine CONLSQ is used to modify (2) to obtain (3). CONLSQ is called once for each constraint and the arguments have the following meaning: C is the matrix generated by DSUMUP; G is the row of $[g]$ corresponding to the constraint currently being added, while Y is the corresponding element of $\{c\}$; NCOF is the number of parameters, while NCON indicates which constraint is currently being added; NCON should start with one and should increase by one for each subsequent constraint.

4. MATRIX INVERSION

The matrix inversion is accomplished by the Gauss-Jordan [2] reduction method. The Crout method has also been used by members of the Cryogenic Data Center and seems to be comparable in speed and yields results of comparable accuracy. The Gauss-Jordan method is used in the subroutine DGJFIT. The arguments are defined as follows:

DGJFIT

C is the augmented normal matrix as formed by DSUMUP and modified by CONLSQ upon entry and the variance - covariance matrix upon exit.

G is the coefficient array upon exit.

NCOF is the number of coefficients.

NCON is the number of constraints.

NPTS, YY, and WTSUM are input variables as generated by DSUMUP.

5. SAMPLE PROBLEM

An example of how these subroutines are used is included below.

Problem:

Fit the function: $Z = a_0 + a_1 x + a_2 x^2 + a_3 x^3$ to 10 data points with equal weights and apply the constraints:

$Z' = a_1 + 2a_2 x + 3a_3 x^2 = 0$ at $x = 2$, and

$Z'' = 2a_2 + 6a_3 x = 0$ at $x = 1$.

```
      :
      :
      DIMENSION X (10), Z(10), C(40,41), G(40)
      TYPE DOUBLE C,G,Y,YY,WTSUM
      NOPTS = 10
      NCOF = 4
  100 FORMAT (2F10.4)
      DO 1  I = 1, NOPTS
      READ 100, X (I), Z(I)
      DO 2 J = 1, NCOF
    2 G(J) = X(I)**(J-1)
      WT = 1.0
      Y = Z(I)
    1 CALL DSUMUP (C,G,NCOF,Y,YY,WTSUM NPTSUM)
      NCON = 1
      G(1) = 0.0
      G(2) = 1.0
      G(3) = 4.0
      G(4) = 12.0
      Y = 0.0
      CALL CONLSQ (C,G,NCOF,NCON,Y)
      NCON = 2
      G(1) = 0.0
      G(2) = 0.0
      G(3) = 2.0
      G(4) = 6.0
```

```
      Y = 0.0
      CALL CONLSQ (C,G,NCOF,NCON,Y)
      CALL DGJFIT (C,G,NCOF,NCON,NOPTS,YY,WTSUM)
      DO 3 I = 1, NOPTS
    3 ZCALC = G(1) + G(2)* X (I) + G(3)* X (I)**2 + G(4)* X (I)**3
      .
      .
      .
```

6. REFERENCES

[1] Hust, J. G., and R. D. McCarty, Cryogenics (to be published).

[2] Golden, J. T., (1965), FORTRAN IV - Programming and Computing (Prentice Hall, Inc., Englewood Cliffs, N. J.).

7. APPENDIX - COMPUTER CODE LISTINGS

The computer codes for the subroutines described in this report are listed below. Duplicate decks of these subroutines may be obtained from the Cryogenic Data Center, Cryogenics Division, National Bureau of Standards, Boulder, Colorado.

```
      SUBROUTINE CONLSQ(C,G,NCOF,NCON,Y)
C     C IS THE MATRIX BEING CONSTRAINED, G IS THE CURRENT ADDITION TO C,
C     Y IS THE CURRENT AUGMENTING COMPONENT ,NCOF IS THE NUMBER OF
C     COEFFICIENTS, NCON IS THE NUMBER OF THE CURRENT CONSTRAINT BEING
C     ADDED.
      DIMENSION G(40),C(40,41)
      TYPE DOUBLE C,G,Y
      NN=NCOF+NCON+1
      MM=NN-1
      JJ=NN-2
      II=NCOF+1
      DO 15 I=1,JJ
   15 C(I,NN)=C(I,MM)
      DO 50 I=1,NCOF
      C(I,MM)=G(I)
      C(MM,I)=G(I)
   50 CONTINUE
      DO 60 I=II,JJ
      C(I,MM)=0.0D
   60 CONTINUE
      C(MM,NN)=Y
      RETURN
      END
```

```
      SUBROUTINE DGJFIT(C,G,NCOF,NCON,NPTS,YY,WTSUM)
C     NCOF IS THE NUMBER OF COEFFICIENTS. NCON IS THE NUMBER OF
C     CONSTRAINTS.
C
C     MATRIX INVERSION WITH ACCOMPANYING SOLUTION OF LINEAR EQUATIONS
C
      DIMENSION IPIVOT(40),INDEX(40,2)
      DIMENSION           B(40),PIVOT(40),C(40,41),G(40)
      TYPE DOUBLE   B,PIVOT,C,G,DETERM,AMAX,SWAP,T,R1,YY,FPTS,FN,VARFIT,
     1    SDG,WTSUM
C     SAVE NORMAL MATRIX
      N=NCOF+NCON
      NN=N+1
      DO 6 I=1,N
    6 B(I)=C(I,NN)
      WRITE OUTPUT TAPE 61,101
      WRITE OUTPUT TAPE 61,201
      DO 851 I=1,N
  851 WRITE OUTPUT TAPE 61,301,(C(I,J),J=1,NN)
C
C     INITIALIZATION
C
   10 DETERM=1.0D
   15 DO 20 J=1,N
   20 IPIVOT(J)=0
   30 DO 550 I=1,N
C
C     SEARCH FOR PIVOT ELEMENT
C
   40 AMAX=0.0D
   45 DO 105 J=1,N
   50 IF (IPIVOT(J)-1) 60, 105, 60
   60 DO 100 K=1,N
   70 IF (IPIVOT(K)-1) 80, 100, 740
   80 IF (ABSF(AMAX)-ABSF(C(J,K))) 85, 100, 100
   85 IROW=J
   90 ICOLUM=K
   95 AMAX=C(J,K)
  100 CONTINUE
  105 CONTINUE
  110 IPIVOT(ICOLUM)=IPIVOT(ICOLUM)+1
C
C     INTERCHANGE ROWS TO PUT PIVOT ELEMENT ON DIAGONAL
C
  130 IF (IROW-ICOLUM) 140, 260, 140
  140 DETERM=-DETERM
  150 DO 200 L=1,N
  160 SWAP=C(IROW,L)
  170 C(IROW,L)=C(ICOLUM,L)
  200 C(ICOLUM,L)=SWAP
  205 CONTINUE
```

```
  210 CONTINUE
  220 SWAP=B(IROW)
  230 B(IROW)=B(ICOLUM)
  250 B(ICOLUM)=SWAP
  260 INDEX(I,1)=IROW
  270 INDEX(I,2)=ICOLUM
  310 PIVOT(I)=C(ICOLUM,ICOLUM)
  320 DETERM=DETERM*PIVOT(I)
C
C     DIVIDE PIVOT ROW BY PIVOT ELEMENT
C
  330 C(ICOLUM,ICOLUM)=1.0D
  340 DO 350 L=1,N
  350 C(ICOLUM,L)=C(ICOLUM,L)/PIVOT(I)
  355 CONTINUE
  360 CONTINUE
  370 B(ICOLUM)=B(ICOLUM)/PIVOT(I)
C
C     REDUCE NON-PIVOT ROWS
C
  380 DO 550 L1=1,N
  390 IF(L1-ICOLUM) 400, 550, 400
  400 T=C(L1,ICOLUM)
  420 C(L1,ICOLUM)=0.0D
  430 DO 450 L=1,N
  450 C(L1,L)=C(L1,L)-C(ICOLUM,L)*T
  455 CONTINUE
  460 CONTINUE
  500 B(L1)=B(L1)-B(ICOLUM)*T
  550 CONTINUE
C
C     INTERCHANGE COLUMNS
C
  600 DO 710 I=1,N
  610 L=N+1-I
  620 IF (INDEX(L,1)-INDEX(L,2)) 630, 710, 630
  630 JROW=INDEX(L,1)
  640 JCOLUM=INDEX(L,2)
  650 DO 705 K=1,N
  660 SWAP=C(K,JROW)
  670 C(K,JROW)=C(K,JCOLUM)
  700 C(K,JCOLUM)=SWAP
  705 CONTINUE
  710 CONTINUE
  740 CONTINUE
C     PLACE SOLUTION MATRIX INTO G
      MM=NCOF
      DO 749 I=1,MM
  749 G(I)=B(I)
C     NOW CALCULATE THE SUM OF SQUARES
      R1=0.0D
```

```
      DO 752 I7=1,N
  752 R1=R1-    C(I7,NN)*B(I7)
      R1=R1+YY
      WRITE OUTPUT TAPE 61,1100,DETERM
      WRITE OUTPUT TAPE 61,401,R1
C     CALCULATE THE VARIATION OF THE FIT AND THE RMS DEVIATION
      FPTS=NPTS
      FN=N
      VARFIT=R1/(FPTS-NCOF+NCON)
      WMSDEV=R1/WTSUM
      WRITE OUTPUT TAPE 61,1001,VARFIT
      WRITE OUTPUT TAPE 61,701,WMSDEV
      WRITE OUTPUT TAPE 61,1201,NPTS
       WRITE OUTPUT TAPE 61,1301,WTSUM
      WRITE OUTPUT TAPE 61,801
      DO 742 I=1,N
      WRITE OUTPUT TAPE 61,301,(C(I,J),J=1,N)
  742 CONTINUE
C     CALCULATE VARIANCE-COVARIANCE MATRIX
      DO 743 I=1,N
      DO 743 J=1,N
      C(I,J)=VARFIT*C(I,J)
  743 CONTINUE
      WRITE OUTPUT TAPE 61,901
      DO 744 I=1,N
      WRITE OUTPUT TAPE 61,301,(C(I,J),J=1,N)
  744 CONTINUE
      WRITE OUTPUT TAPE 61,501
      DO 852 I=1,MM
C     CALCULATE THE STANDARD DEVIATION OF THE COEFFICIENTS
      SDG=(C(I,I))**0.5
      RSDG=ABSF(SDG/G(I) )
      WRITE OUTPUT TAPE 61,301,G(I),SDG,RSDG
  852 CONTINUE
  101 FORMAT(31H1 AUGMENTED MATRIX TO BE SOLVED)
  201 FORMAT(1H0)
  301 FORMAT(8E16.8)
  401 FORMAT(38H0WEIGHTED SUM OF SQUARED DEVIATIONS = ,E16.8)
  501 FORMAT(  *1COEFFICIENTS, STANDARD DEVIATIONS AND RELATIVE STANDARD
     1 DEVIATIONS OF INPUT TERMS IN SAME ORDER*)
  601 FORMAT(1H1)
  701 FORMAT(43H0WEIGHTED MEAN OF THE SQUARED DEVIATIONS = ,E16.8)
  801 FORMAT(25H1INVERSE MATRIX           )
  901 FORMAT(43H1VARIANCE-COVARIANCE MATRIX OF COEFFICIENTS)
 1001 FORMAT(23H0VARIANCE OF THE FIT = ,E16.8)
 1100 FORMAT(40H1THE DETERMINANT OF THE NORMAL MATRIX = ,E16.8)
 1201 FORMAT(27H0NUMBER OF POINTS FITTED = ,I6)
 1301 FORMAT(18H0SUM OF WEIGHTS = ,E16.8)
      RETURN
      END
```

```
      SUBROUTINE DSUMUP(C,G,NCOF,Y,YY,WT,WTSUM,NPTSUM)
C     IF DIMENSION CARD IS CHANGED CHANGE CARD 20 ALSO
C     IF N=0 THE C MATRIX, NPTSUM,WTSUM AND YY ARE EQUATED TO ZERO
      DIMENSION C(40,41),G(40),GWT(40)
      TYPE DOUBLE C,G,Y,YY ,WT      ,GWT,YWT,WTSUM
      IF(NCOF)20,20,3
    3 CONTINUE
      N=NCOF
      NN=N+1
      NPTSUM=NPTSUM+1
      WTSUM=WTSUM+WT
      DO 5 I=1,N
      GWT(I)=G(I)*WT**0.5
    5 CONTINUE
      YWT=Y*WT**0.5
   10 DO 2 I=1,N
      DO 1 J=1,N
    1 C(I,J)=C(I,J)+GWT(I)*GWT(J)
    2 C(I,NN)=C(I,NN)+YWT*GWT(I)
      YY=YY+YWT**2
      RETURN
   20 M=40
      NN=M+1
      DO 30 I=1,M
      DO 30 J=1,NN
   30 C(I,J)=0.0D
      NPTSUM=0
      WTSUM=0.0
      YY=0.0D
      RETURN
      END
```

U.S. DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20230

OFFICIAL BUSINESS

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF COMMERCE