SHARE PROGRAM LIBRARY AGENCY

PROGRAM NUMBER

360D-13.6.003

University of Miami

1365 MEMORIAL DRIVE - CORAL GABLES, FLORIDA
(305) - 284-6257

360D-13.6.003

NONLINEAR PARAMETER ESTIMATION AND PROGRAMMING

Yonathan Bard

December 1967

Direct all inquiries to:

Mr. Yonathan Bard
IBM Cambridge Scientific Center
545 Technology Square
Cambridge, MA 02139

TABLE OF CONTENTS

NONLINEAR PARAMETER ESTIMATION AND PROGRAMMING

Chapter I
PROGRAM BRIEF

Scientists frequently propose mathematical models or equations which attempt to describe the relationships between physically measured variables. These models usually contain parameters or coefficients whose values are unknown. It is the object of parameter estimation to find values of these parameters that cause the proposed equations to give the best possible fit to the data. The measure of fit depends on the residuals, i.e. on the differences between the observed values of certain variables, and the values predicted by the equations. A common such measure is the sum of squares of the residuals, but more sophisticated measures exist, and are described in the theory section of this manual.

The present program is designed to estimate unknown parameters in a variety of models, using a variety of best-fit criteria. The models may consist of single or multiple equations. The equations may be stated explicitly, or implicitly as solutions of sets of differential equations, or as functions of such solutions, with known or unknown initial conditions. The differential equations may, in particular, represent the progress of chemical reactions. The best fit criteria employed may be least squares (for single equations), weighted (by equation, i.e. by independent variable) least squares (with known or unknown weights), and maximum likelihood (with known or unknown, diagonal or full, covariance

matrices). Prior distributions (normal or other) and bounds or constraints on the parameters may be specified for Bayesian estimation. Variable weights by experiment and missing data items may be handled by means of special devices.

The user not interested in the more complex aspects of estimation, and whose problems assume the form of minimizing the sum of squares of explicitly defined functions, may restrict his attention to the material contained in the section "Single Equation Least Squares", starting on page 135.

The program may also be used to solve general nonlinear programming problems, i.e. to find the maximum of a function, with variables that may be subject to constraints. The function and/or constraints may be linear or nonlinear. The program should be efficient (relative to other nonlinear programming codes) where the function is highly nonlinear, and the constraints are relatively simple and few in number. It should perform best for unconstrainted problems, or problems for which the maximum is an interior point of the feasible region.

The program provides two alternative subroutines for finding the maximum: one uses a modified Gauss-Newton method, the other, the Davidson-Fletcher-Powell method.

Program Limitations

As supplied, the program contains the following (easily changed) restrictions on problem size:

20	parameters
5	equations (observed variables)
100	experiments (200 for single equations least squares)
10	variables per experiment
40	constraints

For differential equations:

(a) Parameters: These are constants appearing in the model. Their numerical values are unknown. The vector of parameters is denoted by

$$\theta = \{\theta_1, \theta_2, \ldots \theta_\ell\}$$

(b) Independent Variables: (Also referred to as experimental conditions.) These are variables which either can be fixed arbitrarily for each experiment, or which are known precisely for each observation. The vector of independent variables for the μ-th experiment is denoted by

$$a_\mu = \{a_{\mu 1}, a_{\mu 2}, \ldots a_{\mu m}\} \quad .$$

(c) Dependent Variables: These are the variables whose values the equations of the model seek to predict on the basis of the known values of the independent variables.

(d) Observed Variables: Those of the dependent variables which are actually measured in each experiment are called observed variables. The vector of observed variables for the μ-th experiment is denoted by

$$y_\mu = \{y_{\mu 1}, y_{\mu 2}, \ldots y_{\mu k}\} \quad .$$

If the same dependent variable is measured several times in each experiment, or under identical experimental conditions, we regard each replication as a separate observed variable; e.g., $y_{\mu 1}$ may represent the first measurement of some variable under the μ-th set of experimental conditions, and $y_{\mu 2}$ may represent the second measurement of the same variable under the same experimental conditions.

An experiment consists of the measurement of all observed variables for a given set of values of the independent variables.

Examples: In a chemical reaction, the rate of the reaction may be the observed variable; the concentrations of

the reactants -- the independent variables; and the forward and backward reaction rate constants -- the parameters.

In an economic time series, the date may be the independent variable; population and volume of sales -- the observed variables; and constants relating these -- the parameters.

The model relates the observed variables to the dependent variables and parameters. It has the form:

$$g(y_\mu, a_\mu, \theta) = 0 \qquad (\mu = 1,2,\ldots n) \qquad (1)$$

where g is a vector of functions. Its dimensionality must equal that of the vector of observed variables. Written out fully, equations (1) become:

$$g_1(y_{\mu 1}, y_{\mu 2}, \ldots y_{\mu k}, a_{\mu 1} \ldots a_{\mu m}, \theta_1, \ldots \theta_\ell) = 0 \qquad (\mu = 1,2,\ldots n)$$

$$g_2(y_{\mu 1}, y_{\mu 2}, \ldots y_{\mu k}, a_{\mu 1} \ldots a_{\mu m}, \theta_1, \ldots \theta_\ell) = 0 \qquad (\mu = 1,2,\ldots n)$$

$$g_k(y_{\mu 1}, y_{\mu 2}, \ldots y_{\mu k}, a_{\mu 1} \ldots a_{\mu m}, \theta_1, \ldots \theta_\ell) = 0 \qquad (\mu = 1,2,\ldots n)$$

Specifying the model requires assigning specific mathematical expressions to the functions $g_1, g_2, \ldots g_k$.

Equations (1) are known as the structural equations. In principle they may be solved, at each experiment, for the observed variables, in which case we obtain the reduced form, or explicit equations:

$$y_\mu = f(a_\mu, \theta) \qquad (\mu = 1,2,\ldots n) \qquad (2)$$

In case of replications, several elements of the vector of functions f will be identical in form.

Best Fit Conditions

Ideally one would like to find values of the θ that will satisfy equations (2) exactly for each experiment. Due to

where E denotes expected value and σ is a constant, then it is easy to show that maximum likelihood is equivalent to least squares.

If $k > 1$ and the u's are distributed as before, but with

$$E(u_{\mu i} u_{\eta j}) = \frac{\sigma^2}{W_i} \delta_{\mu\eta} \delta_{ij} \tag{10}$$

then maximum likelihood reduces to weighted least squares.

(c) Bayesian Estimation: An expression such as $p(u,\phi)$ states the probability distribution of the y's given the θ's. In fact, however, the y's are given and the θ's are sought. We may conceive of many possible universes, each with its own values of θ. Given the observations, we may assign relative probabilities to all these universes, and ask which of these universes is the most probable.* We maximize $p(\theta|y_\mu)$, i.e., the probability density of θ given the observations y_μ. According to Bayes' formula:

$$p(\theta|y_\mu) = \frac{1}{c} p(y_\mu|\theta) p_o(\theta) \tag{11}$$

where $c = \int p(y_\mu|\theta) p_o(\theta) d\theta$ is a normalizing constant; $p(y_\mu|\theta)$ is the probability density of y_μ given θ, and is identical with $p(u_\mu, \phi)$ previously defined; and $p_o(\theta)$ is the prior distribution of θ, i.e., it is the probability distribution one would have assigned to θ prior to having made the observations y_μ.

Ignoring the irrelevant constant c, we seek then, to maximize:

$$G_{B.}(\theta,\phi) = \log p(f(a_\mu, \theta) - y_\mu, \phi) + \log p_o(\theta) \tag{12}$$

If absolutely nothing is known a priori about the θ, one assumes a uniform prior distribution. Then $p_o(\theta) = \text{constant}$, and maximizing $p(\theta|y_\mu)$ is equivalent to maximizing $p(u,\phi)$, i.e., the Bayesian

*This heuristic explanation of Bayesian estimation is frowned upon by statistical theory. For a summary of current thought on the subject, see the article by Cornfield [4].

and maximum likelihood estimates will be identical. Appropriate choices of $p_o(\theta)$, however, enable one to confine the θ's to specified regions (e.g., all θ's positive) or in general to influence one's final estimate with any prior knowledge we may possess about the θ's. A particularly useful application is the re-estimation of parameters as the results of additional experiments become available.

This is not the place to discuss the statistical properties of the estimates obtained by application of these criteria. The reader is referred to the standard texts on statistics and data analysis. From the practical point of view, the maximum likelihood principle, particularly in its Bayesian interpretation, is a very flexible tool which permits one to make use of all prior information one possesses in order to obtain estimates of the unknown parameters, both in the model and in the distribution. In principle one should consider the distribution as part of the model.

$$E(u_{\mu i}u_{\mu j}) = V_{ij} \qquad \begin{matrix}(i,j = 1, 2,\ldots k\\ \mu = 1, 2,\ldots n)\end{matrix} \quad (16)$$

thus, the likelihood function assumes the form:

$$G_{M.L.} = -\frac{kn}{2}\log(2\pi) - \frac{n}{2}\log\det V - \frac{1}{2}\sum_{\mu=1}^{n}\sum_{ij=1}^{k}[V^{-1}]_{ij}u_{\mu i}u_{\mu j} \quad (17)$$

The following options are available concerning the covariance matrix V :

1) V is a known diagonal matrix
2) V is a known symmetric matrix
3) V is proportional to a known diagonal matrix
4) V is proportional to a known symmetric matrix
5) V is an unknown diagonal matrix
6) V is an unknown symmetric matrix

The set ϕ of unknown distribution parameters is empty in cases 1) and 2), consists of a constant of proportionality in cases 3) and 4), and of the unknown matrix elements in cases 5) and 6).

Cases 3) and 4) are equivalent to weighted least squares with case 4) taking into account cross product terms.

When $k = 1$, i.e., when only one variable is observed per experiment, the matrix V reduces to a single number σ^2 , and all six cases are equivalent to least squares.

In cases 1) and 2), the total number of observed values $(= nk)$ must at least equal the number of unknown parameters ℓ . In cases 3) and 4) we must have $nk > \ell$, and in cases 5) and 6) the number of experiments n must exceed the largest number of parameters appearing in any one expression $f(a,\theta)$.

(b) Constraints The values that the parameters can assume may be restricted so as to satisfy any constraints the user wishes to specify. These constraints may reflect physical knowledge (e.g., reaction rate constants must be positive), or they may be imposed simply to guard against divergence of the computation. The following options are available:

(b.1) No constraints

(b.2) An upper and lower bound on each parameter:

$$\alpha_i \le \theta_i \le \beta_i \quad (i = 1, 2,\ldots\ell) \quad (18)$$

(b.3) Any arbitrary set of inequality constraints, to be programmed by the user.

(c) The Prior Distribution The function $p_o(\theta)$ is always assumed equal zero outside the range specified by the constraints. Inside the feasible region, the following options are available:

(c.1) A constant, reflecting no additional prior knowledge on the values of the θ's.

(c.2) A multivariate normal distribution with given means $\bar{\theta}$ and covariance martix W ; i.e.,

$$p_o(\theta) = (2\pi)^{-\frac{\ell}{2}}(\det W)^{-1}\exp\Big[-\frac{1}{2}\sum_{i,j}^{\ell}(\theta_i-\bar{\theta}_i)[W^{-1}]_{ij}(\theta_j-\bar{\theta}_j)\Big] \quad (19)$$

This option may be used to advantage in sequential estimation (see page 54).

(c.3) Any arbitrary prior may be coded by the user.

(d) The Model The program accepts only models in reduced form, i.e., having the observed variables given explicitly as functions of the independent variables, as in equation (2). Implicit equations such as (1) may be treated as described on page

For least squares estimation of models containing a single equation, the equation may be in implicit form and no variable need be singled out as the observed variable. The function minimized has the form:

$$G = \sum_{\mu=1}^{n} g(x_\mu, \theta)^2$$

where x_μ is the set of all variables for the μ-th experiment.

The equations making up the model, being arbitrary in form, naturally must be programmed by the user. Special provisions are provided, however, for models involving solutions of ordinary differential equations, and for models involving typical rate equations of chemical kinetics.

At time t_μ (representing the μ-th experiment) we withdraw three samples and perform the following tests on them:

(1) We determine $x_{\mu 1} = x_1(t_\mu)$ directly by titration.

(2) A different person independently determines $x_{\mu 1}$ by titration.

(3) We determine the light absorptivity of the solution. This is assumed to be a linear function of the concentrations with unknown coefficients.

Denoting the results of these three measurements as $y_{\mu 1}$, $y_{\mu 2}$, and $y_{\mu 3}$, we have:

$$\begin{aligned} y_{\mu 1} &= x_{\mu 1} - u_{\mu 1} \\ y_{\mu 2} &= x_{\mu 1} - u_{\mu 2} \\ y_{\mu 3} &= \beta_0 + \beta_1 x_{\mu 1} + \beta_2 x_{\mu 2} + \beta_3 x_{\mu 3} - u_{\mu 3} \end{aligned} \tag{24}$$

where the $u_{\mu i}$ are, as usual, the errors or residuals.

Equations (22), (23), and (24) together constitute the model. The constants ψ, k_1, k_2, k_3, β_0, β_1, β_2, and β_3 together constitute the set of unknown parameters θ. Note that the model equations do indeed specify (though indirectly) the relation between the observed variables and the unknown parameters, as required by equation (2). In fact, where the differential equations can be integrated analytically, equations (2) may be written down explicitly.

A series of experiments performed at different times but with identical initial conditions constitute a <u>run</u>. Among the independent variables we distinguish between those which vary from experiment to experiment (e.g. the time), and those which vary only from run to run (e.g. known initial conditions; the temperature in an isothermal chemical reaction).

The accompanying computer program contains subroutines that numerically integrate the differential equations. There

are three options regarding the initial conditions:

(1) All initial conditions are known for all runs.

(2) Some initial conditions for some runs constitute unknown parameters.

(3) Some initial conditions for some runs are arbitrary functions (to be specified by the user) of some unknown parameters.

Concerning the relation between observed and state variables, i.e., the functions d in equation (21), three options are available:

(1) The observed variables are the first k among the state variables, i.e.,

$$y_{\mu i} = x_i(t_\mu) \qquad (i = 1, 2, \ldots k) \tag{25}$$

(2) The observed variables are linear functions of the state variables, i.e.,

$$y_{\mu i} = b_i + \sum_{j=1}^{s} B_{ij} x_j(t_\mu) \tag{26}$$

The elements of the observation vector b and the observation matrix B may be known, or may (in part) be unknown parameters.

In the above example, we have:

$$b = \begin{bmatrix} 0 \\ 0 \\ \beta_0 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \beta_1 & \beta_2 & \beta_3 \end{bmatrix}$$

with β_0, β_1, β_2, β_3 unknown parameters.

Note that (1) is a special case of (2), with $b_i = 0$, and $B_{ij} = \delta_{ij}$.

$$L = (1 + \sum_k M_k x_k^{m_k})^p \qquad (30)$$

The M_k may have temperature dependencies similar to equation (29). The m_k and p are constants.

The kinetics subroutine allows the determination of the k_i, K_i, and M_i (with or without temperature dependence) as unknown parameters. Determination of the exponents n_{ji}, $n_{\ell i}$, m_k, and p has to be done on a case-study basis. Usually, only integral or simple fractional values are allowed for these exponents. To estimate any of these exponents should they be allowed to assume a continuous range of values, the user must code his own routine FUN.

Only isothermal reactions can be handled by the kinetics subroutine. A different temperature may be specified for each run. For non-isothermal reactions, the user must code his own routine FUN.

D. METHOD OF SOLUTION

Maximization

The central function of the program is to find the values of the parameters that maximize a function (usually a likelihood or log probability density). Two alternative subroutines are provided for effecting the maximization. These are:

(1) The Gauss-Newton method, with modifications by Greenstadt-Eisenpress [6], Bard [1], and Carroll [3].

(2) The Davidon-Fletcher-Powell method [5].

The following is a guide for choosing which method to apply to any given problem.

(1) The Gauss-Newton method is expected to be faster for parameter estimation problems, particularly when the fit of the model to the data is expected to be fairly good.

(2) In nonlinear programming problems, and in Bayesian estimation with a complicated prior destribution, Davidon's method may be preferred because it does not require the user to code second derivatives.

(3) In parameter estimation problems, Davidon's method is expected to give a better estimate of the covariance matrix of the estimates. The advantage increases as the fit becomes poorer.

(4) Davidon's method requires less storage for maximum likelihood and Bayesian estimation problems. See page 101 on how to obtain this advantage.

(5) On any problem for which one of the methods has failed to produce a satisfactory solution, it is worth while to try the other method.

Davidon's method is fully described in reference [5]. We describe below the current implementation of the Gauss-

$\theta^{(o)} + \lambda_3 \Delta\theta$ is infeasible, we truncate λ_3 appropriately.

We consider the following cases:

(1) $\Gamma(\lambda_2) > \Gamma(0)$. If $\frac{|\lambda_3 - \lambda_2|}{\lambda_2} < .1$, we set $\lambda = \lambda_2$. Otherwise we set $\lambda = \lambda_2$ or $\lambda = \lambda_3$ and proceed to the next iteration depending on whether $\Gamma(\lambda_2)$ is or is not greater than $\Gamma(\lambda_3)$.

(2) $\Gamma(\lambda_2) \leq \Gamma(0)$. We set $\lambda_4 = \max(\lambda_3, 1/4\ \lambda_2)$ and compute $\Gamma(\lambda_4)$. If $\Gamma(\lambda_4) > \Gamma(0)$ we set $\lambda = \lambda_4$ and proceed to the next iteration. Otherwise we replace λ_2 with λ_4 , draw a new parabola, and reproceed as before.

Convergence

In principle, (i.e., if the above procedure is repeated ad infinitum) the method should (except in extremely pathological cases) converge to a stationary point of the objective function, provided the latter is smooth (possesses continuous first derivatives) and all values of the function and its first derivatives are computed precisely (second derivatives need be known only approximately). In practice we must stop the iterations at some point, and therefore a termination criterion is required.

Termination

The process is terminated when each component of the vector $\lambda\Delta\theta$ is so small as to satisfy the inequality:

$$|\lambda\Delta\theta_i| \leq .0001(.001 + |\theta_i^{(o)}|) \qquad (i = 1,2, \ldots \ell) \qquad (33)$$

This criterion has been suggested by Marquardt [9]. It does not guarantee that the maximum has been reached, but seems

to work well in practice. If at the time of termination the matrix $-Q$ is not positive definite (i.e., the eigenvalues of Q , which are printed out, are not all negative) it is probable that the process has converged to a boundary point of the admissible region, to a saddle point, or not at all. An indication of convergence to an (at least local) interior maximum is obtained if the elements of p are small compared to the corresponding diagonal elements of $-Q$, i.e., when for $i = 1, 2, \ldots \ell$ we have $Q_{ii} < 0$, $|\frac{p_i}{Q_{ii}}| << 1$.

Penalty Functions

It was stated above that the objective function must be smooth, i.e., that the first derivative be continuous. When bounds (or more complex constraints) are placed on the θ's the prior distribution function $p_o(\theta)$ is discontinuous; it is zero outside the admissible region and finite inside it. We smooth out this jump discontinuity by means of the following device.

Suppose the restrictions placed on the θ's are stated in the form of the inequalities:

$$z_i(\theta) \leq 0 \qquad (i = 1,2, \ldots r) \qquad (34)$$

where the z_i are specified functions. For instance, in the case of bounds $\alpha_i \leq \theta_i \leq \beta_i$ the corresponding z's would be:

$$z_{2i-1}(\theta) = \alpha_i - \theta_i \leq 0$$

$$z_{2i}(\theta) = \theta_i - \beta_i \leq 0$$

We now replace the prior distribution $p_o(\theta)$ by the function $p_o^*(\theta) = p_o(\theta) \exp \sum_{i=1}^{} \frac{\gamma_i}{z_i(\theta)}$, where the γ_i are preassigned constants. If the γ_i are sufficiently small, p_o^* will not

E. DETAILS OF PARAMETER ESTIMATION

The Likelihood Function

For a normal distribution the likelihood function has the form:

$$G_{M.L.} = -\frac{kn}{2}\log(2\pi) - \frac{n}{2}\log\det V - \frac{1}{2}\sum_{ij}[V^{-1}]_{ij}A_{ij} \qquad (38)$$

Where A_{ij} is the moment matrix of the residuals:

$$A_{ij} = \sum_{\mu=1}^{n} u_{\mu i}u_{\mu j} = \sum_{\mu=1}^{n}[f_i(a_\mu,\theta) - y_{\mu i}][f_j(a_\mu,\theta) - y_{\mu j}] \qquad (39)$$

The function $G_{M.L.}$ depends on the θ's only through the elements A_{ij} of the moment matrix. Hence,

$$\frac{\partial G_{M.L.}}{\partial\theta_\alpha} = \sum_{i,j}\frac{\partial G}{\partial A_{ij}}\frac{\partial A_{ij}}{\partial\theta_\alpha} \qquad (40)$$

$$\frac{\partial^2 G_{M.L.}}{\partial\theta_\alpha\partial\theta_\beta} = \sum_{i,j}\frac{\partial G}{\partial A_{ij}}\frac{\partial^2 A_{ij}}{\partial\theta_\alpha\partial\theta_\beta} + \sum_{i,j,k,\ell}\frac{\partial^2 G}{\partial A_{ij}\partial A_{k\ell}}\frac{\partial A_{ij}}{\partial\theta_\alpha}\frac{\partial A_{k\ell}}{\partial\theta_\beta} \qquad (41)$$

Differentiating (39):

$$\frac{\partial A_{ij}}{\partial\theta_\alpha} = \sum_\mu\left(u_{\mu i}\frac{\partial f_{\mu j}}{\partial\theta_\alpha} + u_{\mu j}\frac{\partial f_{\mu i}}{\partial\theta_\alpha}\right) \qquad (42)$$

$$\frac{\partial^2 A_{ij}}{\partial\theta_\alpha\partial\theta_\beta} = \sum_\mu\left(\frac{\partial f_{\mu i}}{\partial\theta_\alpha}\frac{\partial f_{\mu j}}{\partial\theta_\beta} + \frac{\partial f_{\mu i}}{\partial\theta_\beta}\frac{\partial f_{\mu j}}{\partial\theta_\alpha} + u_{\mu i}\frac{\partial^2 f_{\mu j}}{\partial\theta_\alpha\partial\theta_\beta} + u_{\mu j}\frac{\partial^2 f_{\mu i}}{\partial\theta_\alpha\partial\theta_\beta}\right) \qquad (43)$$

If the model fits the data at all well, the values of

$u_{\mu i}$ and $u_{\mu j}$ will be small, and the term involving the second derivatives of the f's may be neglected. In fact, we always neglect these terms. In the least squares case this approximation is known as the Gauss Method. In most cases this generalization of the Gauss Method has required no more iterations than the Newton-Raphson-Greenstadt Method using correct second derivatives. The amount of computation saved by not requiring second derivatives of the f's can be vast.

As will be recalled, the following six options are available for the covariance matrix V :

(1) $V_{ij} = \delta_{ij}v_i$, v_i known

(2) V_{ij} known

(3) $V_{ij} = \delta_{ij}v_i\tau$, v_i known, τ unknown

(4) $V_{ij} = v_{ij}\tau$, v_{ij} known, τ unknown

(5) $V_{ij} = \delta_{ij}v_i$, v_i unknown

(6) $V_{ij} = v_{ij}$, v_{ij} unknown

Where there are unknown distribution parameters, i.e., τ , v_i , or v_{ij} , we proceed by the method of stagewise maximization (Koopmans & Hood [8]). In case (3), for instance, we differentiate (44) with respect to τ and set the derivative equal to zero. Equation (44) was obtained by substituting the proper expression for V in (38).

$$G_{M.L.} = -\frac{kn}{2}\log(2\pi) - \frac{kn}{2}\log\tau - \frac{n}{2}\sum_{i=1}^{k}\log v_i - \frac{1}{2\tau}\sum_{i=1}^{k}\frac{1}{v_i}\sum_{\mu=1}^{n}u_{\mu i}^2 \qquad (44)$$

$$\frac{\partial G}{\partial\tau} = -\frac{kn}{2\tau} + \frac{1}{2\tau^2}\sum_i\frac{1}{v_i}\sum_\mu u_{\mu i}^2 = 0 \qquad (45)$$

(2) Arbitrary Covariance

We define:

$$A_{ij} = \sum_{\mu} u_{\mu i} u_{\mu j} \qquad D = \sum_{ij} T_{ij} A_{ij}$$

$$B_{ij\alpha} = \sum_{\mu} u_{\mu i} \frac{\partial f_{\mu j}}{\partial \theta_\alpha} \qquad E_\alpha = \sum_{ij} T_{ij} B_{ij\alpha}$$

$$C_{ij\alpha\beta} = \sum_{\mu} \frac{\partial f_{\mu i}}{\partial \theta_\alpha} \frac{\partial f_{\mu j}}{\partial \theta_\beta} \qquad F_{\alpha\beta} = \sum_{ij} T_{ij} C_{ij\alpha\beta}$$

$$G_{kh\alpha} = \sum_{m} T_{mk}(B_{hm\alpha} + B_{mh\alpha})$$

Where T_{ij} is defined below for each case. (50)

TABLE 2: Arbitrary Covariance

	Case 2	Case 4	Case 6
$T_{ij} =$	$[v^{-1}]_{ij}$	$[v^{-1}]_{ij}$	$n[A^{-1}]_{ij}$
$L =$	$-\frac{1}{2}D$	$-\frac{kn}{2}\log\frac{1}{kn}D$	$-\frac{n}{2}\log\det\frac{1}{n}A$
$\frac{\partial L}{\partial \theta_\alpha} =$	$-E_\alpha$	$-\frac{kn}{D}E_\alpha$	$-E_\alpha$
$\frac{\partial^2 L}{\partial\theta_\alpha\partial\theta_\beta} =$ (Approximately, neglecting $\frac{\partial^2 L}{\partial\theta_\alpha\partial\theta_\beta}$.)	$-F_{\alpha\beta}$	$-\frac{kn}{D}(F_{\alpha\beta}-\frac{2}{D}E_\alpha E_\beta)$	$-(F_{\alpha\beta}-\frac{1}{2n}\sum_{ik} G_{ik\alpha}G_{ki\beta})$
$G_{M.L.} =$	$-\frac{kn}{2}\log(2\pi)-\frac{n}{2}\log\det v+L$	$-\frac{kn}{2}[1+\log(2\pi)]-\frac{n}{2}\log\det v+L$	$-\frac{kn}{2}[1+\log(2\pi)]+L$
$\tau =$	See note on previous page.	$\exp(-\frac{2L}{kn})$ (Multiply by $\frac{1}{1-\frac{\ell}{kn}}$ to obtain unbiased estimate.)	
$V_{ij} =$	v_{ij}	τv_{ij}	$\frac{1}{n}A_{ij}$

The Prior Distribution

For a normal prior distribution of known means $\bar{\theta}$ and covariance V we have, omitting constants:

$$\log p_o = -\frac{1}{2}\sum_{ij=1}^{\ell} W_{ij}(\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j) \tag{51}$$

$$\frac{\partial \log p_o}{\partial \theta_\alpha} = -\sum_{i=1}^{\ell} W_{\alpha i}(\theta_i - \bar{\theta}_i) \tag{52}$$

$$\frac{\partial^2 \log p_o}{\partial\theta_\alpha\partial\theta_\beta} = -W_{\alpha\beta} \tag{53}$$

where $W_{ij} = [V^{-1}]_{ij}$.

Covariance Estimates

As is well known, the matrix $\left[\frac{\partial^2 G}{\partial\theta_\alpha\partial\theta_\beta}\right]^{-1}$ forms an estimate of the covariance matrix of the parameters. If the covariance of the errors in the observed variables is given in the form $V = \tau v$, where v is a known matrix, then $\exp(-\frac{2Lmax}{kn})$ is the maximum likelihood estimate of τ . This estimate is biased. An unbiased estimate is given by:

$$\tau = \frac{1}{1-\frac{\ell}{kn}}\exp(-\frac{2Lmax}{kn}) \tag{54}$$

where Lmax is the maximum value of L .

If V is unknown, its maximum likelihood estimate is:

$$V_{ij} = \frac{1}{n}A_{ij} = \frac{1}{n}\sum_{\mu} u_{\mu i}u_{\mu j} \tag{55}$$

Let $\frac{dZ}{dt} = \phi(Z,t)$ denote the complete set of equations, with $Z(0) = Z_o$. We select our initial time step τ, and set

$$Z_1 = Z_o + \phi(Z_o,0)\tau$$

$$Z_2 = Z_o + 1/2\ [\phi(Z_o,0) + \phi(Z_1,\tau)]\tau$$

$$Z_3 + Z_o + 1/2\ [\phi(Z_o,0) + \phi(Z_2,\tau)]\tau$$

$$\Delta = \frac{\sum_{i=1}^{s} |Z_3 - Z_2|}{\sum_{i=1}^{s} |Z_3|}$$

If $\Delta > .9X10^{-5}$, we compute

$$\tau_1 = (\frac{.9X10^{-5}}{\Delta})^{\frac{2}{3}}\tau$$

and replace τ with τ_1. We recalculate Z_1, Z_2, Z_3, and Δ and continue until $\Delta \leq .9X10^{-5}$. Now we set $Z(\tau) = \frac{2}{3} Z_3 + \frac{1}{3} Z_2$.

Supposing $Z(t - \tau)$ and $Z(t)$ have been calculated, and the next time step to be taken has length τ_1 (initially $\tau_1 = \tau$). We set

$$\alpha = 1 + \frac{\tau_1}{2\tau}$$

$$\beta = \frac{\tau_1}{3(\tau + \tau_1)}$$

$$Z_o = Z(t)$$

$$Z_{-1} = Z(t - \tau)$$

$$Z_1 = Z_o + [\alpha\phi(Z_o,t) + (1 - \alpha)\phi(Z_{-1},t - \tau\ ]\tau_1$$

$$Z_2 = Z_o + \frac{1}{2}\ [\phi(Z_o,t) + \phi(Z_1,t + \tau_1)]\tau_1$$

$$Z_3 = \beta Z_1 + (1 - \beta)Z_2$$

Then we accept $Z(t + \tau_1) = Z_3$. We now calculate

$$\Delta = \frac{\sum_{i=1}^{s} |Z_{2i} - Z_{1i}|}{\sum_{i=1}^{s} |Z_{2i}|}$$

$$\tau_2 = \left|\frac{.9X10^{-5}(\tau + \tau_1)\tau^2}{\Delta}\right|^{1/3}$$

and replace, for the next intergration step: t with $t + \tau_1$; τ with τ_1 ; and τ_1 with τ_2. If $t < t_\mu \leq t + \tau_1$, we replace τ_1 with $t_\mu - t$, where t_μ is the time for the μ-th experiment. This method will usually result in values of the state variables accurate to four or five decimal places.

H. DEGENERACY

The maximization routine will sometimes fail to converge because of ill conditioning in the model or the data. We discuss the most important cases below, and prescribe remedies to the situations.

(1) Correlated Parameters. It may happen that the data has been obtained in such a way that it is impossible to estimate certain parameters. For instance, suppose our model has the form:

$$y_\mu = \theta_1 a_{\mu 1} + \theta_2 a_{\mu 2} \tag{61}$$

and under all experimental conditions the independent variables were about equal to each other, i.e., $a_{\mu 1} = a_{\mu 2}$ for all μ. Then the equation

$$y_\mu = (\theta_1 + \theta_2) a_{\mu 1} \tag{62}$$

would represent the data just as well as (61). We cannot hope to estimate θ_1 and θ_2 independently, although their sum may be accurately determined. In the course of the calculations, such a situation reveals itself by the matrix $\frac{\partial^2 G}{\partial\theta_i \partial\theta_j}$ becoming singular, or nearly so. In this case, the matrix R (see equation (32)) cannot be formed, or is itself nearly singular -- in which case the vector Rp no longer necessarily leads in a direction in which G increases, and the maximization procedure breaks down.

The precise cause of the breakdown is revealed by means of an eigenvalue analysis of the matrix $\frac{\partial^2 G}{\partial\theta_i \partial\theta_j}$. Eigenvectors corresponding to large (in

absolute value) eigenvalues reveal those parameters (or linear combinations thereof) which are well determined; vectors corresponding to small eigenvalues reveal ill determined combinations. In the above case, we would have one large eigenvalue with eigenvector $\frac{1}{2}(\sqrt{2}, \sqrt{2})$, showing that the quantity $\frac{\sqrt{2}}{2}(\theta_1 + \theta_2)$ (and therefore $\theta_1 + \theta_2$) is well determined. The second eigenvalue would be very much smaller, with eigenvector $\frac{1}{2}(\sqrt{2}, -\sqrt{2})$, showing the quantity $\frac{\sqrt{2}}{2}(\theta_1 - \theta_2)$ to be ill determined.

The eigenvalues and vectors are printed out at the end of the maximization procedure, regardless of whether or not it converged.

No mathematical technique will produce acceptable estimates of ill determined parameters. The experimenter's only recourse is to design his experiments in a suitable way. A method of doing so is given in papers by G. E. P. Box and coworkers [2].

(2) Correlated Variables. In some cases the model equations (2) will be linearly dependent. This will certainly happen when several of the observed variables represent replicate measurements of the same dependent variables. It may also happen in other cases: for instance, when the observed variables are concentrations of several reacting chemicals subject to a material balance. When options (1)-(5) concerning the nature of the covariance matrix of the residuals are used, no problems should arise on this score. However, if option (6) applies, i.e., when the covariance matrix is completely unknown, we find that we must, in effect, minimize the function

$$\det A_{ij} = \det \sum_{\mu=1}^{n} (f_{\mu i} - y_{\mu i})(f_{\mu j} - y_{\mu j})$$

and, from (64),

$$A = D^{-1}BD^{-1} = D^{-1}U\Lambda U^{T}D^{-1} \tag{66}$$

Inverting (66), and since $U^{-1} = U^{T}$ (U being orthogonal):

$$A^{-1} = DU\Lambda^{-1}U^{T}D = V\Lambda^{-1}V^{T} \tag{67}$$

where $V = DU$. The columns of V are not themselves eigenvectors, but they represent the eigenvectors of the scaled matrix B in the unscaled coordinate system of A . We refer to the columns of V as the <u>principal components</u> of A . These are usually not orthogonal.

If in Λ^{-1} the terms λ_i^{-1} are replaced by $|\lambda_i|^{-1}$, equation (67) becomes equivalent to (32), but will usually be more accurate computationally.

Since $\det U = 1$, we have from (66):

$$\det A = \det^{-2}D \det \Lambda = a_{11}a_{22}\cdots a_{nn}\lambda_1\lambda_2\cdots\lambda_n \tag{68}$$

Note that when A is a covariance matrix, B is the corresponding correlation matrix.

The scaling described above is carried out automatically in the program. Therefore the output always contains the eigenvalues of the scaled matrix and the principal components, rather than the eigenvalues and vectors of the original matrix.

Scaling has been found to reduce the required number of iterations in many problems.

Suppose A is the covariance matrix of the parameter estimates. Let B be a matrix transforming the vector θ ; i.e., let

$$\psi = B(\theta - \bar{\theta}) \tag{69}$$

where $\bar{\theta}$ is the expected value of θ . Then

$$E(\psi\psi^{T}) = BE(\theta - \bar{\theta})(\theta - \bar{\theta})^{T}B^{T} = BAB^{T} = BD^{-1}U\Lambda U^{T}D^{-1}B^{T}$$

If we select $B = (D^{-1}U)^{-1} = U^{T}D = V^{T}$, then $E(\psi\psi^{T}) = \Lambda$. Thus the principal components (i.e., the rows of V^{T}) are the coefficients of ℓ statistically independent linear combinations of the θ's whose variances are the eignevalues of the correlation matrix.

The general case that can be treated is the following:

Let V be a known or unknown positive definite, symmetric matrix, and let A_μ $(\mu = 1,2,\ldots n)$ be a set of non-singular square matrices. Then the covariance matrix of the residuals for the μ-th experiment must have the form (in matrix notation):

$$E(u_\mu^* u_\mu^T) = A_\mu V A_\mu \tag{75}$$

where A_μ^T is the transpose of A_μ.

Defining a new vector of residuals by

$$u_\mu^* = A_\mu^{-1} u_\mu \tag{76}$$

we find that:

$$E(u^* u_\mu^{*T}) = V \tag{77}$$

Thus, V is the covariance matrix of the u_μ^*, and is the same for all experiments. The u_μ^* thus form a suitable set of residuals for use in our parameter estimation program. Equation (76) may be rewritten as:

$$u_\mu^* = A_\mu^{-1}(f(a_\mu,\theta) - y_\mu) = A_\mu^{-1} f(a_\mu,\theta) - A_\mu^{-1} y_\mu \quad . \tag{78}$$

We define

$$\begin{aligned} f_\mu^*(A_\mu^{-1}, a_\mu, \theta) &= A_\mu^{-1} f(a_\mu,\theta) \\ y_\mu^* &= A_\mu^{-1} y_\mu \end{aligned} \tag{79}$$

Thus:

$$u_\mu^* = f_\mu^*(A_\mu^{-1}, a_\mu, \theta) - y_\mu^* \tag{80}$$

Thus, we must replace the observed variables y_μ by y_μ^*, and the original model equations $f(a_\mu,\theta)$ by $f_\mu^*(A_\mu^{-1}, a_\mu, \theta)$. The elements of A_μ^{-1} are simply added to the list of independent variables a_μ.

We illustrate by means of a few examples:

(1) Proportional Errors The standard deviation of each measurement is proportional to the magnitude of the measurement, and all measurements are independent. In this case

$$A_{\mu ij} = \delta_{ij} |y_{\mu i}|$$

$V = \sigma^2 I$, where σ^2 is a known or unknown constant, and I is the identity matrix. Hence,

$$y_{\mu i}^* = \frac{y_{\mu i}}{|y_{\mu i}|}$$

and

$$f_i^* = \frac{1}{|y_{\mu i}|} f_i(a_\mu,\theta)$$

The "observed" values $y_{\mu i}^*$ are all $+1$ or -1 depending on whether $y_{\mu i}$ is positive or negative.

(2) Weighted (by experiment) Least Squares In a least squares problem (single observed variable), the standard deviation of the μ-th measurement is known to be χ_μ. In this case

$$J_{\mu ij} = \frac{\partial g_i(y_\mu, a_\mu, \theta)}{\partial y_{\mu j}} \tag{83}$$

Because of the need to code first, and (with the Gauss-Newton method) second derivatives of the prior distribution, the use of this method will probably be impractical except in the simplest cases. At any rate, we have:

$$\frac{\partial \log \det J_\mu}{\partial \theta_\alpha} = \sum_{ij} J^{-1}_{\mu ij} \frac{\partial J_{\mu ji}}{\partial \theta_\alpha} \tag{84}$$

and:

$$\frac{\partial^2 \log \det J_\mu}{\partial \theta_\alpha \partial \theta_\beta} = \sum_{ij} J^{-1}_{\mu ij} \frac{\partial^2 J_{\mu ji}}{\partial \theta_\alpha \partial \theta_\beta} - \sum_{ijk\ell} J^{-1}_{\mu ik} J^{-1}_{\mu \ell j} \frac{\partial J_{\mu ji}}{\partial \theta_\alpha} \frac{\partial J_{\mu k\ell}}{\partial \theta_\beta} \tag{85}$$

5. SOLUTION OF SIMULTANEOUS NONLINEAR EQUATIONS

Solving the set of ℓ simultaneous equations:

$$g_\mu(\theta) = 0 \qquad (\mu = 1,2,\ldots\ell) \tag{86}$$

for the ℓ unknowns $\theta_1, \theta_2, \ldots \theta_\ell$, is equivalent to minimizing

$$G(\theta) = \sum_{\mu=1}^{\ell} g_\mu^2(\theta) \tag{87}$$

Thus, the simultaneous equation problem is equivalent to a least squares problem which is solvable by our program.

6. TWO POINT BOUNDARY VALUE PROBLEMS

Let the state variables x in a dynamic system obey the differential equations

$$\frac{dx_i}{dt} = h_i(x,t) \qquad (i = 1,2,\ldots s) \tag{88}$$

Suppose the initial values of some of the x's are given:

$$x_{i_1}(0) = \xi_1 \;;\; x_{i_2}(0) = \xi_2 \;;\; \ldots\; x_{i_p}(0) = \xi_p \tag{89}$$

where the subscripts $(i_1, i_2, \ldots i_p)$ belong to the set $(1,2,\ldots s)$. Also, let the final values of some other set of state variables be given at time $t = T$, i.e.:

$$x_{j_1}(T) = \eta_1 \;;\; x_{j_2}(T) = \eta_2 \;;\; \ldots\; x_{j_q}(T) = \eta_q \tag{90}$$

where the subscripts $(j_1, j_2, \ldots j_q)$ again belong to the set $(1,2,\ldots s)$. The sets $(i_1 \ldots i_p)$ and $(j_1 \ldots j_q)$ may overlap, and need not exhaust the whole set of state variables. It is generally required that $p + q = s$. The problem of finding the missing initial conditions is called the <u>two point boundary value problem</u> (TPBVP).

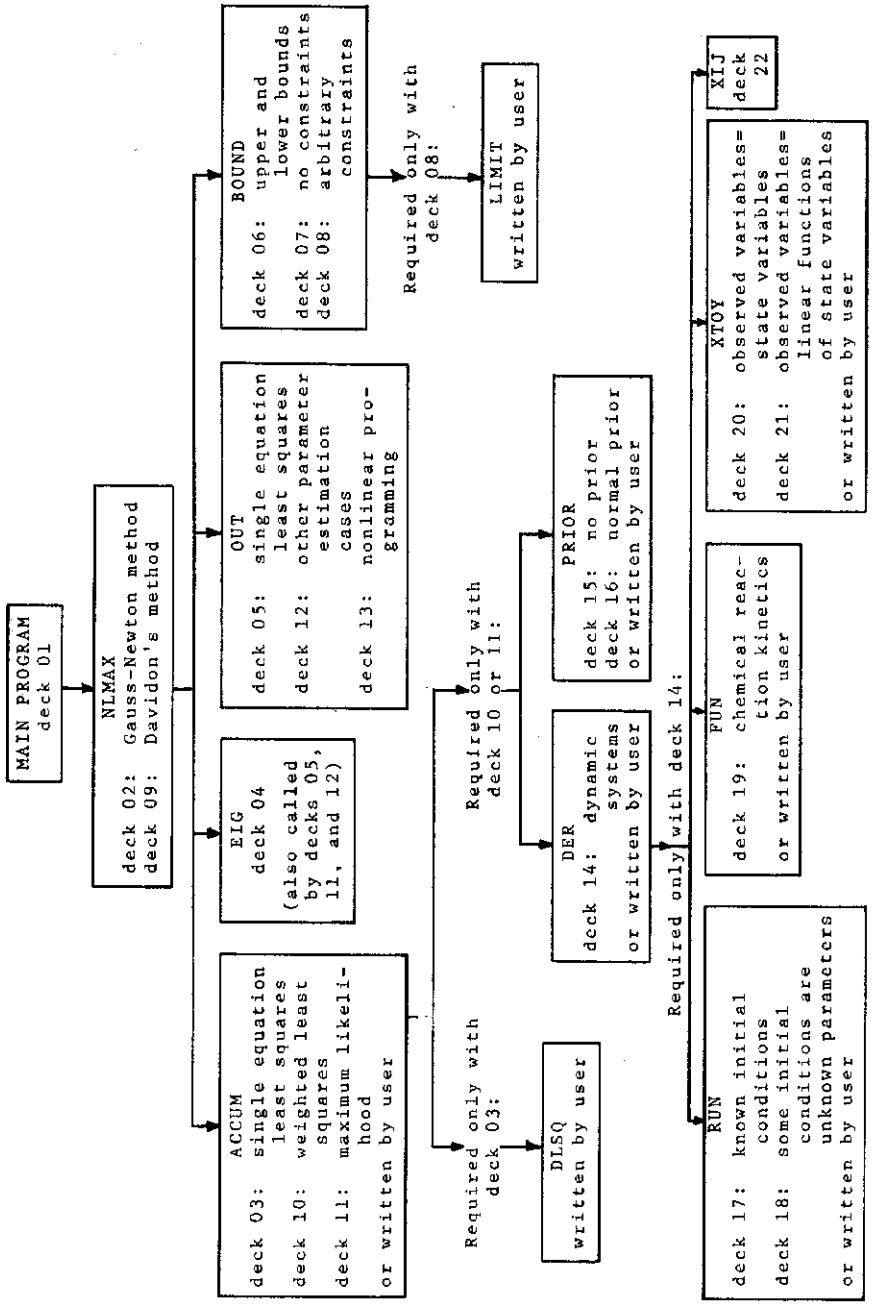
For example, with two state variables, let

$$\frac{dx_1}{dt} = h_1(x_1, x_2, t)$$

$$\frac{dx_2}{dt} = h_2(x_1, x_2, t)$$

and let $x_1(0) = \xi_1$ and $x_1(T) = \eta_1$ be given. The problem is to determine $x_2(0)$.

Figure 1: PROGRAM STRUCTURE



Use deck 10 for multiple equation weighted least squares, and maximum likelihood estimation with diagonal covariance matrix (i.e. weighted least squares with unknown weights). Also use for single equation least squares if the equation involves numerical solution of a differential equation (including kinetics problems), or if a prior distribution is specified. Calls on DER, PRIOR.

Use deck 11 for multiple equation maximum likelihood estimation with non-diagonal (known or unknown) covariance matrix. Calls on EIG, DER, PRIOR.

Write own subroutine for arbitrary nonlinear programming problems.

EIG (deck 04): scales a symmetric matrix, computes eigenvalues and/or eigenvectors, and then rescales the vectors. Uses Threshold-Jacoby method.

OUT: provides additional detailed output after solution has been found.

Use deck 05 in conjunction with subroutine ACCUM deck 03 for least squares problem. Calls on EIG.

Use deck 12 in conjunction with subroutine ACCUM deck 10 or 11 for parameter estimation problems. Calls on EIG.

Use deck 13 otherwise.

BOUND: Computes the constraint penalty functions and their derivatives.

Use deck 06 for specifying lower and upper bounds on each parameter.

Use deck 07 for no constraints.

Use deck 08 for arbitrary constraints. Calls on LIMIT.

The following subroutine is required only if subroutine BOUND deck 08 (arbitrary constraints) is used.

LIMIT: computes constraint functions and their first derivatives. Must be written by user.

B. LIST OF SYMBOLS - COMMON STORAGE

Data is shared among the subroutines through COMMON storage. This is made up of three segments A, B, and C. All subroutines require segment A; all subroutines relating specifically to parameter estimation also require segment B; and all subroutines relating specifically to dynamic systems also require segment C.

In the following list of COMMON variables, the dimensions of the arrays are given in terms of the minimum required capacity, using the following symbols:

- a = number of chemical reactions
- b = number of heterogeneous reaction denominators
- c = number of unknown kinetics parameters
- d = number of unknown initial conditions
- k = number of observed variables
- ℓ = number of unknown parameters
- m = number of independent variables (exclusive of time in a dynamic system)
- n = number of experiments
- p = number of run variables
- q = number of unknown parameters in differential equations and initial conditions*
- r = number of runs
- s = number of state variables
- t = number of constraints
- u = max(k,ℓ)
- v = max(ℓ,t/2) (i.e. 2v must be at least t)

As distributed, the program can be run for problems satisfying:

$a \leq 10$

$b \leq 5$

*In a dynamic system, $\ell \geq q$. If $\ell > q$, then $\theta_{q+1}, \theta_{q+2}, \ldots \theta_{\ell}$ appear only in the equations relating the observed variables to the state variables.

$c \leq 20$

$d \leq 20$

$k \leq 5$

$\ell \leq 20$

$k + m \leq 10$

$n \leq 100$ (200 in the least squares package)

$p + s \leq 10$

$q \leq 20$

$r \leq 20$

$s \leq 10$

$t \leq 40$

The COMMON and DIMENSION statements may be changed easily to accommodate problems with different configurations. For the required changes in DIMENSION statements, see page 99 The required changes in the COMMON statements will be obvious from the following list:

Segment A:

C(u,u), G1(u,u)	for internal use
PSCA	
G(ℓ,ℓ)	$\partial^2 G/\partial\theta_i\partial\theta_j$ in Gauss-Newton; H_{ij} in Davidon
F(ℓ)	$\partial G/\partial\theta_i$
Y(ℓ), EGV(ℓ), FF(ℓ)	for internal use
TITLE(20)	80 characters of BCD identification
CUB(v)	$\partial z/\partial\theta_i$
CLB(v), PLN(ℓ)	for internal use
NCON	t (number of constraints)
LOUT	= 1 if iteration by iteration; 2 if final only; output is desired

C. USER WRITTEN SUBROUTINES

Most variables used or computed by the subroutines are located in COMMON storage, whose structure was described in the preceding section. All variables referred to, which do not appear in the calling sequences, are COMMON variables.

The chapter "PROGRAM STRUCTURE" will guide the user in deciding which of the following subroutines he must code for his specific problem.

The variable II appears in the calling sequence of most subroutines.

II = 1	means that the required function is to be computed
II = 2	means that the function and its first, and, where stated, second, derivatives are to be computed
II = 3	upon the first entry to the subroutine, no calculations are required, but the routine may read input and perform initializations at this time

In all subroutines, if the required calculations cannot be performed, e.g. because some variables exceed the allowed range, the COMMON variable LS should be set to equal 2.

Sometimes one may wish to estimate parameters for several proposed models. Rather than write separate subroutines for each model, the user may combine all models into a single set of subroutines. At running time the number input into location MD will instruct the subroutines which particular model should be used in the current run.

In the following examples, the statement (COMMON SEGMENT A) should be interpreted as:

```
      COMMON  C(20,20),G1(20,20),PSCA,G(20,20),F(20),Y(20),
              EGV(20),FF(20),TITLE(20),CUB(20),CLB(20),
              PNL(20),NCON,LOUT,F3,NTH,F6,F7,METH,NPH,MD,
              LS,C1(20)
```


The statement (COMMON SEGMENT B) as:

```
      COMMON  V(5,5),QY(5),YTH(5,20),A(100,10),ICOV,DET,
              IDER,M,NY,NA
```

The statement (COMMON SEGMENT C) as:

```
      COMMON  NX,NB,NTH2,NTH1,Q(10),P(10),FX(10,10),FTH(10,20),
              BV(20,10),TIME(100),IA(100),FN(10),XTH(10,20),
              XTTH(10,20)
```

(1) SUBROUTINE ACCUM(II)

Requires COMMON segment A.

Purpose: To compute the objective function to be maximized $G(\theta_1,\theta_2,\ldots)$ and/or its first and second derivatives. The value of θ_I is found in location C1(I) .

If II = 1 : place G in location F3.

If II = 2 : place G in location F3.
place $\frac{\partial G}{\partial \theta_I}$ in location F(I), for I = 1,2,...NTH .

The following is required only when using Gauss-Newton method (METH = 1) :

place $\frac{\partial^2 G}{\partial \theta_I \partial \theta_J}$ in location G(I,J) for
I = 1,2,...NTH
J = I,I+1,...NTH

Example: Suppose we wish to maximize

$$G = \theta_1 + 2\theta_1\theta_2 - \theta_2^2$$

The appropriate subroutine would be:

```
      SUBROUTINE ACCUM(II)
      (COMMON SEGMENT A)
      GO TO(1,1,3),II
    1 F3=C1(1)+C1(2)*(2.*C1(1)-C1(2))
      GO TO(3,4),II
    4 F(1)=1.+2.*C1(2)
      F(2)=2.*(C1(1)-C1(2))
```

$$\frac{\partial f_1}{\partial \theta_1} = a_{\mu 1} \quad ; \quad \frac{\partial f_1}{\partial \theta_2} = a_{\mu 2} \quad ; \quad \frac{\partial f_2}{\partial \theta_1} = - a_{\mu 3} e^{-\theta_1 a_{\mu 3}} \quad ;$$

$$\frac{\partial f_2}{\partial \theta_2} = 0 \quad .$$

The observed values of $y_{\mu 1}$ and $y_{\mu 2}$ for $\mu = I$ are stored in A(I,1) and A(I,2) respectively, while $a_{\mu 1}$, $a_{\mu 2}$, and $a_{\mu 3}$ are in A(I,3), A(I,4), and A(I,5) respectively. The subroutine would be:

```
      SUBROUTINE DER(II,I)
      (COMMON SEGMENT A)
      (COMMON SEGMENT B)
      GO TO(1,1,2),II
1     QY(1)=C1(1)*A(I,3)+C1(2)*A(I,4)
      A1=-C1(1)*A(I,5)
      IF(A1-30.)3,3,4
3     QY(2)=EXP(A1)
      GO TO(5,6),II
4     LS=2
      RETURN
6     YTH(1,1)=A(I,3)
      YTH(1,2)=A(I,4)
      YTH(2,1)=-A(I,5)*QY(2)
5     RETURN
2     YTH(2,2)=0.
      RETURN
      END
```

(4) SUBROUTINE PRIOR(II)

Requires COMMON segment A.

Purpose: To compute the logarithm of the prior distribution of the unknown parameters $\log p_o(\theta_1, \theta_2, \ldots)$, and/or its first and second derivatives. The value of θ_I is found in location C1(I).

If II = 1 : add the value of $\log p_o$ to the contents of location F3

If II = 2 : add the value of $\log p_o$ to the contents of location F3

add the value of $\frac{\partial \log p_o}{\partial \theta_I}$ to the contents of location F(I), for I = 1,2,...NTH

The following is required only when using the Gauss-Newton method (METH = 1) :

add the value of $\frac{\partial^2 \log p_o}{\partial \theta_I \partial \theta_J}$ to the contents of location G(I,J) for I = 1,2,...NTH, J = I,I+1,...NTH

Example: Suppose θ_1 and θ_2 are assumed independent, and obeying an exponential and normal distribution, respectively. Thus:

$$p_o = \alpha e^{-\beta\theta_1 - \frac{(\theta_2-\gamma)^2}{2\sigma^2}}$$

with α, β, γ, and σ known constants. Thus,

$$\ell n\, p_o = \ell n\, \alpha - \beta\theta_1 - \frac{(\theta_2-\gamma)^2}{2\sigma^2}$$

The subroutine would be:

```
      SUBROUTINE PRIOR(II)
      (COMMON SEGMENT A)
      GO TO(1,1,2),II
2     READ(5,3)BETA,GAMMA,SIGMA
      A2=1./SIGMA**2
3     FORMAT(3E10.5)
      RETURN
1     A1=(C1(2)-GAMMA)/SIGMA
      F3=F3-BETA*C1(1)-.5*A1**2
      GO TO(4,5),II
```

and for all during initialization (II = 3) . If any of these quantities are constant for all experiments constituting a given run, the corresponding locations may be initialized at the first entrance to the subroutine for the given run: this is characterized by JJ = 2 (on all other occassions JJ = 1). The run number is given by IB.

Example: Suppose the state variables satisfy the following differential equations:

$$\frac{dx_1}{dt} = \theta_1 x_2^{\,2} + \theta_2 t$$

$$\frac{dx_2}{dt} = \theta_2 x_1 x_2 + \theta_3$$

The subroutine would be:

```
      SUBROUTINE FUN(II,JJ,IB,T)
      (COMMON SEGMENT A)
      (COMMON SEGMENT B)
      (COMMON SEGMENT C)
      GO TO(1,1,2),II
2     FX(1,1)=0.
      FTH(1,3)=0.
      FTH(2,1)=0.
      FTH(2,3)=1.
      RETURN
1     P(1)=C1(1)*Q(2)**2+C1(2)*T
      P(2)=C1(2)*Q(1)*Q(2)+C1(3)
      GO TO(3,4),II
4     FX(1,2)=2.*C1(1)*Q(2)
      FX(2,1)=C1(2)*Q(2)
      FX(2,2)=C1(2)*Q(1)
      FTH(1,1)=Q(2)**2
      FTH(1,2)=T
      FTH(2,2)=Q(1)*Q(2)
3     RETURN
      END
```


(7) SUBROUTINE XTOY(II,IB,I)

Requires COMMON segments A, B, C.

Purpose: To compute the predicted values of the observed variables for the I-th experiment in a dynamic system, from the computed values of the state variables: $y_{IJ} = d_J(x_{I1}, x_{I2}, \ldots, \theta_1, \theta_2, \ldots)$, and/or their first derivatives. The values of x_{IJ} will be found in location Q(J) , and that of θ_J in location C1(J) . IB denotes the run number.

If II = 1 : place y_{IJ} in location QY(J) for J = 1,2,...NY

If II = 2 : place y_{IJ} in location QY(J) for J = 1,2,...NY

place $\frac{\partial y_{IJ}}{\partial \theta_K}$ in location YTH(J,K) for J = 1,2,...NY, K = 1,2,...NTH

To compute $\frac{\partial y_{IJ}}{\partial \theta_K}$, remember that

$$\frac{\partial y_{IJ}}{\partial \theta_K} = \sum_{L=1}^{NX} \frac{\partial d_J}{\partial x_L} \frac{\partial x_L}{\partial \theta_K} + \frac{\partial d_J}{\partial \theta_K} \qquad \text{for } K \leq NTH2$$

$$\frac{\partial y_{IJ}}{\partial \theta_K} = \frac{\partial d_J}{\partial \theta_K} \qquad \text{for } NTH1 \leq K \leq NTH \text{ , with } NTH1 = NTH2 + 1$$

The value of $\frac{\partial x_L}{\partial \theta_K}$ will be found in location XTH(L,K) for L = 1,2,...NX ; K = 1,2,...NTH2 .

Example: Let there be three state variables x_1 , x_2 , x_3 which depend on the parameters θ_1 , θ_2 . Only one variable y_1 is observed, and it is thought to be a linear combination of unknown powers of the state variables, with coefficients that vary from run to run:

in one run by several alternative models, it is necessary to reproduce the data deck several times, merely changing the value of MD .

(1) NLMAX, deck 02 or 09

Title card:

Any 80 columns of BCD information, to appear as title for the output:

Format: 20A4

Problem definition card:

Containing three integers:

NTH = number of unknown parameters

LOUT = 1 if iteration by iteration output is desired

2 if only final output is desired

MD = model number

Format: 3I5

Parameter initial guess card(s):

Each containing up to eight fixed or floating point numbers:

C1(I) , I = 1,2,...NTH : initial guesses for the unknown parameters

Format: 8E10.5

(2) ACCUM

(a) deck 03, least squares problems. See page , items (4) and (5).

(b) deck 10, diagonal covariance matrix (weighted least squares):

Problem size card:

Containing four integers:

NY = number of observed variables

M = number of experiments

NA = total number of variables (observed and independent. Do not count time as an independent variable in a dynamic system).

ICOV = 1 if covariance matrix is known

2 if covariance matrix is known up to a multiplicative constant. Use for weighted least squares with known weights.

3 if covariance matrix is unknown. Use for weighted least squares with unknown weights.

Format: 4I5

Data cards:

Each containing up to eight fixed or floating point numbers:

A(I,J) , J = 1,2,...NA : for the I-th experiment, the values of the NY observed variables, followed by the values of the independent variables.

Repeat for I = 1,2,...M . Start a new card for each experiment.

Format: 8E10.5

Covariance card(s):

Each containing up to eight fixed or floating point numbers. Omit if ICOV = 3 .

C(I,I) , I = 1,2,...NY : the diagonal elements of the covariance matrix, i.e. the reciprocals of the weights to be given each observed variable in weighted least squares.

Format: 8E10.5

(c) deck 11, non-diagonal covariance matrix.

Identical to input required by deck 10 above, except for the last item, which becomes:

Covariance cards:

Each containing up to eight fixed or floating point numbers. Omit if ICOV = 3 .

1st card: C(1,1) : the first element of the first row of the covariance matrix.

2nd card: C(2,1),C(2,2) : the first two elements of the second row.

(4) RUN

(a) deck 17, all initial conditions known: no input required.

(b) deck 18, some initial conditions are unknown parameters:

First card, integer:

NTH3 = number of unknown initial conditions (may be = 0)

Format: I5

Subsequent card(s):

Each containing three integers:

IR(I,1) = unknown parameter number
IR(I,2) = run number
IR(I,3) = state variable number

} in I-th unknown initial condition

Repeat for I = 1,2,...NTH3 . Start a new card for each condition.

Format: 3I5

NOTE: A card punched 7 1 3 means that the seventh unknown parameter represents the initial value of the third state variable in the first run.

(5) FUN, deck 19 (chemical kinetics problems)

System definition card:

Containing four integers:

NTH3 = number of unknown kinetic parameters.
NOTE: There may be additional unknown parameters involving initial conditions and the observation function. The parameters must be ordered as follows:
1) kinetic parameters
2) initial condition parameters
3) observation parameters.
NTH3 may equal zero.

NRE = number of reactions.

NDEN = number of distinct denominators (for heterogeneous reactions).

ITEMP = 1 if temperature is not a variable.
2 if temperature is a variable.

NOTE: If ITEMP = 2 , the program expects to find the temperature for the I-th run in location BV(I,NX+1) . I.e., the temperature must be read in as the first run variable.

Format: 4I5

The general form of the rate equation is:

$$\frac{dx_j}{\partial t} = \sum_{i=1}^{NRE} A_{ji} R_i \qquad (j = 1,2,\ldots NX)$$

where

$$R_i = \frac{k_i e^{-e_i/T} (F_i - K_i e^{-E_i/T} B_i)}{G_{n_i}}$$

F_i = 1, or x_{i_1} , or $x_{i_1} x_{i_2}$, or $x_{i_1} x_{i_2} x_{i_3}$ = forward reaction components

B_i = 1, or x_{i_4} , or $x_{i_4} x_{i_5}$, or $x_{i_4} x_{i_5} x_{i_6}$ = reverse reaction components

G_n = 1, or $(1 + \sum_{\ell=1}^{NX} M_{n\ell} e^{-D_{n\ell}/T} x_\ell^{q_{n\ell}})^{P_n}$ = denominator

If ITEMP = 1 , all the e_i , E_i , and $D_{n\ell}$ are assumed equal to zero.

If NDEN = 0 , all G_m are assumed equal to 1.

An unknown kinetics parameter may be a k_i , e_i , K_i , E_i , $M_{n\ell}$, or $D_{n\ell}$.

The numbers $i_1, i_2, \ldots i_6$ are subscripts whose values (not necessarily distinct) are numbers taken from the set 1,2,...NX . For instance, if $F_i = x_2^2 x_4$, then $i_1 = 2$, $i_2 = 2$, $i_3 = 4$.

The number n_i is a subscript whose value is taken from the set 0,1,2,...NDEN .

Format: 8E10.5

(c) Component coefficient card(s), each containing up to eight fixed or floating point numbers:

DECOF(N,L) , L = 1,2,...NX : $M_{n\ell}$, the coefficient of the L-th component in N-th denominator.

Format: 8E10.5

(d) Activation energies card(s), each containing up to eight fixed or floating point numbers. Omit if ITEMP = 1 .

DEX(N,L) , L = 1,2,...NX : $D_{n\ell}$, the activation energy for $M_{n\ell}$.

Format: 8E10.5

(6) XTOY

(a) deck 20, observed variables identical to state variables number 1,2,...NY : no input required.

(b) deck 21, observed variables $y_{\mu i}$ are linear functions of the state variables $x_{\mu j}$, i.e.:

$$y_{\mu i} = \sum_{j=1}^{NX} B_{ij} x_{\mu j} + B_{i,NX+1} \qquad (i = 1,2,...NY)$$

Option card:

Containing one integer:

IMAT = 1 if $B_{ij} = \delta_{ij}$ for $i,j \leq NY$, and $B_{j,NX+1} = 0$. (In this case the program functions precisely as though deck 20 were used.)

= 2 if all B_{ij} are known.

= 3 if some B_{ij} are among the unknown parameters. (These unknown parameters must be the last NTH2 among the NTH unknown parameters.)

Format: I5

Parameter definition card(s):

Each containing up to 16 integers. Use only if IMAT = 3 .

IBTH(1,1) = i
IBTH(1,2) = j
} for first unknown B_{ij}

IBTH(2,1) = i
IBTH(2,2) = j
} for second unknown B_{ij}

etc., for all unknown B_{ij} .

Format: 16I5

NOTE: Suppose NTH = 8 , NTH2 = 3 (at input time), and the parameter definition card reads:

1 3 4 7 4 8

This means that $\theta_6 = B_{1,3}$; $\theta_7 = B_{4,7}$; $\theta_8 = B_{4,8}$.

Observation matrix card(s):

Each containing up to eight fixed or floating point numbers. Use only if IMAT = 2 or 3 .

YX(I,J) , J = 1,2,...,NX + 1 : $B_{I,J}$

Repeat for I = 1,2,...NY . Start a new card for each I .

Format: 8E10.5

(7) PRIOR

(a) deck 15, uniform prior distribution: no input required.

(b) deck 16, normal prior distribution:

Expected values card(s):

Each containing up to eight fixed or floating point numbers:

TAV(I) , I = 1,2,...,NTH : the expected value of the I-th unknown parameter.

Format: 8E10.5

If initial guess θ_I is	Assigned upper bound is
-	0
0	10^{10}
+	100X initial guess

(b) deck 07, no constraints: no input required.

(c) deck 08, arbitrary constraints:

Number of constraints card:

Containing one integer:

NCON = number of constraints.

Format: I5

E. OUTPUT

The output of a run depends on which routines are used. Listed below is the output produced by each routine. Unlisted routines produce no output.

Output designated as initial, intermediate, and final occurs respectively before, during, and after the search for the maximum.

Error messages indicating conditions which prevent further execution of the current problem appear underlined with asterisks in the output. Any further output produced by the problem may be meaningless. Premature termination of one problem does not prevent proper execution of the next one in a batch.

NLMAX, deck 02

Initial output:

Title and model number. Initial guesses of parameter values (when using BOUND deck 08 these may differ slightly from the input values, since an interior point is substituted for one on a constraint).

Intermediate output:

(Omitted when LOUT = 2). Messages indicating that penalty functions are included, reduced, or omitted (these appear only when bounds or constraints are imposed on the parameters). For each iteration (corresponding to a new evaluation of the derivatives) and function evaluation: the value of the objective function and the current values of the parameters. The objective function is the quantity L (see Table 1, page 34 when using ACCUM deck 10, or Table 2, page 35 when using ACCUM deck 11), plus penalty function and log of prior distribution where applicable.

Final output:

Maximum value of objective function, and the corresponding

than ℓ (= number of parameters) iterations after a re-initialization of the matrix. The parameter estimates, however, should still be correct, but not the covariance matrix of the estimates.

Error messages:

RESTART and FEASIBLE PARAMETER VALUES COULD NOT BE FOUND are the same as for deck 02.

ACCUM, deck 10

Initial output:

The diagonal elements of the covariance matrix, when specified; the observations (i.e. values of the observed and independent variables for each experiment) when not using DER deck 14.

Error message:

SINGULAR DIAGONAL MOMENT MATRIX

When the covariance matrix is unknown, this indicates that one of the model equations was fitted exactly. Control is turned over to the final output routine.

ACCUM, deck 11

Initial output:

The covariance matrix, when specified; the observations (i.e. the values of the observed and independent variables for each experiment) when not using DER deck 14.

Error messages:

SINGULAR MOMENT MATRIX

When the covariance matrix is not specified, this indicates that the moment matrix of the residuals has become singular. Control is turned over to the final output routine.

GIVEN COVARIANCE NOT POSITIVE DEFINITE

Self-explanatory. Execution of current problem is terminated.

OUT, deck 12

Final output:

Title and model number; the residuals (computed minus observed values of the observed variables) for each experiment; unbiased covariance matrix and standard deviations of the residuals; log $p(\theta/y)$ (designated "LOG PROBABILITY"); log $p(y/\theta)$ ("LOG LIKELIHOOD", the quantity $G_{M.L.}$ in Table 1 or 2, page 34-35); log $p_o(\theta)$ ("LOG PRIOR", normalizing constants are ignored. Note that log $p(\theta/y)$ = log $p(y/\theta)$ + log $p_o(\theta)$); the determinant of the moment matrix of residuals (= product of diagonal elements when ACCUM deck 10 is used) if covariance matrix is unknown; the weighted sum of squares of the residuals if the covariance matrix is known or known up to a multiplicative constant; unbiased estimate of the multiplier if covariance matrix is known up to a multiplicative constant (this estimate is also provided when the covariance matrix is known; in this case, a value much different from 1 would cast doubt on the claimed knowledge); the estimated values of the parameters and their standard deviations; the covariance matrix of the parameter estimates; the eigenvalues of the correlation matrix and the principal components, with their expected values and standard deviations.

DER, deck 14

Initial output:

The initial conditions and run variables for each run, interspersed with the time, observed, and independent variables for each experiment belonging to the run.

PRIOR, deck 16

Initial output:

The prescribed means and standard deviations or covariance matrix of the prior distribution.

Segment B:

```
COMMON V(k,k),QY(k),YTH(k,ℓ),A(n,k+m),ICOV,
       DET,IDER,M,NY,NA
```

Alternate Segment B:

```
COMMON X,XTH(ℓ),A(n,m),M,NA
```

Segment C:

```
COMMON NX,NB,NTH2,NTH1,Q(s),P(s),FX(s,s),FTH(s,q),
       BV(r,s+p),TIME(n),IA(n),FN(s),XTH(s,ℓ),
       XTTH(s,ℓ)
```

The COMMON and DIMENSION requirements of each deck are listed below:

Main program, deck 01:

None.

NLMAX, deck 02:

```
COMMON Segment A
DIMENSION W(ℓ)
```

ACCUM, deck 03:

```
COMMON Segments A and Alternate B
```

EIG, deck 04:

```
COMMON A(ℓ,ℓ),V(ℓ,ℓ),PSCA
DIMENSION SCA(ℓ)
```

OUT, deck 05:

```
COMMON Segments A and Alternate B
```

BOUND, deck 06:

```
COMMON Segment A
```

BOUND, deck 07:

None.

BOUND, deck 08:

```
COMMON Segment A
```


NLMAX, deck 09:

```
COMMON Segment A
```

ACCUM, deck 10:

```
COMMON Segments A and B
DIMENSION UI(k,ℓ),QQ(k)
DIMENSION ZI(k,ℓ,ℓ)
```

ACCUM deck 11:

```
COMMON Segments A and B
DIMENSION UIK(k,k,ℓ),QQ(k)
DIMENSION VIK(k,k,ℓ(ℓ+1)/2),WIK(k,k,ℓ)
```

OUT, deck 12:

```
COMMON Segments A and B
```

OUT, deck 13:

None.

DER, deck 14:

```
COMMON Segments A, B, and C
DIMENSION QO(s),P1(s),XTTH1(s,q),XTH1(s,q),
          Q1(s)
```

PRIOR, deck 15:

None.

PRIOR, deck 16:

```
COMMON Segment A
DIMENSION TCOV(ℓ,ℓ),TAV(ℓ),TER(ℓ)
```

RUN, deck 17:

None.

RUN, deck 18:

```
COMMON Segments A, B, and C
DIMENSION IR(d,3)
```

$$\theta_1 = k_1$$

$$\theta_2 = e_1$$

$$\theta_3 = M_{11}$$

$$\theta_4 = D_{11}$$

Three isothermal runs are made, each at a different temperature and with different initial concentrations. These initial concentrations are known approximately, as indicated below:

	$x_1(0)$	$x_2(0)$
Run 1	$1 \pm .05$	$1 \pm .05$
Run 2	$1 \pm .05$	0
Run 3	0	$1 \pm .05$

Thus we have four unknown initial condition parameters:

$$\theta_5 = x_1(0) \quad \text{in Run 1}$$

$$\theta_6 = x_2(0) \quad \text{in Run 1}$$

$$\theta_7 = x_1(0) \quad \text{in Run 2}$$

$$\theta_8 = x_2(0) \quad \text{in Run 3}$$

The light absorptivity of the solution is measured at ten different times (including $t = 0$) in the course of each run. The absorptivity y_1 is assumed to be a linear function of the concentrations:

$$y_1 = B_{11}x_1 + B_{12}x_2 + B_{13}$$

With the exception of B_{13} (the absorptivity of pure solvent), the coefficients are known only approximately:

$$B_{11} = 1 \pm .05 \quad ; \quad B_{12} = 2 \pm .05 \quad ; \quad B_{13} = 1$$

Thus we have two unknown observation-matrix parameters:

$$\theta_9 = B_{11}$$

$$\theta_{10} = B_{12}$$

Our approximate knowledge concerning θ_5 to θ_{10} is expressed by specifying a normal prior distribution with the indicated means and standard deviations for these parameters.

Note that, as required, the kinetics parameters appear first, and the observation parameters appear last in the list of unknown parameters.

Since there is only one observed variable (y_1) and no prior distribution is specified, this is a single equation least squares problem. It cannot be solved, however, by means of the least squares package because the latter does not accommodate the integration and kinetics routines. To solve this problem, we use the following routines:

deck 01 - main program
deck 02 - NLMAX, Grauss-Newton method
deck 04 - EIG, required by 02
deck 06 - BOUND, upper and lower bounds
deck 10 - ACCUM, diagonal covariance
deck 12 - OUT, parameter estimation
deck 14 - DER, integrates differential equations
deck 16 - PRIOR, normal prior distribution
deck 18 - RUN, unknown initial conditions
deck 19 - FUN, kinetics reaction rates

card	required by deck	contents
(continued)		
40	"	80., 90. (times for the experiments of first run)
41	"	10 (number of experiments in second run)
42	"	1.*, 0. (initial values of state variables for second run); 400. (temperature for second run)
43	"	0., 2., 4., 6., 8., 10., 12., 14.
44	"	16., 18. (times for the experiments of second run)
45	"	10 (number of experiments in third run)
46	"	0., 1.* (initial values of state variables for third run); 600. (temperature for third run)
47	"	0., .5, 1., 1.5, 2., 2.5, 3., 3.5
48	"	4., 4.5 (times for the experiments of third run)
49	"	10., 10. (upper bounds on $\|x_1\|$, $\|x_2\|$)
50	18(RUN)	4 (number of unknown initial conditions)
51	"	5 (parameter number); 1 (run number); 1 (state variable number); i.e., $\theta_5 = x_1(0)$ in run 1
52	"	6, 1, 2 ($\theta_6 = x_2(0)$ in run 1)
53	"	7, 2, 1 ($\theta_7 = x_1(0)$ in run 2)
54	"	8, 3, 2 ($\theta_8 = x_2(0)$ in run 3)
55	19(FUN)	4 (number of kinetics parameters); 1 (number of reactions); 1 (number of denominators); 2 (temperature varies from run to run)
56	"	1 (θ_1 is a forward rate constant); 1 (reaction number)

card	required by deck	contents
(continued)		
57	19(FUN)	2 (θ_2 is an activation energy); 1 (reaction number)
58	"	5 (θ_3 is a denominator coefficient); 1 (denominator number); 1 (component number)
59	"	6 (θ_4 is a denominator coefficient activation energy); 1 (denominator number); 1 (component number)
60	"	1 (denominator number for reaction 1); 1 (component number in forward reaction); blank; blank; 2, 2 (component numbers in reverse reaction); blank
61	"	-1., 2. (stoichiometric coefficients for reaction 1)
62	"	2.*, 1. (forward rate and inverse equilibrium constant for reaction 1)
63	"	500.*, 1000. (activation energies for the above)
64	"	2. (exponent for denominator 1)
65	"	1., 0. (exponents for the concentrations in denominator 1)
66	"	.5*, 0. (coefficients of the state variables in denominator 1)
67	"	50.*, 0. (activation energies for the above)
68	21(XTOY)	3 (some observation matrix elements are unknown)
69	"	1 (observed variable number); 1 (state variable number for θ_9); 1 (observed variable number); 2 (state variable number for θ_{10})
70	"	1.* (B_{11}); 2.* (B_{12}); 1. (B_{13})

KINETICS TEST PROBLEM

MODEL 0

RUN	EXP.	TIME	OBSERVATIONS		
1		INIT. COND.	0.100000E 01	0.100000E 01	0.200000E 03
	1	0.0	0.398800E 01		
	2	0.100000E 02	0.407300E 01		
	3	0.200000E 02	0.415300E 01		
	4	0.300000E 02	0.423100E 01		
	5	0.400000E 02	0.430900E 01		
	6	0.500000E 02	0.437600E 01		
	7	0.600000E 02	0.445700E 01		
	8	0.700000E 02	0.452200E 01		
	9	0.800000E 02	0.461500E 01		
	10	0.900000E 02	0.466700E 01		
2		INIT. COND.	0.100000E 01	0.0	0.400000E 03
	11	0.0	0.199700E 01		
	12	0.200000E 01	0.214900E 01		
	13	0.400000E 01	0.232000E 01		
	14	0.600000E 01	0.246500E 01		
	15	0.800000E 01	0.261100E 01		
	16	0.100000E 02	0.275400E 01		
	17	0.120000E 02	0.289600E 01		
	18	0.140000E 02	0.303400E 01		
	19	0.160000E 02	0.316600E 01		
	20	0.180000E 02	0.327800E 01		
3		INIT. COND.	0.0	0.100000E 01	0.600000E 03
	21	0.0	0.301200E 01		
	22	0.500000E 00	0.295600E 01		
	23	0.100000E 01	0.292600E 01		
	24	0.150000E 01	0.287700E 01		
	25	0.200000E 01	0.285300E 01		
	26	0.250000E 01	0.282300E 01		
	27	0.300000E 01	0.280000E 01		
	28	0.350000E 01	0.277600E 01		
	29	0.400000E 01	0.276700E 01		
	30	0.450000E 01	0.276000E 01		

UNKNOWN INITIAL CONDITIONS

PARAMETER	RUN	STATE VARIABLE
5	1	1
6	1	2
7	2	1
8	3	2

PARAMETER	LOWER BOUND	UPPER BOUND	PENALTY COEFFICIENT
1	0.0	0.200000E 03	0.200100E-03
2	0.0	0.500000E 05	0.500001E-01
3	0.0	0.500000E 02	0.501000E-04
4	0.0	0.500000E 04	0.500010E-02
5	0.0	0.100000E 03	0.100100E-03
6	0.0	0.100000E 03	0.100100E-03
7	0.0	0.100000E 03	0.100100E-03
8	0.0	0.100000E 03	0.100100E-03
9	0.0	0.100000E 03	0.100100E-03
10	0.0	0.200000E 03	0.200100E-03

PARAMETER INITIAL GUESSES

0.200000E 01 0.500000E 03 0.500000E 00 0.500000E 02 0.100000E 01 0.100000E 01 0.100000E 01
0.100000E 01 0.100000E 01 0.200000E 01

MAXIMUM OF OBJECTIVE FUNCTION 0.1582596E 03

PARAMETERS

0.1393018E 01 0.1140164E 04 0.1822429E 01 0.3669609E 03 0.1005001E 01 0.9988383E 00 0.9858559E 00
0.1018941E 01 0.1010827E 01 0.1975482E 01

FUNCTION EVALUATIONS 165 DERIVATIVE EVALUATIONS 54

PRINCIPAL COMPONENTS

0.928767E-03	0.970220E-01	-0.989227E-04	-0.290807E 00	0.287264E-03	-0.514985E-03	-0.318384E-03
-0.204520E-03	0.328836E-03	0.409790E-03				
-0.102839E-02	0.236736E 00	0.188246E-02	-0.431022E 00	-0.361097E-03	-0.220559E-03	-0.356532E-03
-0.217095E-04	-0.486014E-03	-0.309045E-03				
-0.315479E-03	-0.406386E 00	0.291869E-02	0.436860E-01	0.188292E-03	-0.299851E-03	0.117288E-03
0.130477E-03	0.235297E-03	-0.216545E-03				
0.714255E-03	0.272248E 00	0.291786E-02	0.705684E 00	0.279063E-04	-0.268736E-04	0.217141E-04
-0.195345E-04	-0.230287E-04	0.406646E-04				
-0.659325E-03	0.212519E-01	0.175736E-02	-0.194746E 00	0.610427E-03	0.387803E-03	-0.582691E-03
0.191811E-03	0.154847E-03	0.315441E-03				
0.247170E-03	-0.148919E 00	0.129104E-02	-0.527111E-01	-0.985346E-03	0.170133E-03	-0.275771E-03
-0.180711E-03	0.292128E-03	0.357132E-03				
0.100215E-02	-0.208623E-01	0.181895E-02	-0.444032E 00	0.117298E-03	0.150429E-03	0.691965E-03
-0.688186E-04	-0.606367E-03	0.223376E-03				
-0.178904E-03	0.621455E-01	-0.379301E-03	0.110952E-01	-0.293139E-03	-0.181811E-03	0.192714E-03
0.802916E-03	-0.581698E-04	0.348062E-03				
-0.693904E-03	0.219682E 00	0.752443E-03	-0.251353E 00	-0.428447E-04	0.640434E-04	0.648559E-03
-0.484344E-04	0.932469E-03	-0.818428E-04				
-0.180158E-02	-0.366168E-01	-0.186204E-03	0.263874E 00	0.845823E-04	-0.139252E-03	0.319310E-03
-0.317986E-03	-0.310643E-03	0.496480E-03				

LOG PROBABILITY 0.1156915E 03

LOG LIKELIHOOD 0.1159489E 03

LOG PRIOR -0.2574005E 00

SUM OF SQUARES OF RESIDUALS 0.7953061E-03

PARAMETERS

0.139302E 01	0.114016E 04	0.182243E 01	0.366961E 03	0.100600E 01	0.998838E 00	0.985856E 00
0.101894E 01	0.101083E 01	0.197548E 01				

STANDARD DEVIATIONS OF PARAMETERS

0.212424E 00	0.765678E 02	0.818099E 00	0.198133E 03	0.450762E-01	0.299534E-01	0.260772E-01
0.168838E-01	0.262181E-01	0.322059E-01				

COVARIANCE MATRIX OF PARAMETERS

0.451238E-01	0.154301E 02	0.163838E 00	0.380903E 02	0.181670E-02	-0.257630E-02	0.162178E-04
-0.132008E-02	-0.460465E-04	0.293778E-02				
0.154301E 02	0.586262E 04	0.616384E 02	0.149342E 05	0.928751E 00	-0.754054E 00	0.431560E 00
-0.417762E 00	-0.465158E 00	0.873418E 00				
0.163838E 00	0.616384E 02	0.669286E 00	0.160533E 03	0.409072E-02	-0.551697E-02	0.474369E-02
-0.478879E-02	-0.492901E-02	0.999934E-02				
0.380903E 02	0.149342E 05	0.160533E 03	0.392569E 05	0.139041E 01	-0.111138E 01	0.146178E 01
-0.940347E 00	-0.155484E 01	0.193054E 01				
0.181670E-02	0.928751E 00	0.409072E-02	0.139041E 01	0.203186E-02	-0.788452E-03	0.309089E-03
0.820871E-04	-0.326357E-03	-0.158227E-03				
-0.257630E-02	-0.754054E 00	-0.551697E-02	-0.111138E 01	-0.788452E-03	0.897807E-03	0.341210E-03
0.318851E-03	-0.349764E-03	-0.626083E-03				
0.162178E-04	0.431560E 00	0.474369E-02	0.146178E 01	0.309089E-03	0.341210E-03	0.680009E-03
0.155549E-03	-0.676404E-03	-0.322302E-03				
-0.132008E-02	-0.417762E 00	-0.478879E-02	-0.940347E 00	0.820871E-04	0.318851E-03	0.155549E-03
0.285062E-03	-0.167508E-03	-0.539341E-03				
-0.460465E-04	-0.465158E 00	-0.492901E-02	-0.155484E 01	-0.326357E-03	-0.349764E-03	-0.676404E-03
-0.167508E-03	0.687389E-03	0.347033E-03				
0.293778E-02	0.873418E 00	0.999934E-02	0.193054E 01	-0.158227E-03	-0.626083E-03	-0.322302E-03
-0.539341E-03	0.347033E-03	0.103722E-02				

EXPECTED VALUES

0.615796E 01	-0.320035E 02	0.327072E 02	0.760281E 01	-0.203728E 01	0.163882E 02	0.157372E 02
0.917250E 02	0.317688E 02	-0.859900E 01				

STANDARD DEVIATIONS

0.213854E 01	0.234290E 00	0.147037E-01	0.364011E-02	0.117127E 01	0.513969E-01	0.176718E 01
0.344136E-01	0.100771E 00	0.928785E 00				

B. MAXIMUM LIKELIHOOD TEST PROBLEM WITH MISSING DATA

We consider a three equation model with three observed variables y_1 , y_2 , y_3 , three independent variables a_1 , a_2 , a_3 , and three unknown parameters θ_1 , θ_2 , θ_3 :

$$y_1 = \theta_1 a_1 + \theta_2 a_2 + \theta_3 a_3 = g_1(\theta)$$
$$y_2 = \theta_1 a_2 + \theta_2 a_3 + \theta_3 a_1 = g_2(\theta)$$
$$y_3 = \theta_1 a_3 + \theta_2 a_2 + \theta_3 a_1 = g_3(\theta)$$

Thus:

$$\frac{\partial g_1}{\partial \theta_1} = a_1 \qquad \frac{\partial g_1}{\partial \theta_2} = a_2 \qquad \frac{\partial g_1}{\partial \theta_3} = a_3$$

$$\frac{\partial g_2}{\partial \theta_1} = a_2 \qquad \frac{\partial g_2}{\partial \theta_2} = a_3 \qquad \frac{\partial g_1}{\partial \theta_3} = a_1$$

$$\frac{\partial g_3}{\partial \theta_1} = a_3 \qquad \frac{\partial g_3}{\partial \theta_2} = a_2 \qquad \frac{\partial g_3}{\partial \theta_3} = a_1$$

Twenty experiments yielded values of the y's for different sets of values of the a's. However, the value of y_2 in the third experiment and the value of a_3 in the ninth experiment were missing. We denote these by θ_4 and θ_5 respectively. Now for all experiments but the third, $\partial g_i/\partial\theta_4 = 0$ $(i = 1,2,3)$; at the third experiment, $\partial g_2/\partial\theta_4 = -1$; $\partial g_1/\partial\theta_4 = \partial g_3/\partial\theta_4 = 0$. For all experiments but the ninth, $\partial g_1/\partial\theta_5 = 0$ $(i = 1,2,3)$; at the ninth experiment, $\partial g_1/\partial\theta_5 = \theta_3$; $\partial g_2/\partial\theta_5 = \theta_2$; $\partial g_3/\partial\theta_5 = \theta_1$.

A subroutine DER written especially for this problem is required. This subroutine performs the usual computation of the g_i and $\partial g_i/\partial\theta_\alpha$. In addition, at the start of each iteration (with I in the calling sequence equal to 1), it places the current values of θ_4 and θ_5 (i.e. C1(4) and C1(5)) in the locations reserved for y_2 , third experiment (i.e. A(3,2)), and a_3 , ninth experiment (i.e. A(9,6) ; note that A(I,1), A(I,2), ... A(I,6) contain y_1, y_2, y_3, a_1, a_2, a_3 for the I-th experiment). The subroutine, which appears as deck 23 in the program, reads:

The deck for solving this problem is made up of the following routines:

deck 01 - Main program
deck 02 - NLMAX, Gauss-Newton method
deck 04 - EIG, required by 02
deck 07 - BOUND, no constraints
deck 11 - ACCUM, non-diagonal covariance matrix
deck 12 - OUT, parameter estimation
deck 15 - PRIOR, no prior distribution
deck 23 - DER, written by user for this problem

The input consists of:

card	required by deck	contents
1	02(NLMAX)	LINEAR MAXIMUM LIKELIHOOD
2	"	5 (number of unknown parameters); 1 (full output); blank (no model number)
3	"	0., 0., 0., 0., 0. (initial guesses of parameter values)
4	11(ACCUM)	3 (number of observed variables); 20 (number of experiments); 6 (number of variables per experiment); 3 (unknown covariance matrix)
5	"	4.36, 5.21, 5.35(y_1 , y_2 , y_3); .871, .643, .550 (a_1 , a_2 , a_3 for first experiment)
6	"	4.99, 3.30, 3.10, .228, .669, .854
7	"	1.67, 0. (unknown value), 2.75, .528, .229, .170
8	"	2.17, 1.48, 1.49, .110, .354, .337
9	"	2.98, 4.69, 4.23, .911, .056, .493
10	"	4.46, 3.87, 3.15, .476, .154, .918
11	"	1.79, 3.18, 3.57, .655, .421, .077
12	"	1.71, 3.13, 3.07, .649, .140, .199
13	"	3.07, 5.01, 4.58, .995, .045, 0 (unknown value)
14	"	.94, .93, .74, .130, .016, .195
15	"	4.97, 5.37, 5.35, .823, .690, .690
16	"	4.32, 4.85, 5.46, .768, .992, .389
17	"	2.17, 1.78, 2.43, .203, .740, .120
18	"	2.22, 2.18, 2.44, .302, .519, .221
19	"	2.88, 4.90, 5.11, .991, .450, .249
20	"	2.29, 1.94, 1.46, .224, .030, .502
21	"	3.76, 3.39, 2.71, .428, .127, .772
22	"	1.99, 2.93, 3.31, .552, .494, .110
23	"	4.95, 4.08, 4.19, .461, .824, .714
24	"	2.96, 4.26, 4.48, .799, .494, .295

LINEAR MAXIMUM LIKELIHOOD

MODEL 0

OBSERVATIONS

1	0.436000E 01	0.521000E 01	0.535000E 01	0.871000E 00	0.643000E 00	0.550000E 00
2	0.499000E 01	0.330000E 01	0.310000E 01	0.228000E 00	0.669000E 00	0.854000E 00
3	0.167000E 01	0.0	0.275000E 01	0.528000E 00	0.229000E 00	0.170000E 00
4	0.217000E 01	0.148000E 01	0.149000E 01	0.110000E 00	0.354000E 00	0.337000E 00
5	0.298000E 01	0.469000E 01	0.423000E 01	0.911000E 00	0.560000E-01	0.493000E 00
6	0.446000E 01	0.387000E 01	0.315000E 01	0.476000E 00	0.154000E 00	0.918000E 00
7	0.179000E 01	0.318000E 01	0.357000E 01	0.655000E 00	0.421000E 00	0.770000E-01
8	0.171000E 01	0.313000E 01	0.307000E 01	0.649000E 00	0.140000E 00	0.199000E 00
9	0.307000E 01	0.501000E 01	0.458000E 01	0.995000E 00	0.450000E-01	0.0
10	0.940000E 00	0.930000E 00	0.740000E 00	0.130000E 00	0.160000E-01	0.195000E 00
11	0.497000E 01	0.537000E 01	0.535000E 01	0.823000E 00	0.690000E 00	0.690000E 00
12	0.432000E 01	0.485000E 01	0.546000E 01	0.768000E 00	0.992000E 00	0.389000E 00
13	0.217000E 01	0.178000E 01	0.243000E 01	0.203000E 00	0.740000E 00	0.120000E 00
14	0.222000E 01	0.218000E 01	0.244000E 01	0.302000E 00	0.519000E 00	0.221000E 00
15	0.288000E 01	0.490000E 01	0.511000E 01	0.991000E 00	0.450000E 00	0.249000E 00
16	0.229000E 01	0.194000E 01	0.146000E 01	0.224000E 00	0.300000E-01	0.502000E 00
17	0.376000E 01	0.339000E 01	0.271000E 01	0.428000E 00	0.127000E 00	0.772000E 00
18	0.199000E 01	0.293000E 01	0.331000E 01	0.552000E 00	0.494000E 00	0.110000E 00
19	0.495000E 01	0.408000E 01	0.419000E 01	0.461000E 00	0.824000E 00	0.714000E 00
20	0.296000E 01	0.426000E 01	0.448000E 01	0.799000E 00	0.494000E 00	0.295000E 00

PARAMETER INITIAL GUESSES

0.0 0.0 0.0 0.0 0.0

ITERATION 1

FUNCTION -0.1917259E 02 EVALUATION 1
PARAMETERS

0.0 0.0 0.0 0.0 0.0

FUNCTION -0.3775360E 02 EVALUATION 2
PARAMETERS

-0.3564905E 01 0.2887874E 01 0.3737406E 01 0.1687765E 01 0.0

FUNCTION 0.7134838E 01 EVALUATION 3
PARAMETERS

-0.1491981E 01 0.1208631E 01 0.1564176E 01 0.7063622E 00 0.0

```
ITERATION      7

FUNCTION     0.2733870E 03   EVALUATION     15
PARAMETERS
   0.9907422E 00     0.2009041E 01     0.4002668E 01     0.2670377E 01     0.4984117E 00

FUNCTION     0.2752251E 03   EVALUATION     16
PARAMETERS
   0.9932513E 00     0.2004161E 01     0.3998457E 01     0.2686703E 01     0.4976868E 00

ITERATION      8

FUNCTION     0.2752251E 03   EVALUATION     17
PARAMETERS
   0.9932513E 00     0.2004161E 01     0.3998457E 01     0.2686703E 01     0.4976868E 00

FUNCTION     0.2758589E 03   EVALUATION     18
PARAMETERS
   0.9925833E 00     0.2005203E 01     0.3999859E 01     0.2680016E 01     0.4977505E 00

ITERATION      9

FUNCTION     0.2758589E 03   EVALUATION     19
PARAMETERS
   0.9925833E 00     0.2005203E 01     0.3999859E 01     0.2680016E 01     0.4977505E 00

FUNCTION     0.2758613E 03   EVALUATION     20
PARAMETERS
   0.9924690E 00     0.2005357E 01     0.3999756E 01     0.2680399E 01     0.4977777E 00

ITERATION     10

FUNCTION     0.2758613E 03   EVALUATION     21
PARAMETERS
   0.9924690E 00     0.2005357E 01     0.3999756E 01     0.2680399E 01     0.4977777E 00

MAXIMUM OF OBJECTIVE FUNCTION     0.2758613E 03

PARAMETERS
   0.9924690E 00     0.2005357E 01     0.3999756E 01     0.2680399E 01     0.4977777E 00

FUNCTION EVALUATIONS     21   DERIVATIVE EVALUATIONS     10
```

LINEAR MAXIMUM LIKELIHOOD

MODEL 0

RESIDUALS (COMPUTED-OBSERVED)

1	-0.625038E-02	0.148897E-01	-0.309114E-01
2	-0.634384E-02	-0.115204E-01	0.109482E-02
3	-0.679111E-02	-0.342369E-03	-0.101833E-01
4	-0.301552E-02	-0.128880E-01	-0.566959E-02
5	0.831795E-02	-0.200367E-02	0.153637E-01
6	-0.698471E-02	0.276394E-01	-0.262060E-01
7	0.123024E-01	0.120821E-01	-0.294857E-01
8	0.108137E-01	0.385284E-02	0.409222E-02
9	-0.126362E-02	0.126400E-01	-0.159740E-01
10	0.105894E-02	-0.310773E-02	0.558531E-02
11	-0.967121E-02	-0.970268E-02	0.102978E-01
12	-0.125666E-01	-0.135756E-01	-0.128040E-01
13	-0.459576E-02	0.702000E-02	-0.149908E-01
14	0.445175E-02	-0.137987E-01	0.280418E-01
15	0.188637E-02	0.970173E-02	0.329304E-02
16	0.349998E-03	-0.759315E-02	-0.567436E-02
17	0.726795E-02	-0.352628E-02	0.227613E-01
18	-0.115385E-01	-0.112667E-01	-0.231743E-02
19	0.157671E-01	0.135050E-01	0.149221E-01
20	0.355625E-02	0.176649E-01	-0.770569E-03

COVARIANCE MATRIX OF RESIDUALS

0.690971E-04	0.297891E-04	0.531780E-04
0.297891E-04	0.159574E-03	-0.105259E-03
0.531780E-04	-0.105259E-03	0.286837E-03

STANDARD DEVIATIONS OF RESIDUALS

0.831246E-02	0.126323E-01	0.169363E-01

PRINCIPAL COMPONENTS

0.180641E 03	0.153643E 03	0.139201E 03	-0.414448E 01	0.431799E 02
-0.168178E 03	0.138111E 03	-0.114866E 03	-0.184861E 00	0.205584E 03
-0.754416E 02	-0.219429E 02	0.245709E 03	0.756943E 02	0.109695E 03
0.392928E 02	0.511886E 02	-0.183007E 03	0.778373E 02	-0.217428E 03
0.128923E 03	-0.635564E 02	-0.145326E 03	0.261363E 02	0.338137E 03

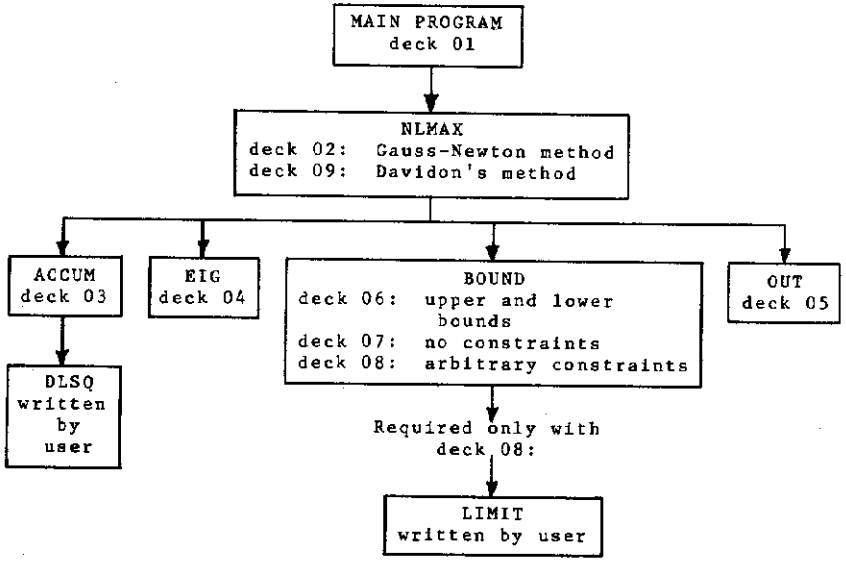
EXPECTED VALUES

0.105454E 04	-0.247545E 03	0.112139E 04	-0.489930E 03	-0.342396E 03

STANDARD DEVIATIONS

0.509559E 00	0.141422E 01	0.110646E 01	0.908202E 00	0.831404E 00

Figure 2: LEAST SQUARES PACKAGE STRUCTURE



As supplied, the program may handle up to 20 parameters, 10 variables per experiment, and 200 experiments. Instructions for modifying the capacity will be found on page 146.

The structure of the least squares subpackage is shown in Figure 2. The following routines are required:

MAIN Program, deck 01:

Calls on subroutine NLOLS.

NLMAX, deck 02:

Reads general input. Finds and prints out minimum of sum of squares function, using the Gauss-Newton method.

ACCUM, deck 03:

Computes sum of squares function and its derivatives.

EIG, deck 04:

Computes eigenvalues and vectors of a real, symmetric matrix. Incorporates automatic scaling.

OUT, deck 05:

Prints out final output, including covariance of parameter estimates.

BOUND, deck 06:

Handles the upper and lower bounds on the parameters.

NOTE: The use of deck 06 is generally recommended. Deck 07 may, however, be substituted if no bounds at all are to be placed on the parameters. Deck 08 may be used if constraints other than upper and lower bounds are to be placed on the parameters. In this case, the user must code a subroutine LIMIT according to specifications on page 69. Deck 02 may be replaced by deck 09 if one wishes to use the Davidon-Fletcher-Powell method. See page 24 for a discussion of the relative merits of the two methods.

DLSQ:

Must be written by the user to compute the function to be fitted and its derivatives.

```
      GO TO(5,6),II
6     A3=-AMD*A2/A1
      XTH(2)=A3*A(I,3)
      XTH(3)=A3*A(I,4)
5     RETURN
      END
```


B. INPUT

The input data for several problems may be stacked together, provided all the problems can be run with the same set of routines. The use of the COMMON parameter MD enables one to code subroutine DLSQ to handle several alternative equations. If the same data are then to be fitted in one run by several alternative equations, it is necessary to reproduce the data deck several times, merely changing the value of MD (see item (2) below).

(1) Title card:

Any 80 columns of BCD information to appear as title for the output.

Format: 20A4

(2) Problem definition card:

Containing three integers:

NTH = ℓ , number of unknown parameters.

LOUT = 1 if iteration by iteration output is desired.

= 2 if only final output is desired.

MD = model number.

Format: 3I5

(3) Parameter initial guess card(s):

Each containing up to eight fixed or floating point numbers:

C1(I) , I = 1,2,...NTH : initial guesses for the unknown parameters.

Format: 8E10.5

(4) Problem size card:

Containing two integers:

M = m , number of experiments.

NA = k , number of variables per experiment.

Format: 2I5

C. OUTPUT

Initial output:
Title; model number; values of the x_i for each experiment. When using BOUND deck 06, the prescribed or assigned lower and upper bounds on the parameters; the penalty function coefficients (γ_i of equation (35)); initial guesses of the parameter values (when using BOUND deck 08 these may differ slightly from the input values).

Intermediate output (omitted when LOUT = 2):
Messages indicating that penalty functions are included, reduced, or omitted (these appear only when bounds or constraints on the parameters are prescribed).

For each iteration (corresponding to a new evaluation of the derivatives) and function evaluation: the value of the objective function (= - sum of squares of residuals) and the current values of the parameters.

Analysis of objective function at its maximum:
Maximum value of objective function, and the corresponding values of the parameters; number of function and derivative evaluations required; the gradient of the objective function and its approximate Hessian (second derivative matrix) if using NLMAX deck 02, or minus the inverse of the Hessian if using NLMAX deck 09; the eigenvalues of the scaled Hessian (or its negative inverse) and the corresponding principal components (see page 46 for explanation).

Analysis of fit:
Title and model number; the residuals (i.e. the value of $g(x,\theta)$ for each experiment); sum of squares and unbiased standard deviation of residuals; the estimated values of the parameters; the standard deviations of the parameter estimates; the covariance matrix of the parameter estimates; the eigenvalues of the correlation matrix of the parameter

estimates; the principal components of the parameter estimates, with their expected values and standard deviations.

Error messages:

RESTART

Indicates that some variables exceeded the permitted magnitude (e.g. an argument of an exponentiation routine). The program attempts to continue calculations with altered values of the parameters.

FEASIBLE PARAMETER VALUES COULD NOT BE FOUND

The RESTART procedure has failed. The program abandons the current problem.

INITIAL GUESSES VIOLATE CONSTRAINT J VALUE XXX

(Produced only by BOUND deck 08.) The initial guesses of the parameters give the J-th constraint function the positive value XXX . The program abandons the current problem.

(6) No input required by DLSQ.

(7) Parameter lower bound card: 0., 0., 0.

(8) Parameter upper bound card: 0., 0., 0. (The program will replace these values with 100., 100., 100.)

The following output was produced by running the sample problem on an IBM System/360 Model 50, under OS control, using the FORTRAN G compiler.

```
PENALTY FUNCTION INCLUDED

ITERATION      1

FUNCTION    -0.4168193E 02    EVALUATION      1
PARAMETERS
   0.1CCCCCCE 01     C.1CCCCCCE 01     C.1000000E 01

FUNCTION    -0.1078524E 02    EVALUATION      2
PARAMETERS
   0.5CCCCCCE 00     0.11C1175E 01     C.1362021E 01

FUNCTION    -0.3675735E 01    EVALUATION      3
PARAMETERS
   0.25CCCCCE 00     0.1151763E 01     C.1543C32E 01

ITERATION      2

FUNCTION    -0.3675735E C1    EVALUATION      4
PARAMETERS
   0.25CCCCCE 00     0.1151763E 01     0.1543032E 01

FUNCTION    -0.4133791E 00    EVALUATION      5
PARAMETERS
   0.125CCCCE 00     0.1174004E 01     0.1974931E 01

FUNCTION    -0.1529593E CC    EVALUATION      6
PARAMETERS
   0.9596395E-01     0.1179170E 01     0.2075255E 01

ITERATION      3

FUNCTION    -0.1529593E 00    EVALUATION      7
PARAMETERS
   0.9596395E-01     0.1179170E 01     0.2075255E 01

FUNCTION    -0.1000329E-01    EVALUATION      8
PARAMETERS
   0.84C8290E-01     0.117C512E 01     C.2294042E 01

ITERATION      4

FUNCTION    -0.1000329E-01    EVALUATION      9
PARAMETERS
   0.84C8290E-01     0.117C512E 01     0.2294042E 01

FUNCTION    -0.9549953E-02    EVALUATION     10
PARAMETERS
   0.8412999E-01     0.1167720E 01     0.2311341E 01

ITERATION      5

FUNCTION    -C.9549953E-C2    EVALUATION     11
PARAMETERS
   0.8412999E-01     0.1167720E 01     C.2311341E 01

NO PENALTY FUNCTION

ITERATION      6

FUNCTION    -0.8228C71E-C2    EVALUATION     12
PARAMETERS
   0.8412999E-01     0.1167720E 01     C.2311341E 01

FUNCTION    -0.8214876E-C2    EVALUATION     13
PARAMETERS
   0.8242196E-01     0.1132924E 01     C.2343819E 01

ITERATION      7

FUNCTION    -0.8214876E-C2    EVALUATION     14
PARAMETERS
   0.8242196E-01     0.1132924E 01     C.2343819E 01

FUNCTION    -0.8214895E-C2    EVALUATION     15
PARAMETERS
   0.8241028E-C1     0.1133032E 01     C.2343698E 01

MAXIMUM OF OBJECTIVE FUNCTION    -0.8214876E-02

PARAMETERS
   0.8242196E-01     0.1132924E C1     C.2343819E 01

FUNCTION EVALUATIONS     15    DERIVATIVE EVALUATIONS     7
```

LEAST SQUARES SAMPLE PROBLEM

MODEL 1

RESIDUALS (COMPUTED-OBSERVED)

-0.566551E-02 -0.247657E-03 0.293016E-03 0.655943E-02 -0.806212E-03 -0.128460E-02 -0.445008E-02
-0.199524E-01 0.822264E-01 -0.182018E-01 -0.148018E-01 -0.147014E-01 -0.112009E-01 -0.419998E-02
0.680447E-02

SUM OF SQUARES OF RESIDUALS 0.8214876E-02

STANDARD DEVIATION 0.2616434E-01

PARAMETERS

0.824220E-01 0.113292E 01 0.234382E 01

STANDARD DEVIATIONS OF PARAMETERS

0.123729E-01 0.307817E 00 0.296197E 00

COVARIANCE MATRIX OF PARAMETERS

0.153089E-03 0.286855E-02 -0.265531E-02
0.286855E-02 0.947515E-01 -0.909339E-01
-0.265531E-02 -0.909339E-01 0.877329E-01

EIGENVALUES OF CORRELATION MATRIX

0.265652E 01 0.341744E 00 0.173319E-02

PRINCIPAL COMPONENTS

0.431217E 02 0.195221E 01 -0.200936E 01
-0.682608E 02 0.11C861F 01 -0.139293E 01
-0.362314E 01 0.234814F 01 0.232819E 01

EXPECTED VALUES

0.105631E 01 -0.763457E 01 0.781849E 01

STANDARD DEVIATIONS

0.162988E 01 0.584589E 00 0.416316E-01