THIS BOOK
CONTAINS
NUMEROUS PAGES
WITH THE ORIGINAL
PRINTING BEING
SKEWED
DIFFERANTLY FROM
THE TOP OF THE
PAGE TO THE
BOTTOM.

THIS IS AS RECEIVED
FROM THE
CUSTOMER.

NONLINEAR ESTIMATION, GAUSS-NEWTON PROCEDURE

by

STEVEN DEANE McFARLAND

B.S., Kansas State University, 1975

A REPORT

submitted in partial fulfillment of the

requrements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1976

Approved by:

Major Professor

TABLE OF CONTENTS

THIS BOOK CONTAINS NUMEROUS PAGES WITH DIAGRAMS THAT ARE CROOKED COMPARED TO THE REST OF THE INFORMATION ON THE PAGE.

THIS IS AS RECEIVED FROM CUSTOMER.

CHAPTER I. BACKGROUND

Introduction

The purpose of this report is to discuss and investigate an estimation procedure for cases involving models which are nonlinear in the parameters. The Gauss-Newton procedure will be the procedure to be discussed.

General Model: The general model to relate a set of n observed values of the random variable y to values of p mathematical variables and p parameters has the form

$$f(y_i, \underline{X}_i, \underline{\beta}, \varepsilon_i) = 0 \quad , \quad i = 1,2,\ldots,n \tag{1-1}$$

where y_i = ith component of $\underline{Y}' = (y_1, y_2, \ldots, y_n)$;

$\underline{X}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$ = ith vector of observations on the set of mathematical variables;

$\underline{\beta}' = (\beta_1, \beta_2, \ldots, \beta_p)$ = vector of parameter values;

ε_i = i-th component of the "error" vector, $\underline{\varepsilon}' = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$;

f = designates some known functional form.

A general term for $\underline{\varepsilon}$ is the error structure for the model. The error structure will generally be composed of a vector but it is possible for it to be an array. The error structure deserves special attention since the way it enters the functional form of the model can affect the ease of the estimation process. For example, consider an enzyme kinetic model,

$$y_i = \theta(x_{i1} + \varepsilon_{i1})/(\mu + x_{i2} + \varepsilon_{i2}) \quad , \quad i = 1,2,\ldots,n \ .$$

Note that the error structure enters the model in a non-additive form, it does not enter with simple addition. This type of model is difficult to analyze. Therefore, this report will consider only models with additive error structure.

General Model with Additive Error Structure: Consider the model,

$$y_i = f(\underline{\beta}, \underline{X}_i) + g(\underline{\varepsilon}_i) \quad , \quad i = 1,2,\ldots,n \tag{1-2}$$

where f = known functional form;

g = known linear functional form.

A special case of (1-2) is the General Linear Model:

$$y_i = \sum_{j=1}^{p} x_{ij}\beta_j + \varepsilon_i \quad , \quad i = 1,2,\ldots,n \tag{1-3}$$

or, in matrix notation,

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \; . \tag{1-3a}$$

The name linear comes from the fact that the functional form of the model is linear in the unknown parameters $\underline{\beta}$. The mathematical variables, $\underline{X}_i$, can enter the model in any fashion as long as they do not affect the linearity of $\underline{\beta}$. For example, consider the model,

$$y_i = \beta_1 x_{i1} + \beta_2 \log x_{i2} + \beta_3 x_{i3} x_{i2} + \beta_4 \sin x_{i4} + \varepsilon_i \quad , \quad i = 1,2,\ldots,n \; .$$

$\underline{X}_i$ enters the model in many ways, but $\underline{\beta}$ is always linear. To contrast this, consider the model,

$$y_i = \beta_1 x_{i1} + \exp(\beta_2 x_{i2}) + \varepsilon_i \quad , \quad i = 1,2,\ldots,n \; .$$

In this case both $\underline{X}_i$ and $\underline{\beta}$ enter in a nonlinear fashion. The first model can be analyzed with linear methods but the second model cannot.

The last model definition is for a Nonlinear Model with Additive Error Structure. Consider the model of (1-2),

$$y_i = f(\underline{\beta}, \underline{X}_i) + g(\underline{\varepsilon}_i) \; , \tag{1-4}$$

but with f redefined as,

f = known functional form, but nonlinear with respect to at least one β_i.

Techniques for Estimation, Linear Models

Consider the general linear model (1-3). There are three widely used techniques of estimating the unknown parameters $\underline{\beta}$:

1) Least-squares;

2) Gauss-Markov, Best Linear-Unbiased Estimates;

3) Maximum Likelihood.

The exact derivation of these three techniques will not be discussed here, but can be found in Grayhill (1961), or Searle (1971). It will be sufficient to note that under the assumption that $\underline{\varepsilon}$ is distributed multivariate normal with mean vector $\underline{0}$ and covariance structure $\sigma^2 I$, the three techniques give the same results.

While the techniques do give the same results, the least-squares method is the most commonly used. Consider the general linear model (1-3). The β_j's can be estimated from the observed responses, y_i, and associated inputs x_{ij}, $i = 1,2,\ldots,n$; $j = 1,2,\ldots,p$. This process examines the differences $y_i - f(\underline{X}_i,\underline{\beta})$ where $f(\underline{X}_i,\underline{\beta})$ is the computed response for a given x_{ij}, $i = 1,2,\ldots,n$; $j = 1,2,\ldots,p$. The sums of squares of these differences, the error sum of squares,

$$Q(X,\underline{\beta}) = \sum_{i=1}^{n} (y_i - f(\underline{X}_i,\underline{\beta}))^2 \qquad (1\text{-}5)$$

is a p variable function of $\underline{\beta}$ and is minimized with respect to $\underline{\beta}$. The minimization process is done by taking the p partial derivatives of Q and setting them equal to zero. The resulting set of p linear equations is the set of normal equations, and the normal equations are linear in the β_i's. In matrix notation, consider the general linear model (1-3a), the normal equations are:

$$(X'X)\underline{\beta} = X\underline{Y} \ , \tag{1-6}$$

and from this set of equations, point estimates for $\underline{\beta}$ and σ^2 can be calculated.

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{Y} \ ; \tag{1-7}$$

$$\begin{aligned} \hat{\sigma}^2 &= (\underline{Y}'(I - X(X'X)^{-1}X)\underline{Y})/(n-p) \ ; \\ &= (\underline{Y}'\underline{Y} - \hat{\beta}X'\underline{Y})/(n-p) \ ; \\ &= ESS/(n-p) \ . \end{aligned} \tag{1-8}$$

These estimates can be shown to have several optimum properties, Graybill (1961), Searle (1971), including independence, sufficiency, and completeness. Additionally, hypothesis testing and confidence intervals are easy to apply from the least squares approach. To look at hypothesis testing from an intuitive approach, the Principle of Conditional Error is ideal. In this manner, the hypothesis to be tested is:

$$H_o: \ H\underline{\beta} = \underline{h} \quad \text{vs.} \quad H_a: \ H\underline{\beta} \neq \underline{h} \ , \tag{1-9}$$

where H is a k by p matrix and $\underline{h}$ is a k×1 vector, both specified by the null hypothesis. The restricted model is formed such that $\underline{\beta}$ is restricted to values assigned under H_o. Normal equations are formed for this new model and a restricted $\hat{\underline{\beta}}$ is calculated. Also the Error Sums of Squares for the restricted case, ESS_r, is calculated by (1-8). The test statistic for this hypothesis is:

$$w = (ESS_r - ESS)/(p-q)\hat{\sigma}^2 \ , \tag{1-10}$$

where p is the number of β_i's in the original model and q is the number of β_i's in the restricted model. The critical point for w is the α percentage point of the F-distribution, with (p-q,n-p) degrees of freedom, where α is the type I error rate desired. An equivalent formula for w is:

$$w = (H\hat{\beta} - \underline{h})'(H(X'X)^{-1}H')^{-1}(H\hat{\beta} - h)/(q\hat{\sigma}^2) \ . \qquad (1\text{-}11)$$

Confidence intervals can be found by an expansion of the hypothesis testing method. Confidence intervals are given by:

$$p(\underline{\ell}'\underline{\beta}\varepsilon(\underline{\ell}'\hat{\underline{\beta}} \pm (t_{\alpha/2}(n-p))(\hat{\sigma}^2\underline{\ell}'(X'X)^{-1}\underline{\ell}))^{\frac{1}{2}})) = 1 - \alpha \ ,$$

where $\underline{\ell}$ is specified by the experimenter. An example follows:

Example 1, Linear Model Analyzed

Consider the model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \quad ; \quad i = 1,2,\ldots,n \qquad (1\text{-}12)$$

where the ε_i's are NID$(0,\sigma^2)$.

It is desired that the following data be fitted by (1-12), a hypothesis about $\underline{\beta}$ tested, and confidence intervals calculated for $\underline{\beta}$.

Data:

y_i:	12.1	5.5	4.6	4.5	10.8	4.9	6.0	4.2	5.3	6.7	4.0	6.1
x_{i1}:	.870	.202	.203	.198	.730	.510	.205	.670	.205	.211	.203	.264
x_{i2}:	1.69	1.17	1.17	1.21	1.63	1.59	1.14	1.92	1.22	1.71	1.11	1.37

Normal Equations:

$$\begin{vmatrix} 12 & 4.531 & 16.981 \\ 4.531 & 2.388 & 7.453 \\ 16.981 & 7.0153 & 24.864 \end{vmatrix} \begin{vmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{vmatrix} = \begin{vmatrix} 74.7 \\ 33.2144 \\ 108.93 \end{vmatrix}$$

Estimate $\underline{\beta}$:

$$\hat{\beta} = (X'X)^{-1}X'\underline{Y} = \begin{vmatrix} 4.18 & 2.66 & -3.61 \\ 2.66 & 4.13 & -2.98 \\ -3.61 & -2.98 & 3.34 \end{vmatrix} \begin{vmatrix} 74.7 \\ 33.2144 \\ 108.93 \end{vmatrix} = \begin{vmatrix} 7.8881 \\ 11.0737 \\ -4.1303 \end{vmatrix}$$

Estimate σ^2:

$$\hat{\sigma}^2 = (1.7) = (538.55 - 507.13)/9 = 3.49$$

As an example of hypothesis testing, consider,

$$H_o: \ \beta_0 - 2\beta_2 = 0 \quad ; \quad \beta_0 = 8.00 \quad \text{vs.} \quad H_a: \ \text{not } H_a \ ,$$

In matrix notation,

$$H = \begin{vmatrix} 1 & 0 & 0 \\ 1 & 0 & -2 \end{vmatrix} \quad ; \quad \underline{h} = \begin{vmatrix} 8.0 \\ 0 \end{vmatrix} \quad ; \quad n = 12; \ p = 3; \ q = 2 \ .$$

Then,

$$w = (1.10) = [-0.1119 \quad 16.487] \begin{vmatrix} 8.1404 & -2.9003 \\ -2.9003 & 1.0646 \end{vmatrix} \begin{vmatrix} -0.1119 \\ 16.487 \end{vmatrix} \frac{1}{(2)(3.49)}$$

$$= 41.29 \quad \text{and} \quad F(2,9) = 4.26 \ , \quad (\alpha = 0.05).$$

Since w is greater than F, the hypothesis H_o would be rejected.

This example illustrates application of some mathematical results which are much more difficult to derive for nonlinear models. The remaining discussion deal with problems encountered and techniques for solving them.

Transformations and Rudimentary Nonlinear Estimation

An Additivity Transformation transforms a general model to a linear model with additive error structure. This type of transformation is useful for some nonlinear models and some models which do not have additive error structure. If such an additivity transformation exists, it should be utilized since the transformed model can be analyzed with linear model

techniques which are based on exact small sample theory. If a particular model does not fit the linear models' assumptions, then large sample asymtotic theory is the only alternative.

Consider the model

$$y_i = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{2i} x_{1i} + \varepsilon_i) \quad , \quad i = 1,2,\ldots,n \ .$$

If a natural log transformation is used, then,

$$\ell n \ y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{2i} x_{1i} + \varepsilon_i \quad , \quad i = 1,2,\ldots,n \ ,$$

and both an additive error structure and a model linear in the parameters is achieved.

Another model to consider is,

$$y_i = x_i^{\beta} + \varepsilon_i \quad , \quad i = 1,2,\ldots,n \ .$$

This presents a problem. The model already exists with an additive error structure, so the model cannot be transformed to a linear model. This means nonlinear techniques will have to be used to analyze the model. Here is an example of how the error structure affects the ease of estimation. If the error structure was multiplicative and distributed log normal, then a log transformation would work as an additivity transformation.

A rudimentary estimation technique is a technique which is crude in its theoretical background and practical application. One such rudimentary technique for nonlinear estimation will be illustrated using the model

$$y_i = (\beta_1 x_i)/(x_i + \beta_2) + \varepsilon_i \quad , \quad i = 1,2,\ldots,n \ . \qquad (1\text{-}13)$$

The procedure is:

1) select an initial value of β_2, the parameter which "causes" the nonlinearity;

2) estimate β_1 from the linear models techniques described earlier;

3) calculate the residual sum of squares;

4) vary the value of β_2, and estimate β_1 again;

5) calculate the residual sums of squares and compare with the last value to determine the direction of correction for β_2;

6) continue iterations until a desired degree of accuracy is achieved.

The residual sum of squares will decrease if $\underline{\beta}$ is approaching the "correct" value. This technique will be applied to (1-13) with the data

x_i:	1	2	3	4	5
Y_i:	3.5	5.0	6.0	6.7	7.1 .

The process will be stopped when β_2 is estimated to the nearest 0.01. The results are summarized in Table 1.1.

Table 1.1. Search Procedure

Iteration	β_1	β_2	Residual S.S.	Comments
1	5.66	0	8.3720	Initial guess
2	8.05	1	0.6325	
3	10.03	2	0.0294	
4	11.84	3	0.4594	β_2 too large
5	9.65	1.8	0.0121	β_2 in correct direction
6	9.45	1.7	0.0196	β_2 too small
7	9.74	1.85	0.0127	β_2 too large
8	9.66	1.81	0.01197	
9	9.68	1.82	0.01199	stop

The process stopped at the ninth iteration since the residual sum of squares for three values of β_2 had a minimum at the middle point. The final estimated model would be:

$$y_i = (9.66\ x_i)/(x_i + 1.81) .$$

The rudimentary process gave estimates but it was quite cumbersome and difficult to use. It illustrates the need for general procedures which, during an iteration, indicate the direction and distance for the correction for a new estimate of a parameter. The procedures should also handle more than 2 or 3 parameters and be somewhat independent of the form of the model. Finally, the nonlinear procedures should have a rule to stop the procedure. A procedure which fulfills such considerations will be explored and examined.

CHAPTER II. THE GAUSS-NEWTON (OR LINEARIZATION) PROCESS

Introduction

An iterative estimation procedure is a procedure whereby estimates for the parameters are not arrived at by solving a single set of normal equations, but they are arrived at in a repetitive procedure, using the same mathematical operation each time. This method takes current values of the parameters and tries to move toward a "better" solution, in the sense that Q, the error sum of squares, is reduced. Each repetition of the procedure is called an iteration.

The General Gauss-Newton Process

One iterative estimation procedure is called Linearization, or Taylor Series, or the Gauss-Newton, process. It utilizes the results of linear least squares in each iteration. The model being considered is the Nonlinear Model with Additive Error Structure. $\underline{\beta}_o$ is a p by 1 vector of initial guesses of the parameter vector $\underline{\beta}$. An initial guess is to be improved in an iterative manner as follows. Consider the Taylor Series expansion of $f(X,\underline{\beta})$. Let

$$f(\underline{X}_i,\underline{\beta}) = f(\underline{X}_i,\underline{\beta}_o) + \sum_{j=1}^{p} Z_{oij}\gamma_{oj} \quad ; \quad i = 1,2,\ldots,n \,, \qquad (2\text{-}1)$$

which is the Taylor Series truncated after the first term of the expansion. Z_{oji} is the first partial derivative, with respect to β_j, of $f(\underline{X}_i,\underline{\beta})$ evaluated at $\underline{\beta}_o$. The expansion is about $\underline{\beta}_o$, so $\gamma_{oj} = \beta_j - \beta_{oj}$. With this expansion, the model now becomes,

$$y_i - f(\underline{X}_i,\underline{\beta}_o) = \sum_{j=1}^{p} \gamma_{oj}Z_{oji} + \epsilon_i \quad ; \quad i = 1,2,\ldots,n \,, \qquad (2\text{-}2)$$

where Z_o is the n by P matrix of Z_{oij}, $\underline{Y}$ is an n×1 vector of observed values, and $f(\underline{X},\underline{\beta}_o)$ is the n by 1 vector of functional values evaluated at $\underline{\beta}_o$ with known values of $\underline{X}_i$ (i = 1,2,...,n). Since $\underline{\gamma}$ is the difference between the true parameter vector and the initial guesses, then $\hat{\underline{\gamma}}$ is the best approximation of this difference. The best updated estimate of $\underline{\beta}$, $\underline{\beta}_1$, comes from $\hat{\underline{\gamma}}_o = \underline{\beta}_1 - \underline{\beta}_o$, which implies $\underline{\beta}_1 = \underline{\beta}_o + \hat{\underline{\gamma}}_o$. Now $\underline{\beta}_1$ can undergo the same process to obtain another set of new estimates, $\underline{\beta}_2$. In general,

$$\underline{\beta}_{m+1} = \underline{\beta}_m + \hat{\underline{\gamma}}_m$$

$$= \underline{\beta}_m + (Z_m'Z_m)^{-1}Z_m'(\underline{Y} - f(\underline{X},\underline{\beta}_m)) \ , \qquad (2\text{-}4)$$

where m = 1,2,... is the number of the current iteration.

The stopping rule approximates a test for convergence. An example of a stopping rule is:

$$\text{Stop if } \left|(\beta_{(m+1)j} - \beta_{mj})/\beta_{mj}\right| < \delta \quad , \quad j = 1,2,\ldots,p \ ,$$

$$\text{holds for all } j \ ,$$

where δ is some predetermined value, usually small. The procedure is not completely fool-proof, (Draper and Smith (1966)). Problems which may cause difficulties are:

1) The solution may converse slowly or take many iterations to stablized even though the error sum of squares decreases consistently;
2) The solution may oscillate widely, increasing and decreasing the error sum of squares, but the solution may stabilize;
3) The solution may diverge, increasing the sum of squares without bound.

The Modified Gauss-Newton Process

To combat the drawbacks of the Generalized Gauss-Newton Process, Hartley (1961) modified the process. For his modifications he made several assumptions.

1) The following derivatives exist:

   (a) $f_h(\underline{X}_i,\underline{\beta})$ is the first partial derivative with respect to β_h, evaluated at $\underline{X}_i$;

   (b) $f_{h,k}(\underline{X}_k,\underline{\beta})$ is the mixed partial derivative with respect to β_h and β_k, evaluated at $\underline{X}_i$;

   (c) $Q_h(\underline{X}_i,\underline{\beta})$ is the first partial derivative of $Q(\underline{X}_i,\underline{\beta})$, (1-5), with respect to β_h;

$$Q_h(\underline{X}_i,\underline{\beta}) + -2\ \Sigma_i(y_i - f(\underline{X}_i,\underline{\beta}))f_h(\underline{X}_i,\underline{\beta})\ ;$$

   (d) $Q_{h,k}(\underline{X}_i\ \underline{\beta})$ is the mixed partial derivative of $Q(\underline{X}_i,\underline{\beta})$, (1-5), with respect to β_h and β_k;

$$Q_{h,k}(\underline{X}_i,\underline{\beta}) = -2\ \Sigma_i(y_i - f(\underline{X}_i,\underline{\beta}))f_{h,k}(\underline{X}_i,\underline{\beta}) + 2\ \Sigma_i f_h(\underline{X}_i,\underline{\beta})f_k(\underline{X}_i,\underline{\beta})\ .$$

2) For any non-trivial set of u_j, $u = 1,2,..,p$, with $\Sigma_j u_j^2 > 0$,

$$\sum_{i=1}^{n}\ (\sum_{j=1}^{p}\ u_j f_j(\underline{X}_i,\underline{\beta}))^2 > 0\ ,$$

for all $\underline{\beta}$ in the bounded convex set S of the parameter space. This assumption guarantees the nonsingularity of the Z'Z matrix.

3) Let Q = *limit over* S^c of $Q(X,\underline{\beta})$, where S^c is the compliment of S, then assume that there exists a vector $\underline{\beta}_o$ in S such that,

$$Q(X;\underline{\beta}_o) < Q\ .$$

This assumption assures convergence of the solution.

In addition to the above assumptions, Hartley made a change in the procedure. The updated estimate of $\underline{\beta}$, $\underline{\beta}_{m+1}$, is now defined as:

$$\underline{\beta}_{m+1} = \underline{\beta}_m + \lambda \hat{\underline{\gamma}}_m \ . \tag{2-5}$$

The method of determining λ is:

1) Define a function $H(\lambda)$:

$$H(\lambda) = Q(X, \underline{\beta}_m + \lambda \hat{\underline{\gamma}}_m)$$

$$= \Sigma_i (y_i - f(\underline{X}_i, \underline{\beta}_m + \lambda \hat{\underline{\gamma}}_m))^2 \ .$$

2) Determine λ to be the value which minimizes $H(\lambda)$. An approximation to the minimum may be obtained by fitting a parabola through $H(0)$, $H(\frac{1}{2})$, $H(1)$, and using the minimum of the parabola, i.e.,

$$\lambda = \frac{1}{2} + \frac{1}{4} (H(0) - H(1))/(H(1) - 2H(\tfrac{1}{2}) + H(0)) \ .$$

The rest of the procedure remains unchanged.

Differences between the General and Modified Gauss-Newton Procedures

The actual application of either procedure is the same except for the one change in procedure made by Hartley and assumption number 3. These two changes guarantee convergence in areas where the generalized procedure would not. If assumption 3 is not violated, then $0 \leq \lambda \leq 1$, but if assumption 3 is violated then λ can become larger than 1. In the latter case, the generalized procedure would converge to a local minima, but the modified procedure may "step-over" this local minima and being the estimates into the convex bounded set S, but this does not necessarily occur. If β_m does belong to S, with an upper bound, call it Q, to the Error Sum of Squares, $Q(X,\underline{\beta})$, then $0 \leq \lambda \leq 1$. This means

$$Q(X,\underline{\beta}_{m+1}) = Q(X,\underline{\beta}_{m}) \quad \text{if} \quad \lambda = 0 \ ;$$

$$Q(X,\underline{\beta}_{m+1}) \leq Q(X,\beta_{m}) \quad \text{if} \quad \lambda > 0 \ ,$$

so $\underline{\beta}_{m+1}$ lies on the interior of S and the process continues. Since the generalized process does not make assumption 3 and it does not restrict $Q(X,\underline{\beta}_{m})$, the process can diverge or oscillate widely. Divergence occurs when the solution approaches the minimal vector, the value of $\underline{\beta}$ which minimize $Q(X,\underline{\beta})$. An iteration may "push" the solution past the minimal vector. If the new solution is now geometrically further away from the minimal vector, then the next iteration may "push" even further past the minimal vector. This leads to divergence; an increasing $Q(X,\underline{\beta})$. If the $Q(X,\underline{\beta})$ slowly converges with oscillation of the estimates, then the process will eventually converge. This is caused by a mild case of the divergence situation. The iterations do not push the solution further away from the minimal vector, but they push almost the same distance. Hartley (1961) shows that the solution vector in the modified procedure is limited by the least-squares solution and the solution is unique.

One problem with both procedure is the choice of the initial parameter vector β_{o}. If assumption 3 is violated, the modified version will sometimes converge to a local minima and the generalized version will always converge to a local minima. Hartley and Booker (1965) suggest searching the parameter space in a wide grid. Presumably this means that if an experimenter has doubts about the initial estimates, then several processes can be done with different initial estimates. If different final estimates are found, then the set with the smallest Error Sum of Squares should be used. Finally, if the processes all turn up similar final estimates, then $Q(X,\underline{\beta})$ is a fairly stable function and the results can be used with some confidence.

CHAPTER III. NUMERICAL APPLICATIONS

Introduction

Both the generalized and modified procedures were programmed in APL for the I.B.M. 370/158. Numerical examples were run to illustrate usage of the procedures and their differences.

Program Descriptions

For clarity in making comparisons, both of the main programs use like names for variables which have the same function, but this does not affect the program usage. There are two main programs, GN and MOD GN, and three subprograms, DIFF, FUNC, and H. Copies of these routines can be found in the Appendix.

GN (Program number 2, Appendix): This is the generalized Gauss-Newton procedure. It follows the algorithm described in Chapter II, and it uses two subprograms, FUNC and DIFF. There are five initial input steps:

1) 3 values: number of parameters to be estimated,
   number of mathematical variables,
   number of observations;
2) dependent variable vector;
3) observed value array of mathematical variables;
4) initial values of parameters, (note, 0 is an illegal initial value);
5) number of iterations per step, (NIT),

The sixth input is used in conjunction with the number of iterations. The program does not have a test for divergence, so there will be NIT iterations and then an option for NIT more iterations. This allows the user to stop a

run if the residual sum of squares is diverging by entering an N for No or a Y for Yes when asked if more iterations are desired. There are three lines of output per iteration:

1) iteration number;

2) residual (error) sum of squares;

3) vector of current parameter estimates.

MOD GN (Program 1, Appendix): This is the modified Gauss-Newton procedure described in Chapter II. The two subroutines, FUNC and DIFF are used in their original definition, and another routine H is used. The input is identical to input for GN, and output is identical with one additional; namely the scalar multiplier λ, which is outputed on the line after the residual sum of squares.

FUNC (Program number 4, Appendix): This is a user written routine which supplies the functional values for a given parameter vector and a given observed mathematical variable array. Program number 4 was actually the routine used for Example 1.

DIFF (Program number 3, Appendix): This is a user written routine which provides the matrix of partial derivatives, Z. The i-th column is the partial derivative of the function with respect to the i-th parameter. Program number 3 is an example, and was actually used for Example 1.

H (Program number 5, Appendix): This is not user written. It used by MODGN for the sums of squares function for given dependent vector, parameter vector, and observed mathematical variable array. H uses FUNC and is used for calculating λ.

Example 1, Fertilizer

Consider a fertilizer experiment where six responses of wheat yield have been measured corresponding to six rates of fertilizer application. The model considered is:

$$y_i = L + B \exp(Kx_i) + \varepsilon_i \quad , \quad i = 1,2,\ldots,6 \ .$$

The following data were analyzed:

x_i:	-5	-3	-1	1	3	5
y_i:	127	151	379	421	450	426 .

The initial parameters were guessed to be:

$$L = 580 \quad , \quad B = -180 \quad , \quad K = -.160 \ .$$

Example No. 1, Gauss-Newton Program, in the Appendix, shows the computer printout and the results are summarized in Table 3.1.

Table 3.1

Iteration	Res. S.S.	L	B	K
0	27376.61865	580.000	-180.000	-.160
1	14585.83999	490.418	-121.114	-.223
2	13778.76184	528.693	-163.786	-.185
3	13407.53547	515.917	-148.537	-.207
4	13393.81272	525.637	-157.748	-.196
5	13390.53578	522.019	-155.445	-.201
6	13390.17003	523.807	-157.540	-.199
7	13390.10676	523.082	-156.685	-.1999
8	13390.09559	523.399	-157.058	-.1996
9	13390.09357	523.265	-156.901	-.1997

To contrast the generalized and modified versions, the same model, data, and initial parameter values were used in a modified run, (See Example No. I, Modified Gauss-Newton Program, Appendix). The results are summarized in Table 3.2.

Table 3.2

Iteration	Res. S.S.	λ	L	B	K
0	27376.62	1.00	580.00	-180.00	-.1600
1	14590.57	0.95	495.20	-124.26	-.2197
2	13638.93	0.90	525.08	-159.41	-.1905
3	13394.54	0.96	519.36	-152.42	-.2036
4	13390.33	0.73	523.26	-156.94	-.1995
5	13390.09	0.98	523.23	-156.86	-.1997
6	13390.09		523.31	-156.95	-1.997

The contrast between the two procedures can be observed from the two tables. Both procedures arrived at the same solution, but the modified procedure took three less iterations than the generalized procedure. Number of iterations are not quite so important now, since the computer is readily available. The important point is the treatment of the residual sum of squares. Note that the reduction is faster while the estimate values are not so drastically affected. This is the important idea to guard against divergence.

Example 2, Steam

Consider a physics model which represents a possible relationship between pressure and temperature in saturated steam. The model under consideration is:

$$y_i = \beta_1 \, 10^{\beta_2 \, x_i/(\beta_3 + x_i)} + \epsilon_i \quad ; \quad i = 1,2,\ldots,14 \, .$$

The following data were analyzed:

x_i (in degrees Centigrade):

0 10 20 30 40 50 60

70 80 85 90 95 100 105 ,

y_i (pressure):

4.14	8.52	16.31	32.18	64.62	98.76	151.13
224.74	341.35	423.36	522.78	674.32	782.04	920.01

For comparison, both procedures were run with identical input. The initial estimates were found by examination of the data. When $x_i = 0$, $y_i = \beta_1$, so from the data a guess of $\beta_1 = 4$ was made. Using this estimate and two more observations, the other starting values were estimated to be $\beta_2 = 5$ and $\beta_3 = 150$. The results are summarized in Tables 3.3 and 3.4 for the generalized and modified versions, respectively.

Table 3.3. (Generalized)

Iteration	Res. S.S.	β_1	β_2	β_3
0	566161.9986	4.0000	5.0000	150.0000
1	2803.0458	8.5253	9.9407	407.2510
2	2301.6771	5.0257	8.3977	282.4967
3	1720.4980	5.2625	8.5587	294.5511
4	1718.2108	5.2674	8.5652	295.0012

Table 3.4. (Modified)

Iteration	Res. S.S.	λ	β_1	β_2	β_3
0	566161.9986	1.0000	4.0000	5.0000	150.0000
1	2687.7815	0.9892	8.4763	9.8872	404.4627
2	1942.5688	0.6472	6.2481	8.9393	326.4159
3	1720.5250	0.9018	5.3480	8.5971	297.7571
4	1718.2108		5.2681	8.5659	295.0423

In this case the sum of square function behaved so well, that the modified procedure did not converge faster. Since this was the case, initial estimates were chosen "badly," to examine what happens in this situation.

$\underline{\beta}_0'$ = [2 2 75] was used. This choice was extremely bad. The generalized procedure diverged so badly that the machine computations went out of bounds. The modified procedure converged, after 23 iterations to a local minima, $\underline{\beta}_{23}'$ = [39.67 - 4869.21 -. 390489.51] with residual sum of squares 55616.94. This clearly shows that when assumption 3 is violated, there are problems with the estimation process.

CHAPTER IV. INFERENCE

Introduction

Confidence intervals and regions, and hypothesis testing are very useful tools in modeling. Through these techniques, a parameter in a model can be examined for its contribution to the model. For example consider the linear model case,

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \ .$$

A possible hypothesis of interest might be to examine β_2 to see if it contributes to the model:

$$H_o: \ \beta_2 = 0 \quad \text{vs.} \quad H_a: \ \beta_2 \neq 0 \ .$$

If H_o is not rejected, then x_2 might not provide useful information. If a nonlinear case is to be considered, the parameters do not always correspond to an information variable. Consider the steam example of Chapter III,

$$y_i = \beta_1 \, 10^{\beta_2 \, x_i/(\beta_3 + x_i)} + \varepsilon_i \quad , \quad i = 1,2,\ldots,14 \ .$$

If a hypothesis of the form,

$$H_o: \ \beta_1 = 1 \quad \text{vs.} \quad H_a: \ \beta_1 \neq 1 \ ,$$

were tested, there the form of the model and not a specific information variable could be tested.

The theory and application are not as developed as some of the estimation procedures, but Bard (1974) and DeBruin (1971) have done some work in this area.

Confidence Regions

Bard (1974) showed that for the nonlinear case, an adaptation of linear confidence regions will work. The procedure is to select n points of $\underline{\beta}$ to satisfy

$$(\underline{\beta} - \hat{\underline{\beta}})^1 Z'Z(\underline{\beta} - \hat{\underline{\beta}}) = p\hat{\sigma}^2 F_\alpha(p,n-p) = d^2 \ , \qquad (4\text{-}1)$$

where Z is the matrix of derivatives used in the Taylor's expansion, p is the number of parameters, $\hat{\sigma}^2$ is given by (1-8), and $F_\alpha(p,n-p)$ is the α-th percentile point of the F distribution with p and n-p degrees of freedom. The resulting figure will be ellipsoidal. The procedure is summarized below.

1) Scale the parameter space by units of standard deviations of the elements of $Z'Z = C$. Let D be a p by p diagonal matrix with diagonal elements of C^{-1}.

2) Transform the parameters to $\underline{\mu} = D^{-\frac{1}{2}}(\underline{\beta} - \hat{\underline{\beta}})$, which implies selection of points in μ space satisfying

$$\mu' \, D^{-\frac{1}{2}} \, Z'Z \, D^{\frac{1}{2}} \, \mu = d^2$$

3) Let U be a p by p matrix whose columns are the characteristic vectors of $H = D^{\frac{1}{2}}Z'ZD^{\frac{1}{2}}$. The characteristic roots of H appear on the diagonal of U'HU.

4) Now the area can be transformed to a unit circle, or p-dimensional sphereoid, so the points, $\underline{t}$, can be chosen on the axes and then transformed back to the ellipsoid. Let $\underline{\mu} = d \, U(U'1 + U)^{\frac{1}{2}}\underline{t}$, and chose $\underline{t}$ at the end points of the axes of the sphereoid. There are $p^2 + p = p(p-1) + 2p$ points to chose, 2p points at the ends of the axes for the spheroid, and p(p-1) points symetrically between the axes.

5) to transform back to the confidence region (limit values of $\underline{\beta}$), use:

$$\underline{\beta} = \hat{\underline{\beta}} + Qt$$

$$Q = d \, D^{\frac{1}{2}} \, U(U'HU)^{-\frac{1}{2}}$$

$$\underline{\beta} = \underline{\beta}^2 + dD^{\frac{1}{2}} \, U(U'HU)^{-\frac{1}{2}} \, \underline{t}_u \quad ; \quad u = 1,2,\ldots,p(p+1) \ .$$

This procedure gives exact regions but the confidence coefficients are approximate. The region is simultaneously true for all β_i, therefore projection from the region to the i-th axis gives a conservative confidence interval for β_i.

Hypothesis Testing, Principle of Conditional Error

The principle of conditional error was the basis for DeBruin's (1971) dissertation. Consider the null hypothesis:

$$H_o: \; g_1(\underline{\beta}) = g_2(\underline{\beta}) = \ldots = g_r(\underline{\beta}) = 0 \; , \tag{4-2}$$

where r is the rank of X. The principle of conditional error states that the sum of squares due to H_o, SSH_o, is equal to the residual sum of squares for the restricted model, ESS_r, minus the residual sum of squares for the unrestricted model, ESS.

DeBruin examined three test statistics for (4-3),

$$T_1 = S^2/\hat{\sigma}^2 \; ; \tag{4-4}$$

$$T_2 = S^2/r\hat{\sigma}^2 \; ;$$

$$T_3 = (n-p)S^2/rn\hat{\sigma}^2 \; ,$$

where $S^2 = (ESS_r - ESS)$. He also showed that T_1 had an asymptotic Chi-Square distribution, T_2 was distributed approximately F(r,n), and T_3 was distributed approximately F(r,n-p). After simulation of the enzyme kinetic model and the application of the three test statistics, DeBruin concluded that T_3 behaved the best. This method of hypothesis testing would usually require two, or more, evaluations of the nonlinear estimation procedure to get the necessary restricted sum of square.

CHAPTER V. CONCLUDING REMARKS

Modeling has long been an effective tool in statistics. Linear modeling is the usual form of modeling applied, and since it is based on exact small sample theory, it is good to use. Sometimes linear models cannot do the job. When this happens, nonlinear techniques are used, but only if an additivity transformation does not exist. Some popular nonlinear techniques are the Method of Steepest Descent [Draper and Smith (1966)], Marquardt's Compromise [Marquardt (1963)], and the Modified Gauss-Newton [Hartley (1961)].

The Modified Gauss-Newton has been shown to guarantee convergence under some assumptions and it has been shown to give a unique solution. Problems encountered with this technique are choice of initial parameter estimates and convergence to local minima. This procedure has enjoyed some success and is available in a computer package offered through I.B.M.'s sharing library or through Stilson's NONLIN routine (1976).

To help with modeling problems, inference techniques have been developed. Bard (1974) illustrates an extension of linear theory to nonlinear cases, and DeBruin (1971) has researched the Principle of Conditional Error for the nonlinear case.

REFERENCES

1. Bard, Y. (1974), Nonlinear Parameter Estimation, Academic Press, New York and London.

2. DeBruin, R. L. (1971), Testing Hypothesis for Nonlinear Models, Doctoral Dissertation, Kansas State University.

3. Draper, N. R. and Smith H. (1966), Applied Regression Analysis, Wiley, New York.

4. Graybill, F. A. (1961), An Introduction to Linear Statistical Models, Vol. I, McGraw-Hill, New York.

5. Hartley, H. O. (1961), "The modified Gauss-Newton method for the fitting of nonlinear regression functions by least squares," Technometrics, 2:269-280.

6. Hartley, H. O. and Booker, A. (1965), "Nonlinear least squares estimation," Annals of Mathematical Statistics, 36:638-650.

7. Marquardt, D. W. (1963), "An algorithm for least-squares estimation of nonlinear parameters," Journal of the Society of Industrial and Applied Mathematics, 11:431-441.

8. Searle, S. R. (1971), Linear Models, Wiley, New York.

9. Stilson, M. L. (1976), "A Study of Three Algorithms for Nonlinear Least Squares Parameter Estimation," A Master's Report, Kansas State University.

APPENDIX

Referenced from Chapter III

Chapter 3: Program No. 1 -- This is the APL program. MOD GN, Modified Gauss-Newton procedure.

Chapter 3: Program No. 2 -- This is the APL program GN, Gauss-Newton procedure, generalized.

Chapter 3: Program No. 3 -- This is the APL program DIFF, provides matrix of derivatives, user written.

Chapter 3: Program No. 4 -- This is the APL program FUNC, provides the functional values, user written.

Chapter 3: Program No. 5 -- This is the APL program H, the sum of squares function.

Example No. 1: Hartley's Fertilizer; Program: Gauss-Newton. This is an example of the generalized routine doing the analysis.

Example No. 2: Hartley's Fertilizer; Program: Modified Gauss-Newton. This is an example of the modified procedure in analysis.

```
      ⍝ CHAPTER 3;  PROGRAM NO. 1

      ∇MODGN[⎕]∇
    ∇ MODGN
[1]   'INPUT:  NO. PARMS, NO. IND. VAR., NO. OBS.'
[2]   N←⎕
[3]   'INPUT:  DEPENDENT VARIABLE VECTOR'
[4]   Y←(N[3],1)ρ⎕
[5]   'INPUT:  INDEPENT VARIABLE ARRAY'
[6]   X←(N[3],N[2])ρ⎕
[7]   'INPUT:  INITIAL GUESS OF PARMS.'
[8]   B←(N[1],1)ρ⎕
[9]   Z←(N[3],N[1])ρ1
[10]  'INPUT:  NO. ITERATIONS PER STEP'
[11]  NOIT←⎕
[12]  DELTA←N[1]ρ0.001
[13]  DELTA←,DELTA
[14]  LAMBDA← 1 1 ρ1E¯27
[15]  KIT←0
[16]  'OUTPUT:  ITERATION NO., RES. S.S., LAMBDA'
[17] GNTWO:LIM←KIT+NOIT
[18] GNONE:' '
[19]  YD←Y-B FUNC X
[20]  RSS←(⍉YD)+.×YD
[21]  '       ';KIT;RSS;LAMBDA
[22]  'CURRENT PARMS:  ';(,B)
[23]  KIT←KIT+1
[24]  B DIFF X
[25]  ZP←⍉Z
[26]  GAMMA←(⌹(ZP+.×Z))+.×ZP+.×YD
[27]  H0←H 0
[28]  H1←H 1
[29]  LAMBDA←(0.5)+((0.25)×((H0-H1)÷((H1+H0)-2×H 0.5)))
[30]  →(LAMBDA≤1)/GNTRE
[31]  LAMBDA←1
[32] GNTRE:BA←B+(LAMBDA×GAMMA)
[33]  →((+/((,(⍉(BA-B)))÷(,(⍉B)))≥DELTA))=0)/GNEND
[34]  B←BA
[35]  →(KIT<LIM)/GNONE
[36]  'INPUT:  MORE ITERATIONS? (Y/N)'
[37]  MIT←⍞
[38]  →(MIT='Y')/GNTWO
[39] GNEND:' '
[40]  YD←Y-BA FUNC X
[41]  RSS←(⍉YD)+.×YD
[42]  'FINAL RESULTS:'
[43]  '   ITERATIONS = ';KIT
[44]  '   PARAMETERS = ';(,BA)
[45]  '   RES. S.S.  = ';(,RSS)
    ∇
```

```
⍝ CHAPTER 3;  PROGRAM NO. 2

      ∇GN[⎕]∇
    ∇ GN
[1]   'INPUT:  NO. PARMS, NO. IND. VAR., NO. OBS.'
[2]   N←⎕
[3]   'INPUT:  DEPENDENT VARIABLE VECTOR'
[4]   Y←(N[3],1)⍴⎕
[5]   'INPUT:  INDEPENT VARIABLE ARRAY'
[6]   X←(N[3],N[2])⍴⎕
[7]   'INPUT:  INITIAL GUESS OF PARMS.'
[8]   B←(N[1],1)⍴⎕
[9]   Z←(N[3],N[1])⍴1
[10]  'INPUT:  NO. ITERATIONS PER STEP'
[11]  NOIT←⎕
[12]  DELTA←N[1]⍴0.001
[13]  DELTA←,DELTA
[14]  KIT←0
[15]  ⍝ FUNC:  USER WRITTEN
[16]  ⍝        SUPPLIES FUNCTIONAL VALUES
[17]  ⍝ DIFF:  USER WRITTEN
[18]  ⍝        FORMS Z MATRIX
[19]  ⍝        MATRIX OF PARTIAL DIFFERENTIALS
[20] GNTWO:LIM←KIT+NOIT
[21] GNONE:' '
[22]  YD←Y-B FUNC X
[23]  RSS←(⍉YD)+.×YD
[24]  'ITERATION:  ';KIT
[25]  'RES. S.S.:  ';RSS
[26]  'CURRENT PARMS:  ';(,B)
[27]  KIT←KIT+1
[28]  B DIFF X
[29]  ZP←⍉Z
[30]  BA←B+(⌹(ZP+.×Z))+.×ZP+.×YD
[31]  →((+/(((,(⍉(BA-B)))÷(,(⍉B)))≥DELTA))=0)/GNEND
[32]  B←BA
[33]  →(KIT<LIM)/GNONE
[34]  'INPUT:  MORE ITERATIONS? (Y/N)'
[35]  MIT←⎕
[36]  →(MIT='Y')/GNTWO
[37] GNEND:'FINAL RESULTS:  '
[38]  '   ITERATIONS = ';KIT
[39]  '   PARMS  = ';(,BA)
[40]  YD←Y-BA FUNC X
[41]  RSS←(⍉YD)+.×YD
[42]  '   RES. S.S. = ';RSS
    ∇
```

```
⍝ CHAPTER 3;  PROGRAM NO. 3

      ∇DIFF[⎕]∇
    ∇ B DIFF X
[1]   ⍝ USER WRITTEN
[2]   ⍝ FORMS MATRIX OF PARTIAL DERIVATIVES
[3]   ⍝ I TH COLUMN HAS DER. W.R.T. ITH PARM.
[4]   Z[;1]←1
[5]   Z[;2]←(2.718281828459045)*(B[3;1]×X)
[6]   Z[;3]←B[2;1]×(,X)×(,Z[;2])
    ∇
```

```
⍝ CHAPTER 3;  PROGRAM NO. 4

      ∇FUNC[⎕]∇
    ∇ F←B FUNC X
[1]   ⍝ USER WRITTEN
[2]   ⍝ PROVIDES FUNCTIONAL VALUE FOR CURRENT
[3]   ⍝ PARAMETER ESTIMATES.
[4]   F←B[1;1]+B[2;1]×(2.718281828459045)*(B[3;1]×X)
    ∇
```

```
⍝ CHAPTER 3;  PROGRAM NO. 5

      ∇H[⎕]∇
    ∇ Q←H L
[1]   ⍝
[2]   BL←B+(L×GAMMA)
[3]   F1←BL FUNC X
[4]   FD←Y-F1
[5]   Q←(⌹FD)+.×FD
    ∇
```

```
      A EXAMPLE NO. 1:  HARTLEY'S FERTILIZER
      A PROGRAM:  GAUSS-NEWTON


      GN
INPUT:  NO. PARMS, NO. IND. VAR., NO. OBS.
?:
      3  1  6
INPUT:  DEPENDENT VARIABLE VECTOR
?:
      127  151  379  421  460  426
INPUT:  INDEPENT VARIABLE ARRAY
?:
      -5  -3  -1   1   3   5
INPUT:  INITIAL GUESS OF PARMS.
?:
      580  -180  -.160
INPUT:  NO. ITERATIONS PER STEP
?:
      5

ITERATION:  0
RES. S.S.:
     27376.61865
CURRENT PARMS:  580  -180  -0.16

ITERATION:  1
RES. S.S.:
     14585.83999
CURRENT PARMS:  490.4176683  -121.1135077  -0.223115794

ITERATION:  2
RES. S.S.:
     13778.76184
CURRENT PARMS:  528.6931634  -163.7855205  -0.1850563037

ITERATION:  3
RES. S.S.:
     13407.53547
CURRENT PARMS:  515.9166641  -148.5372768  -0.2068142273

ITERATION:  4
RES. S.S.:
     13393.81272
CURRENT PARMS:  525.6374671  -159.7481189  -0.1964706932
INPUT:  MORE ITERATIONS? (Y/N)
```

```
      A EXAMPLE NO. 1 CONT.


Y

ITERATION:  5
RES. S.S.:
     13390.53578
CURRENT PARMS:  522.0188398  -155.4449678  -0.2010284034

ITERATION:  6
RES. S.S.:
     13390.17003
CURRENT PARMS:  523.8065332  -157.5398606  -0.199080187

ITERATION:  7
RES. S.S.:
     13390.10676
CURRENT PARMS:  523.0822019  -156.6952673  -0.1999131286

ITERATION:  8
RES. S.S.:
     13390.09559
CURRENT PARMS:  523.3989402  -157.057892  -0.1995586171
FINAL RESULTS:
   ITERATIONS = 9
   PARMS  = 523.2654373  -156.9006384  -0.199709681
   RES. S.S. =
     13390.09357
```

```
      A EXAMPLE NO. 1:  HARTLEY'S FERTILIZER
      A PROGRAM:  MODIFIED GAUSS-NEWTON
      .

      MODGN
INPUT:  NO. PARMS, NO. IND. VAR., NO. OBS.
?:
      3  1  6
INPUT:  DEPENDENT VARIABLE VECTOR
?:
      127 151 379 421 460 426
INPUT:  INDEPENT VARIABLE ARRAY
?:
      -5 -3 -1  1  3  5
INPUT:  INITIAL GUESS OF PARMS.
?:
      580 -180 -.160
INPUT:  NO. ITERATIONS PER STEP
?:
      10
OUTPUT:  ITERATION NO., RES. S.S., LAMBDA

      0
     27376.61865
  1.0000000000E-27
CURRENT PARMS:  580  -180  -0.16

      1
     14590.57124
    0.9465676858
CURRENT PARMS:  495.2042596  -124.2599493  -0.219743371

      2
     13638.93004
    0.9032855026
CURRENT PARMS:  525.0783865  -159.4127474  -0.1905301221

      3
     13394.54389
    0.9658452445
CURRENT PARMS:  519.3629237  -152.4211883  -0.2035945252

      4
     13390.32512
    0.7331797808
CURRENT PARMS:  523.2579101  -156.9351206  -0.1994577839

      5
     13390.0947
    0.9762112072
CURRENT PARMS:  523.2278457  -156.8574421  -0.1997459928

FINAL RESULTS:
   ITERATIONS = 6
   PARAMETERS = 523.305171  -156.9477255  -0.1996631598
   RES. S.S.  = 13390.09313
```

ACKNOWLEDGEMENTS

The author wishes to thank his major professor, Dr. Arlin M. Feyerherm, who provided valuable counseling and encouragement throughout the preparation of this report and my academic career at Kansas State University.

Appreciation is also expressed to Dr. H. C. Fryer, Dr. S. K. Perng, and Dr. D. Johnson for their instruction in statistical theory and application. A special thanks is offered to Dr. G. A. Milliken who allowed me to audit his nonlinear instruction.

A deep debt of gratitude is also felt to my wife, Connie, who offered patience and encouragement during our years at Kansas State University.

Thanks is also due to Jan Gaines for doing a fine job of typing this report.

NONLINEAR ESTIMATION, GAUSS-NEWTON PROCEDURE

by

STEVEN DEANE McFARLAND

B.S., Kansas State University, 1975

AN ABSTRACT OF A REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1976

ABSTRACT

Linear modeling techniques have been available for some time. Estimation techniques and inference procedures are generally well-known and relatively easy to apply. However, there are cases where the desired model is not linear in the parameters. This report was concerned with this topic.

Nonlinear models can sometimes be handled by transformations, but this requires an assumption about the error structure. To analyze a nonlinear model, the least squares procedure called Gauss-Newton was initially studied. To further study nonlinear cases, Hartley's Modified Gauss-Newton procedure was examined and compared with the original version. IBM's 370/158 version of APL was used to apply both procedures to the computer, and two examples were given.

The final section of this report, dealt with an introduction to nonlinear inference. Two types of inference are examined, Bard's Confidence Regions and DeBruin's Principle of Conditional Error.