An Algorithm for Efficient Maximum Likelihood Estimation
and Confidence Interval Determination
in Nonlinear Estimation Problems

By

Patrick Charles Murphy

B.S. June 1974, University of Virginia
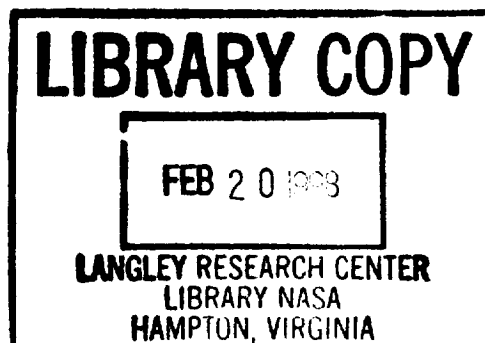M.E. June 1976, Stevens Institute of Technology

A Dissertation submitted to

The Faculty of

The School of Engineering and Applied Science
of The George Washington University in partial satisfaction
of the requirements for the degree of Doctor of Science

April 1985

Dissertation directed by
Dr. Vladislav Klein
Professor of Engineering and Applied Science

ABSTRACT

An algorithm for maximum likelihood (ML) estimation is developed with an efficient method for approximating the sensitivities. The algorithm was developed for airplane parameter estimation problems but is well suited for most nonlinear, multivariable, dynamic systems. The ML algorithm relies on a new optimization method referred to as a modified Newton-Raphson with estimated sensitivities (MNRES).

MNRES determines sensitivities by using slope information from local surface approximations of each output variable in parameter space. The fitted surface allows sensitivity information to be updated at each iteration with a significant reduction in computational effort. MNRES determines the sensitivities with less computational effort than using either a finite-difference method or integrating the analytically determined sensitivity equations. MNRES eliminates the need to derive sensitivity equations for each new model, thus eliminating algorithm reformulation with each new model and providing flexibility to use model equations in any format that is convenient.

A random search technique for determining the confidence limits of ML parameter estimates is applied to nonlinear estimation problems for airplanes. The confidence intervals obtained by the search are compared with Cramer-Rao (CR) bounds at the same confidence level. It is observed that the degree of nonlinearity in the estimation problem is an important factor in the relationship between CR bounds and the error bounds determined by the search technique. The CR bounds were found to be close to the bounds determined by the search when the degree of nonlinearity was small. Beale's measure of nonlinearity is developed in this study for airplane identification problems; it is used to empirically correct confidence levels for the parameter confidence limits. The primary utility of the measure, however, was found to be in predicting the degree of agreement between Cramer-Rao bounds and search estimates.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

LIST OF FIGURES

LIST OF TABLES

SYMBOLS

a_k incoming-parameter vector for MNRES algorithm at kth iteration

a_k^o outgoing-parameter vector for MNRES algorithm at kth iteration

a_y acceleration in y-direction, g units

A   system matrix for state equation

b   arbitrary vector through confidence region

b_w wing span, m

B   control-distribution matrix for state equation

$B_{i/i-1}$ covariance matrix for the measurement $(z_i/z_{i-1}, \theta)$

C_D drag coefficient, F_D/qS

C_L lift coefficient, F_L/qS

C_ℓ rolling-moment coefficient, M_X/qSb_w

$C_{\ell,0}, C_{m,0}, C_{n,0}$ aerodynamic moments for trimmed flight

C_n yawing-moment coefficient, M_Z/qSb_w

C_m pitching moment coefficient, M_Y/qSb_w

C_T thrust coefficient, F_T/qS

C_X longitudinal-force coefficient, F_X/qS

$C_{X,0}, C_{Y,0}, C_{Z,0}$ aerodynamic force for trimmed flight

C_Y lateral-force coefficient, F_Y/qS

C_Z vertical-force coefficient, F_Z/qS

d   squared distance between $P(\theta_w)$ and $T(\theta)$

d_{ij} ij element of the inverse of H or M

D   sum of squares of the squared distance from $P(\theta_w)$ to $P(\hat{\theta})$

e   vector sum of process noise, w_p and measurement noise, v_y

e_i scalar equation error at ith data point

E{ }   expectation operator

f	general function, or probability density function
F_{α_p}	F statistic at confidence level $1 - \alpha_p$
F_L, F_D, F_T	forces along the lift, drag and thrust vectors, N
F_X, F_Y, F_Z	forces along X, Y, and Z body axes, N
g	gravity ($1g = 9.81$ m/sec^2)
G_i	sensitivity matrix at ith data point
h	general function
H	Hessian matrix
H_α	$= \left\{ n_p s^2 F_{\alpha_p} (n_p,\ Nn_o - n_p) \right\}^{-1} H$
H_0, H_1	null and alternative hypothesis
I	arbitrary interval
I_{ij}	moments of inertia, where i,j are x, y, or z
J	cost function
k	$= n_p F_{\alpha p}(n_p,\ N - n_p)$
ℓ_{ij}	direction cosines where i,j are x, y, or z
L_1	likelihood function
L	$-\ln L_1$
m	mass
M	Fisher information matrix
M_X, M_Y, M_Z	rolling, pitching and yawing moments, N-m
n	number of surface fit points
n_o	number of outputs
n_p	number of parameters
N	number of data points
N_ψ	nondimensional nonlinearity measure
N_ϕ	nondimensional, intrinsic nonlinearity measure

n_s	number of states
p	roll rate, rad/sec
P_k	least-squares-parameter covariance matrix at kth iteration divided by σ
$P(\theta)$	estimation space
$Pr\{\ \}$	probability of $\{\ \}$
q	pitch rate, rad/sec
$\bar{q}$	dynamic pressure, $1/2\ \rho V^2$, Pa
Q_ψ	dimensional nonlinearity measure
Q_ϕ	dimensional intrinsic nonlinearity measure
r	yaw rate, rad/sec
R	measurement-noise covariance matrix
R_c	confidence region
s^2	$J(\hat\theta)/(N-n_p)$
S	vector of sensitivities
S_w	wing area, m^2
S_{1i}	vector of sensitivities for k=1 element of Y at the ith data point
S_k	vector of sensitivities for kth element of Y
$s_{k\ell}$	sensitivity of kth element of Y to ℓth element of θ
t	time, sec
Δt	processing time for Control Data Corporation CYBER 175 digital computer, sec
$T(\theta)$	tangent plane to $P(\theta)$ at $P(\hat\theta)$
u	input function or velocity along X-body axis, m/sec
U	input vector
v	velocity along Y-body axis, m/sec
v_i	vector of measurement noise at ith data point

v_x, v_y	measurement noise vectors for predictor variable X_i and response variable Y
V	airspeed, m/sec
w	index for different values of θ during N_ϕ computation or velocity along Z-body axis, m/sec
w_p	process noise vector
W	number of different θ chosen for N_ϕ computation
x_i	scalar element of X_i
X_i	vector of predictor variables at ith data point
X	matrix of predictor variables: parameters in MNRES; state variables in LR
X_s	vector of states
y_k	kth element of output vector Y
y_{ki}^j	kth element of Y, at the ith data point, evaluated at the jth surface fitting point, θ^j
y	scalar output
Y	vector of n_o outputs or N values of response variable
Y_{1i}	vector of j values of y_{ki}^j where $k = 1$
z_j	jth element of Z_i
Z_i	vector of measured outputs at ith data point
Z_i'	$= Z_i, Z_{i-1}, \ldots, Z_1$
α	angle of attack, rad
α_p	probability of type I error, $1 - \alpha_p$ is the confidence level associated with F
α_T	angle of thrust, rad
β	sideslip angle, rad
Δ	incremental value
δ_a	aileron deflection, rad
δ_{ij}	Kronecker delta
δ_r	rudder deflection, rad

ϵ	vector of equation error
θ	vector of unknown parameters
$\theta*$	vector of optimal parameters
Θ	pitch angle
λ	ratio of likelihood functions, L
ν_i	vector of residuals at ith data point
ρ	air density, kg/m^3
σ	standard error
Φ	value of ψ which minimizes Q_ψ
ϕ	roll angle, rad
μ	general statistic equal to $J(\theta) - J(\hat{\theta})$
ψ	$\theta_w - \hat{\theta}$
ω	subset of Ω
Ω	parameter space

Subscripts:

E	measured quantity
H,L	highest and lowest cost, respectively
i,j,k,ℓ	general indices
r,k	iteration number (except as δ_r which is rudder deflection)
t	true value
T	thrust
0	initial condition

Superscript:

j	index of surface-fitting points

Abbreviations:

CPU	central processing unit
FPS	flexible polyhedron search

LR	linear regression
MAX	ML program using MNRES algorithm
MAXLIK	ML program using MNR algorithm
ML	maximum likelihood
MNR	modified Newton-Raphson
MNRES	modified Newton-Raphson with estimated sensitivities
MSR	modified stepwise regression
NR	Newton-Raphson

Matrix exponents:

T	transpose matrix
-1	inverse matrix

Mathematical notation:

^	estimated quantity when over symbol
•	derivative with respect to time when over symbol
∇	gradient operator
Δ	delta indicates incremental value

The following aerodynamic derivatives are referenced to a system of body axes with the origin at the airplane center of gravity:

$$C_{Y_\beta} = \frac{\partial C_Y}{\partial \beta} \qquad C_{Y_p} = \frac{\partial C_Y}{\partial \frac{pb}{2V}} \qquad C_{Y_r} = \frac{\partial C_Y}{\partial \frac{rb}{2V}}$$

$$C_{Y_{\delta a}} = \frac{\partial C_Y}{\partial \delta_a} \qquad C_{Y_{\delta r}} = \frac{\partial C_Y}{\partial \delta_r} \qquad C_{Y_{\alpha\beta}} = \frac{\partial^2 C_Y}{\partial \alpha \partial \beta}$$

$$C_{\ell_\beta} = \frac{\partial C_\ell}{\partial \beta} \qquad C_{\ell_p} = \frac{\partial C_\ell}{\partial \frac{pb}{2V}} \qquad C_{\ell_r} = \frac{\partial C_\ell}{\partial \frac{rb}{2V}}$$

$$C_{\ell_{\delta a}} = \frac{\partial C_\ell}{\partial \delta_a} \qquad C_{\ell_{\delta r}} = \frac{\partial C_\ell}{\partial \delta_r} \qquad C_{\ell_{\alpha\beta}} = \frac{\partial^2 C_\ell}{\partial \alpha \partial \beta}$$

$$C_{\ell_{\alpha p}} = \frac{\partial^2 C_\ell}{\partial \alpha \partial \frac{pb}{2V}} \qquad C_{n_\beta} = \frac{\partial C_n}{\partial \beta} \qquad C_{n_p} = \frac{\partial C_n}{\partial \frac{pb}{2V}}$$

$$C_{n_r} = \frac{\partial C_n}{\partial \frac{rb}{2V}} \qquad C_{n_{\delta a}} = \frac{\partial C_n}{\partial \delta_a} \qquad C_{n_{\delta r}} = \frac{\partial C_n}{\partial \delta_r}$$

$$C_{Y_{\beta^3}} = \frac{1}{6} \frac{\partial^3 C_Y}{\partial \beta^3} \qquad C_{n_{\beta^3}} = \frac{1}{6} \frac{\partial^3 C_n}{\partial \beta^3} \qquad C_{n_{\alpha r}} = \frac{\partial^2 C_n}{\partial \alpha \partial \frac{rb}{2V}}$$

$$C_{X_\alpha} = \frac{\partial C_X}{\partial \alpha} \qquad C_{X_{\delta e}} = \frac{\partial C_X}{\partial C_e} \qquad C_{Z_\alpha} = \frac{\partial C_Z}{\partial \alpha}$$

$$C_{Z_{\delta e}} = \frac{\partial C_Z}{\partial C_e} \qquad C_{Z_q} = \frac{\partial C_Z}{\partial \frac{q\bar{c}}{\partial v}} \qquad C_{m_\alpha} = \frac{\partial C_m}{\partial \alpha}$$

$$C_{m_{\delta e}} = \frac{\partial C_m}{\partial \delta_e} \qquad C_{m_q} = \frac{\partial C_m}{\partial \frac{q\bar{c}}{\partial v}}$$

Chapter I

INTRODUCTION

Problems in dynamics may be divided into three categories. By considering a general dynamical system, f, with input, U, and output, Y, the categories can be defined as: (1) the classical problem where input, U, and system, f, are given and the response, Y, is to be determined; (2) the controls problem where system, f, and the desired response, Y, are given and input, U, is to be determined; and (3) the identification problem where input, U, and output, Y, have been measured and the system, f, is to be modelled.

The theory for system identification provides a way for modelling an unknown system based upon input and output information. The identification theory incorporates a priori knowledge of the dynamic processes and stochastic processes involved; thus, the identification problem is not usually characterized as a black box problem. In fact, system identification problems are usually characterized, as done in reference 1, by three factors: (1) class of models; (2) class of inputs; and (3) a criterion for state and parameter estimation. The models and inputs may be deterministic or stochastic and the criterion (cost function) may be based on statistical theory or numerical considerations.

Implementation of the identification theory usually follows four basic stages. The first stage requires the design of an experiment. This requires the identification objectives to be specified, system configuration and conditions to be stated, and an input form selected. Determining an optimal input for identification can be critical to identification success; all the modes of a system must be excited in order to identify the system correctly and completely. The second stage is for model structure determination (a more comprehensive term is model characterization). The model is assumed to be linear or nonlinear, time varying or time invariant, with or without process noise, and with or without measurement noise, etc. The unknown parameters in the model may include system parameters as well as initial conditions, bias terms, measurement and process noise characteristics. The third stage involves parameter and state estimation. Parameter and state estimation provide mean values and standard error estimates; these are obtained by finding an extremum of some optimality

criterion. State estimation can be better characterized as a filtering problem; a Kalman filter is commonly used for this problem. The fourth stage is verification. This is accomplished by comparing estimates from different data sets and different estimation techniques. In addition, other sources provide comparisons; in the case of airplanes, both wind tunnel and theoretical predictions are used. Verification is also accomplished through sensitivity analysis and through analysis of residuals and model predictive capabilities.

The importance of system identification theory to aircraft technology has developed for several reasons. A primary reason is that it provides an alternate approach to determining aircraft characteristics (parameters). Comparing results with other techniques is always good scientific practice. Purely theoretical approaches or purely experimental approaches (wind tunnels) have in many instances failed to accurately predict prototype characteristics. Flight testing offers an opportunity to observe actual vehicle performance resulting in better calibration and understanding of wind tunnel results and more accurate modelling for ground-based simulators.

The development of aircraft parameter estimation paralleled the developments in estimation and system theory. Early flight test studies centered on steady-state maneuvers and free oscillations. These studies were time consuming and provided limited information. The main interest was to obtain basic aerodynamic parameters, termed stability and control derivatives, from linear dynamic models combined with linear aerodynamic models. In the early 1950's Greenberg and Shinbrot developed a least squares approach to analyze simple transient maneuvers (ref. 2,3,4). However, without computers the simplest flight test problem with only four unknown parameters took 24 hours to analyze (ref. 5). A major development for aircraft parameter estimation occurred in the mid 1960's. This development was the introduction of large-capacity, high-speed digital computers and highly automated data acquisition systems. In 1968, when Larson applied the method of quasi-linearization (ref. 6) and Taylor and Iliff (ref. 7) introduced the modified Newton-Raphson method, a new stimulus

was given to parameter estimation. Other contributions came in the early 1970's from Mehra (ref. 8), Mehra and Stephner (ref. 9), and Rault (ref. 10).

During the past decade an increasing concern has been the application of estimation theory to nonlinear systems. Much of this has been stimulated by aerospace applications. Today, incorporating nonlinear dynamics with linear aerodynamic models is commonly performed in flight test data analysis. The techniques are well established for flight regimes where the aircraft aerodynamic model can be expressed as a linear function of states and control inputs. However, modelling the combination of nonlinear dynamics with nonlinear aerodynamics and estimating the parameters associated with that model present many difficulties. The need to identify the best mathematical representation (model structure) and estimate the associated parameters for nonlinear flight regimes has motivated further development of identification and estimation techniques.

A new approach to airplane parameter estimation and confidence interval determination is offered in this study as a contribution toward building a more general and unified airplane identification methodology. The more general methodology starts with the work done in reference 11. In reference 11 a useful technique for model structure determination, where nonlinear aerodynamic effects are present, is suggested. The suggested technique uses a Modified Stepwise Regression (MSR), along with several testing criteria to determine a parsimonious, yet adequate, model. The limitation of this technique (as with any least squares method) is that the estimates are asymptotically biased and variance estimates are based on simplifying assumptions which are valid only for the "classical" linear regression. This limitation can be skirted by applying the commonly used Maximum Likelihood (ML) technique using the model structure determined by the regression and the regression estimates as an initial guess. The ML approach has much more favorable asymptotic properties (ref. 12), and it provides estimates of the Cramer-Rao (CR) bounds for the parameter variance.

There is a computational cost, however, for the more favorable asymptotic properties of the ML technique. Dynamic systems, such as aircraft systems, require substantial computational effort at each step of the optimization process. At each step the equations of motion must be integrated to obtain time histories of each state and output variable. In addition, most ML algorithms use a Modified Newton-Raphson (MNR) optimization scheme which requires integrating sensitivity equations. This accounts for most of the computational effort since the number of state and sensitivity equations to be integrated at each iteration is equal to the number of states plus the product of the number of states and the number of unknown parameters. Several states and 20 to 30 parameters are not uncommon for one flight condition. If a model is desired throughout the entire flight envelope, the computational requirements become overwhelming since analysis of various flight conditions may require more than one candidate model. A very efficient ML estimation algorithm is desirable to reduce the computational requirements for processing a large number of parameters and candidate models.

Besides the greater computational cost associated with the ML/MNR algorithm an additional difficulty in using the algorithm is that it requires the user to have prior knowledge of the model structure to formulate the sensitivity equations and, thus, to formulate the algorithm. This can be very burdensome when modelling aircraft in nonlinear flight regimes since model structure may change significantly from one flight condition to another. Therefore, it is very advantageous to have an algorithm which is independent of sensitivity equations.

Reducing computational requirements of the ML method requires careful examination of the optimization methods utilized in the algorithm. Although nonlinear, unconstrained optimization problems have been studied quite extensively (ref. 13), little has been done to improve the optimization techniques as they apply to aircraft estimation problems. Gupta and Mehra considered the numerical aspects of computing ML estimates for linear dynamic systems in state-vector form and methods for speeding up convergence (ref. 14). Trankle, et al., considered the difficulties associated

with the use of a nonlinear dynamic model in ML parameter estimation and parameter covariance estimation; sensitivity calculation methods were also considered (ref. 15). More recently, Trankle, et al., considered the overall methodology of system identification for nonlinear aerodynamic models including computational aspects of the problem (ref. 16). In reference 17, a nonlinear least-squares algorithm is developed which uses a linear-surface approximation of a scalar-response variable to eliminate derivative calculations altogether. The algorithm is applied to test problems which do not involve dynamic systems. Presented in the current study is a significantly improved maximum likelihood algorithm which relies on an optimization scheme referred to as a modified Newton-Raphson method with estimated sensitivities (MNRES). A surface approximation is also used in MNRES; however, it is treated differently by developing an algorithm which retains derivative information in a Newton-Raphson method for multivariable, dynamic systems. This is done to provide directional information for the convergence process and to provide covariance information. With the MNRES approach, sensitivity equations are eliminated and a significant reduction in computational demand is obtained.

Another difficulty in using the ML technique is that the CR inequality provides only a lower bound measure of precision for an unbiased estimator. It is known from practical application of ML that this lower bound can differ from the variance obtained, for example, by repeated measurements (Klein ref. 18). Attempts have been made, therefore, to either modify the CR bounds by considering a band-limited measurement noise (Balakrishna, ref. 19, and Maine and Iliff, ref. 20) or to estimate the parameter variance directly from measured data (Rault ref. 10). Advances in statistical methods also came about with the availability of high-speed computers. Beale, in 1960, considered the problem in nonlinear estimation of determining the approximate parameter confidence regions using likelihood ratios (ref. 21). In addition, a measure of nonlinearity was developed to assess the quality of the approximation. Surprisingly, Beale's work has had very little application since it was published (ref. 23). In 1979, Mereau and Provost (ref. 23) made use of the likelihood ratio

approach to determine confidence regions for aircraft systems. In 1980, Mereau and Raymond (ref. 24) developed a search procedure to find the "iso-distances" defining the confidence regions.

The goal of the current study is to provide improved techniques for estimating parameters and their confidence limits in nonlinear, multivariable dynamic systems, in particular, aircraft systems. The improved techniques will provide: (1) increased efficiency for the estimation process; (2) elimination of the need for a priori knowledge of sensitivity equations; (3) more accurate assessment of the parameter error bounds than obtained using Cramer-Rao bounds; and (4) an adaptation of Beale's approach to the airplane estimation problem. In addition, a unified methodology for solving nonlinear airplane identification problems is inherent in this study.

The development of this work begins in chapter II with a description of the airplane model and the regression method used to determine it. Chapter III describes the parameter estimation techniques used and their statistical properties. The primary estimation method used to determine airplane parameters, in this study, is maximum likelihood. Linear regression methods are also presented since special forms are developed for use in MNRES and because they are used in the model structure determination scheme of Chapter II. In Chapter IV the MNRES method is presented with a discussion of its various forms and properties. Also presented briefly are some commonly used optimization methods which are used to compare with MNRES. Chapter V develops the theory for confidence interval estimation and the adaptation of Beale's work to the airplane problem. Finally, chapter VI presents the results and discussion for the application of these methods to simulated and real data.

Chapter II

MODEL CHARACTERIZATION

Model characterization, as discussed in the introduction, establishes the known or assumed characteristics of the model to be used in the identification process. Since system identification usually does not involve a black box problem where nothing is known about the model in advance, qualitative statements describing the class of model, optimal inputs and statistical properties of the measurements are normally provided.

Generally, the more information available to characterize the model the greater the likelihood of successful identification. Of course, attempting a complete representation of a dynamic system, such as an airplane, is extremely difficult, if not impossible. Actually, a complete model is unnecessary. The objective in identification is to select the simplest model that allows proper determination of the desired unknown parameters from measured data. The principle of parsimony is usually applied. This principle states that given a choice of two models having equal residual variances choose the model with fewer parameters. Therefore, the objective in identification is to choose a parsimonious, yet adequate, model. Very complex models may be justified to obtain an accurate description of the system motion but it is clearly detrimental to the estimation process. If the information content in the measured data is very limited, or if too many parameters are required, the estimation algorithm may provide inaccurate estimates or it may fail.

The models considered in this study represent dynamic systems. Dynamic systems are characterized by having derivatives with respect to time included in the model in addition to the dependent and independent variables. One of the possible general forms for these systems is

$$\dot{X}_s = f(X_s, U, \theta, t) \qquad X_s(0) = X_{s0} \qquad (2\text{-}1)$$

$$Y = h(X_s, U, \theta, t) \qquad (2\text{-}2)$$

where X_s is a vector of state variables, Y is a vector of output variables, U is a vector of input variables, and θ is a vector of unknown parameters. The time variable, t, may or may not appear explicitly. This form is not as restrictive as it first appears; many problems can be cast into the matrix differential form above.

Several difficulties arise when estimation techniques are applied to dynamic systems. A major difficulty is the significantly greater computational demand associated with solving matrix differential equations. In an estimation algorithm these equations are solved repetitively. A difficulty can also arise when integrating the equations of motion because the boundary conditions or initial conditions are not always known exactly. Therefore, in many estimation problems the initial conditions are treated as unknown parameters. Another difficulty is that the solution to the differential equations can be very different depending on sometimes very small changes in the unknown parameters. For example, a first degree system is stable or unstable depending only on the sign of the damping term. Nonlinear systems can amplify this type of problem. The success of the estimation can depend on the initial guesses for the parameters since failure may occur when a parameter is outside a stability boundary. Unfortunately, obtaining stability boundaries is really only practical for linear, time-invariant systems. Finally, numerical difficulties with truncation and rounding errors are always present where numerical differentiation and integration are performed.

A. AIRPLANE EQUATIONS OF MOTION

The particular dynamic system of interest to this study is the airplane, modelled by equations in the general form

$$\dot{X}_s = f(X_s, U, \theta) \qquad X_s(0) = X_{s0} \qquad (2\text{-}3)$$

$$Y = g(X_s, U, \theta) \qquad (2\text{-}4)$$

The equations of motion used are referred to a body axes system (See fig. 1). The equations were developed with the following assumptions:

1. The airplane is a rigid body.

2. The effect of spinning rotors is negligible.

3. The airplane has a plane of symmetry in xz plane.

4. There are no external disturbances to the airplane.

5. Thrust is accounted for as part of C_Z and C_X where

$$C_X = C_T \cos \alpha_T + C_L \sin \alpha - C_D \cos \alpha$$

$$C_Z = C_T \sin \alpha_T - C_L \cos \alpha - C_D \sin \alpha$$

The resulting nine equations represent a six degree-of-freedom, coupled, nonlinear system where the kinematic relations are expressed in terms of direction cosines. They are given as follows:

EQUATIONS OF MOTION

$$\dot{u} = -qw + rw + g\ell_{xz} + \frac{\bar{q}S_w}{m} C_X \qquad (2\text{-}5)$$

$$\dot{v} = -ru + pw + g\ell_{yz} + \frac{\bar{q}S_w}{m} C_Y \qquad (2\text{-}6)$$

$$\dot{w} = -pw + qu + g\ell_{zz} + \frac{\bar{q}S_w}{m} C_Z \qquad (2\text{-}7)$$

$$\dot{p} = \frac{I_z}{I_x I_z - I_{zx}^2} F_1 + \frac{I_{zx}}{I_x I_z - I_{zx}^2} F_2 \qquad (2\text{-}8)$$

$$\dot{q} = \frac{I_{zx}}{I} \left(r^2 + p^2 \right) + pr \frac{I_z - I_x}{I_y} + \frac{\bar{q} S_w \bar{c}}{I_y} c_m \qquad (2\text{-}9)$$

$$\dot{r} = \frac{I_x}{I_x I_z - I_{zx}^2} F_2 + \frac{I_{xz}}{I_x I_z - I_{xz}^2} F_1 \qquad (2\text{-}10)$$

$$\dot{\ell}_{xz} = r \ell_{yz} - q \ell_{zz} \qquad (2\text{-}11)$$

$$\dot{\ell}_{yz} = -r \ell_{xz} + p \ell_{zz} \qquad (2\text{-}12)$$

$$\dot{\ell}_{zz} = p \ell_{xz} - p \ell_{yz} \qquad (2\text{-}13)$$

where

$$F_1 = \left(I_y - I_z\right)q\, r + I_{zx}\, p\, q + \bar{q}\, S_w\, b_w\, C_\ell \qquad (2\text{-}14)$$

$$F_2 = \left(I_x - I_y\right)p\, q - I_{zx}\, q\, r + \bar{q}\, S_w\, b_w\, C_n \qquad (2\text{-}15)$$

The nondimensional aerodynamic forces and moments, C_X, C_Y, C_Z, C_ℓ, C_m, C_n (shown in fig. 1), are usually approximated by a Taylor series expansion around steady trimmed flight conditions or by polynomial splines (see ref. 25). The form of the aerodynamic model equations is

$$y(t) = \theta_0 + \theta_1\, x_1 + \theta_2\, x_2 + \ldots + \theta_{n_{p-1}}\, x_{n_{p-1}} \qquad (2\text{-}16)$$

or in vector form

$$y_i = X_i\, \theta \qquad (2\text{-}17)$$

where $y(t)$ or y_i represent one of the nondimensional aerodynamic forces or moments at time t or at the ith data point. The stability and control derivatives are represented by θ_1 to θ_{n_p-1} and corresponding to an initial trim flight condition the trim forces or moments are represented by

θ_0. The x_1 to x_{n_p-1} represent any function of the state and control variables chosen for the model. The row vector X_i is given as

$$X_i = \begin{bmatrix} 1 & x_1 & x_2 & \cdots & x_{n_{p-1}} \end{bmatrix}$$

(2-18)

In general, the form of the aerodynamic equation is unknown; however, for estimation the form must be postulated. The form may vary significantly from one flight condition to another.

The output equations for this study are as follows:

$$V = \overline{u^2 + v^2 + w^2}$$

(2-19)

$$\beta = \sin^{-1}(v/V)$$

(2-20)

$$\alpha = \tan^{-1}(w/u)$$

(2-21)

$$\Theta = \sin^{-1}(-\ell_{xz})$$

(2-22)

$$\phi = \tan^{-1}\left[\ell_{yz}/\ell_{zz}\right]$$

(2-23)

$$a_x = \frac{1}{g}\left[\dot{u} + q\,w - r\,v - q\,\ell_{xz}\right]$$

(2-24)

$$a_y = \frac{1}{g}\left[\dot{v} + r\,u - p\,w - g\,\ell_{yz}\right]$$

(2-25)

$$a_z = \frac{1}{g}\left[\dot{w} + p\,v - q\,u - g\,\ell_{zz}\right]$$

(2-26)

The airplane identification problem can be made more tractable by treating longitudinal and lateral cases separately. This is accomplished by providing the required lateral information to the longitudinal case (or the required longitudinal information to the lateral case) in the form of measured input variables. This has been used successfully in many other studies, for example, reference 18. Thus, the states, outputs, and inputs for the two cases are given as follows:

for the longitudinal case,

$$X_s = \begin{bmatrix} u & w & q & \ell_{xz} & \ell_{yz} & \ell_{zz} \end{bmatrix}^T \tag{2-27}$$

$$Y = \begin{bmatrix} V & \alpha & q & \Theta & a_x & a_z \end{bmatrix}^T \tag{2-28}$$

$$U = \begin{bmatrix} \delta_e & \beta_E & v_E & p_E & r_E & \phi_E \end{bmatrix}^T \tag{2-29}$$

and for the lateral case,

$$X = \begin{bmatrix} v & p & r & \ell_{xz} & \ell_{yz} & \ell_{zz} \end{bmatrix}^T \tag{2-30}$$

$$Y = \begin{bmatrix} \beta & p & r & \phi & a_y \end{bmatrix}^T \tag{2-31}$$

$$U = \begin{bmatrix} \delta_a & \delta_r & u_E & w_E & \Theta_E & q_E & \alpha_E \end{bmatrix}^T \tag{2-32}$$

where the subscript E indicates a measured quantity.

B. MODEL STRUCTURE DETERMINATION

The goal of model structure determination is to determine an analytical representation of the system which can be classified as an adequate

model. An adequate model is one which sufficiently fits the data, allows successful estimation of the parameters, and has good prediction capabilities. In aeronautical applications the form of the rigid body equations of motion is known. The only uncertainty, with regard to model structure, is in the aerodynamic model equations (eq. (2-16)). One of the successful methods for determining the model structure of these equations from measured data is based on stepwise regression.

In the stepwise regression approach, after postulating the aerodynamic model equation, the determination of significant terms among the candidate variables and estimation of corresponding parameters follows. The variable chosen for entry into the regression equation is the one that has the largest correlation with y after adjusting for the effect on y of the variables already selected. The parameters are estimated by the least squares technique. At every step of the regression, the variables incorporated into the model in previous stages and a new variable entering the model are reexamined. Any variable which provides a nonsignificant contribution (due to correlation with more recently added terms) is removed from the model. The process of selecting and checking variables continues until no more variables are admitted to the equation and no more are rejected. Experience shows, however, that the model based only on the significance of individual parameters in model equation (2-16) can still include too many terms and, therefore, may have poor prediction capabilities. Several criteria for the selection of an adequate model are introduced in reference 11 and the details of the whole procedure are explained in references 11 and 26.

Chapter III

PARAMETER ESTIMATION

In this study Maximum Likelihood (ML) and Linear Regression (LR) techniques are used to estimate parameters. ML is used to estimate both airplane parameters and their standard errors (Cramer-Rao lower bounds) from flight data. The ML algorithm is used with various optimization schemes which are described in chapter IV. LR is used for three different applications in this study. The three applications are: (1) estimating aerodynamic model structure; (2) estimating airplane parameters (starting values for ML); and (3) estimating sensitivities in MNRES. The first and second applications of LR were accomplished using stepwise regression as described in the last chapter. The third application was accomplished using an algorithm developed in this study.

A. LINEAR REGRESSION

Linear regression analysis is a part of statistical theory which generally deals with the determination of relationships between response and predictor variables. One application of LR theory is curve fitting or surface fitting. In this application, the predictor variables (independent variables) are assumed to be deterministic and known without error; response variables (dependent variables) may have error. A numerical method commonly used in curve fitting to compute empirical coefficients is the method of Least Squares (LS). In this method, the same model form as equation (2-16) can be used to fit the curve or surface. The solution for the unknown parameters or coefficients are found by minimizing the sum of squares of the error between known data points and computed data points determined by the model. The LS method is valid only for linear problems; that is, problems where the unknown parameters occur linearly in the model regardless of whether the model structure itself is linear or nonlinear. LS can be solved in a batch mode or recursive mode and both modes have application for determining the sensitivities in MNRES.

- 14 -

A-1 BATCH PROCESSING

Batch processing of data in the LS method is probably the most commonly used approach for curve fitting problems. The model form given by equation (2-17) can be written as

$$y_i = X_i S + e_i \qquad i = 1, 2, \ldots, N \qquad (3-1)$$

where y_i is the ith value of one response variable; X_i is the ith vector of predictor variables; S is the vector of unknown coefficients; and e_i is the equation error at the ith data point. This error may contain measurement noise, process noise, and/or modelling error. However, no assumptions are made about the statistical properties of e. In application to MNRES, y_i represents one element of the output vector, $Y(\theta)$; X_i represents the ith set of values for the vector of unknown parameters, θ; and S is the n_p vector of coefficients to be computed. If a first degree n_p-polynomial expansion is chosen for X_i, then each element of S will be the desired sensitivities (slopes).

Applying the least squares criterion which requires minimization of the mean square error gives the cost function as

$$J(S) = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} [y_i - X_i S]^2 \qquad (3-2)$$

and minimization requires

$$\frac{\partial J(S)}{\partial S} = 0 = \sum_{i=1}^{N} X_i^T [y_i - X_i S] \qquad (3-3)$$

Solving for S gives

$$\hat{S} = [\sum_{i=1}^{N} X_i^T X_i]^{-1} \sum_{i=1}^{N} X_i^T y_i \qquad (3-4)$$

A-2 RECURSIVE PROCESSING

Recursive processing provides a significant reduction in memory requirements for the MNRES algorithm. However, a specialized form of recursive least squares is needed for surface fitting in MNRES. Normally in a recursive least squares problem the purpose is to update parameter estimates based on N data points with some new information so that the updated estimates are based on N+1 data points. In the following derivation a least squares recursive algorithm is designed specifically for the MNRES algorithm. MNRES requires the parameters be updated using both incoming new information and outgoing old information so that the estimates are always based on a constant number of data points.

As in the batch mode the surface fitting is performed to obtain slope or derivative information. Consider the least squares problem formulated as

$$Y = XS + \epsilon \qquad (3\text{-}5)$$

where Y is a vector of n data points on a surface to be fit by the model given as XS. S is a vector of n_p unknown coefficients (slopes) and X is an n by n_p matrix defined as

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1j} \\ 1 & x_{21} & x_{22} & \cdots & x_{2j} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ 1 & x_{i1} & x_{i2} & \cdots & x_{ij} \end{bmatrix} \qquad i=n,\ j=n_p \qquad (3\text{-}6)$$

Least squares estimation gives a solution as

$$\hat{S} = \left[X^T X\right]^{-1} X^T Y \qquad n \geq n_p \qquad (3\text{-}7)$$

- 16 -

Now defining a recursive relation for the k+1 iteration

$$P_{k+1} = [X_k^T X_k]^{-1} \qquad (3-8)$$

and the updating equation as

$$P_{k+1} = [P_k^{-1} + a_k^T a_k - a_k^{oT} a_k^{o}]^{-1} \qquad (3-9)$$

where the row vector a_k is the new set of θ to be included in X and a^o is the outgoing set of θ to be removed from X which produced the highest value of the cost function. The recursive relation for S is

$$S_{k+1} = S_k - P_{k+1}[a_k^{oT} y_k^{o} - a_k^T y_k + (a_k^T a_k - a_k^{oT} a_k^{o})S_k] \qquad (3-10)$$

where y_k^{o} and y_k are one of the scalar elements of the outgoing and incoming vector Y, respectively, at the kth iteration.

The derivation for equations (3-9) and (3-10) is as follows. Define Z as the common elements of X between two iterations. Partitioning X for the k-1 and kth iterations results in

$$X_{k-1} = \begin{bmatrix} a_k^{o} \\ Z_k \end{bmatrix} \qquad (3-11)$$

$$X_k = \begin{bmatrix} a_k \\ Z_k \end{bmatrix} \qquad (3-12)$$

By using equations (3-11) and (3-12), the following relations can be written:

$$X_{k-1}^T X_{k-1} = a_k^{oT} a_k^{o} + Z_k^T Z_k \qquad (3-13)$$

- 17 -

$$X_k^T X_k = a_k^T a_k + Z_k^T Z_k \tag{3-14}$$

From equation (3-13)

$$Z_k^T Z_k = X_{k-1}^T X_{k-1} - a_k^{oT} a_k^o \tag{3-15}$$

Substituting equation (3-15) into (3-14) gives

$$X_k^T X_k = X_{k-1}^T X_{k-1} - a_k^{oT} a_k^o + a_k^T a_k \tag{3-16}$$

Substituting equation (3-16) into (3-8) gives

$$P_{k+1} = \left[X_{k-1}^T X_{k-1} - a_k^{oT} a_k^o + a_k^T a_k \right]^{-1} \tag{3-17}$$

which can also be written as

$$P_{k+1} = \left[P_k^{-1} - a_k^{oT} a_k^o + a_k^T a_k \right]^{-1} \tag{3-18}$$

Applying the same development to equation (3-7) gives

$$S_{k+1} = P_{k+1} \left[X_{k-1}^T y_{k-1} - a_k^{oT} y_k^o + a_k^T y_k \right] \tag{3-19}$$

and substituting equations (3-7) and (3-8) delayed a step into (3-19):

$$S_{k+1} = P_{k+1} \left[P_k^{-1} S_k - a_k^{oT} y_k^o + a_k^T y_k \right] \tag{3-20}$$

Expanding equation (3-20) gives

$$S_{k+1} = P_{k+1} P^{-1} S_k - P_{k+1} a_k^{oT} y_k^o + P_{k+1} a_k^T y_k \tag{3-21}$$

Noting that

$$S_k - P_{k+1} \ P_{k+1}^{-1} \ S_k = 0 \qquad (3-22)$$

and then adding equation (3-21) to (3-22) gives

$$S_{k+1} = S_k - P_{k+1} \ P_{k+1}^{-1} \ S_k + P_{k+1} \ P_k^{-1} \ S_k - P_{k+1} \ a_k^{oT} \ y_k^o + P_{k+1} \ a_k^T \ y_k \qquad (3-23)$$

$$S_{k+1} = S_k - P_{k+1} \left[P_{k+1}^{-1} \ S_k + P_k^{-1} \ S_k + a_k^{oT} \ y_k^o - a_k^T \ y_k \right] \qquad (3-24)$$

Combining terms in (3-24) gives

$$S_{k+1} = S_k - P_{k+1} \left[\left(P_{k+1}^{-1} - P_k^{-1} \right) S_k - a_k^{oT} \ y_k^o - a_k^T \ y_k \right] \qquad (3-25)$$

and using equation (3-18) yields the desired relation

$$S_{k+1} = S_k - P_{k+1} \left[\left(a_k^T \ a_k - a_k^{oT} \ a_k^o \right) S_k + a_k^{oT} \ y_k^o - a_k^T \ y_k \right] \qquad (3-26)$$

A-3 STATISTICAL PROPERTIES OF LS ESTIMATES

Although the least squares technique, a numerical procedure, is not based on any statistical formulation, the least squares estimator is often characterized in statistical terms since the estimates can be treated as random variables.

In the general least squares problem both process noise and measurement noise occur in the data. The model has the form

$$Y_t = X_t \ \theta + w_p \qquad (3-27)$$

where Y_t is an N by 1 vector of the response variable for the N data measurement points. Subscript t indicates this is the true value of the variable without noise. X_t is an N by n_p matrix of the state and input

variables and θ is the n_p by 1 vector of unknown parameters. w_p is the N by 1 vector of process noise. The measurements provide

$$Y = Y_t + v_y \tag{3-28}$$

$$X = X_t + v_x \tag{3-29}$$

where v_y and v_x are the measurement noise in Y and X. A typical assumption made about the process and measurement noise is that they are stationary, zero mean and independent random processes. The solution to this least squares problem (from the last section) is

$$\hat{\theta} = \left[X^T X\right]^{-1} X^T Y \tag{3-30}$$

Premultiplying equation (3-27) by $\left[X^T X\right]^{-1} X^T$ and substituting equations (3-28), (3-29), and (3-30) it is found that

$$\hat{\theta} = \theta + \left[X^T X\right]^{-1} X^T \left[w_p + v_y - v_x \theta\right] \tag{3-31}$$

Therefore, the expected value of the estimate error has the form

$$E\{\hat{\theta} - \theta\} = -E\{\left[X^T X\right]^{-1} X^T v_x\} \theta \tag{3-32}$$

and the covariance matrix is

$$\text{cov}(\hat{\theta} - \theta) = E\{(X^T X)^{-1} X^T e \, e^T X(X^T X)^{-1}\}$$

$$+ E\{(X^T X)^{-1} X^T v_x \, \theta \, \theta^T \, v_x X(X^T X)^{-1}\} \tag{3-33}$$

where

$$e = w_p + v_y.$$

From these equations it can be seen that the least squares estimator is biased even if the measurement noise v_x, v_y, and process noise, w_p, are zero mean and independent. Only with the additional assumption that X is known without error (i.e., $v_x = 0$), as might be the case in a curve fitting problem, will the estimates be unbiased. Note the covariance matrix is affected by all the measurement and process noise.

B. MAXIMUM LIKELIHOOD (ML)

The general parameter estimation problem for an airplane involves solving the nonlinear estimation problem in the presence of both process and measurement noise while modelling the airplane with the coupled, non-linear equations of motion. One of the advanced techniques commonly used for this problem because of its superior statistical properties is maximum likelihood.

B-1 ALGORITHM DEVELOPMENT

Assume the outcome of an experiment is N observations of the $(n_0 \times 1)$ output vector Z_i, $i=1,2,\ldots,N$, which depends on the unknown parameter vector θ. In general, the unknown parameters are the aerodynamic parameters, initial conditions, and measurement and process noise statistics. Let $f(Z_1,Z_2,\ldots Z_N/\theta)$ be the conditional probability density function for the measurements given θ. The maximum likelihood estimate is the estimate for which the outcome of the experiment Z_i, $i=1,2,\ldots,N$ is most likely to occur; that is, the probability density is maximized. The problem is stated as

$$\hat{\theta} = \max f(Z_1,Z_2,\ldots,Z_N/\theta) \tag{3-34}$$

Using the property of joint probability density functions that

$$f(Z_1,Z_2,Z_3/\theta) = f(Z_3/Z_2,Z_1,\theta)\ f(Z_2/Z_1,\theta)\ f(Z_1/\theta) \tag{3-35}$$

the density function can be written as

$$f(Z_i, i=1, N/\theta) = f(Z_N/Z'_{N-1}, \theta) f(Z_{N-1}/Z'_{N-2}, \theta) .. f(Z_2/Z_1, \theta) \; f(Z_1/\theta) \qquad (3\text{-}36)$$

$$= \prod_{i=1}^{N} f(Z_i/Z'_{i-1}, \theta) \qquad (3\text{-}37)$$

where

$$Z'_{i-1} = Z_{i-1}, \; Z_{i-2}, \ldots, Z_1 \qquad (3\text{-}38)$$

If the Z_i measurements are treated as fixed, the density function becomes a function of θ only. This function is usually referred to as the likelihood function; that is,

$$L_1(\theta) = \prod_{i=1}^{N} f(Z_i/Z'_{i-1}, \theta) \qquad (3\text{-}39)$$

Consequently, the problem of finding a maximum likelihood estimate becomes the problem of finding the θ which maximizes the likelihood function.

To define the likelihood function, the distribution of the measurements given θ must be defined. If the distribution of the measurement and process noise is Gaussian, then the distribution of the measurements given θ will also be Gaussian and can be uniquely determined by computing the mean and covariance. The mean is given as

$$E\{ Z_i/Z'_{i-1}, \; \theta \} = \hat{Y}_{i/i-1} \qquad (3\text{-}40)$$

where $\hat{Y}_{i/i-1}$ is the best estimate of the measurement at time i given measurements up to and including the previous point. By definition, the covariance is given as

$$\text{cov}\{Z_i/Z'_{i-1}, \theta\} = E\{[Z_i - \hat{Y}_{i/i-1}][Z_i - \hat{Y}_{i/i-1}]^T\}$$

$$= E\{\nu_i\nu_i^T\} = B_{i/i-1} \qquad (3-41)$$

where ν_i are the residuals. Now the problem is to compute the conditional mean, $\hat{Y}_{i/i-1}$, and covariance, $B_{i/i-1}$. For systems including process noise, these values can be obtained by use of a Kalman filter. It has been shown (ref. 8) that for high sampling rates (as is commonly used to collect flight test data), the residuals ν_i tend toward a Gaussian distribution. Therefore, the distributions for both ν_i and for $(Z_i/Z_{i-1}, \theta)$ are reasonably assumed to be Gaussian. In systems without process noise, some simplifications are possible. In particular, the residual error may be written as

$$\nu_i = Z_i - \hat{Y}_i \qquad (3-42)$$

where Y_i is the predicted value of the output vector at time i. Also, the mean and covariance of the residuals may be assumed constant in time and written as

$$E\{\nu_i\} = 0 \qquad (3-43)$$

$$E\{\nu_i\nu_j^T\} = R\delta_i; \qquad (3-44)$$

where δ_{ij} is the Kronecker delta. With these assumptions, the conditional probability density function can be written as

$$f(Z_i/Z'_{i-1}, \theta) = (2\pi)^{-n_o/2} |R|^{-1/2} \exp\{-1/2\ \nu_i^T R^{-1}\ \nu_i\} \qquad (3-45)$$

- 23 -

Hence the likelihood function using equation (3-45) is

$$L_1(\theta) = \prod_{i=1}^{N} f(Z_i/Z'_{i=1}, \theta) = (2\pi)^{-Nn_0/2} \prod_{i=1}^{N} |R|^{-1/2}$$

$$\exp\left\{-1/2 \sum_{i=1}^{N} v_i^T R^{-1} v_i\right\} \qquad (3\text{-}46)$$

The negative log of the density function is more convenient to use and since the function is monotonic, there is no change to the problem except that maximizing the density function equates to minimizing the negative log of the density function. Thus, the more commonly used negative log likelihood function becomes

$$L(\theta) = \frac{1}{2} \sum_{i=1}^{N} \left(Z_i - \hat{Y}_i\right)^T R^{-1} \left(Z_i - \hat{Y}_i\right) + \frac{N}{2} \ln|R| + \text{constant} \qquad (3\text{-}47)$$

The unknown R can be estimated by minimization of the likelihood function with respect to R. This produces

$$\hat{R} = \frac{1}{N} \sum_{i=1}^{N} v_i v_i^T \qquad (3\text{-}48)$$

where

$$v_i = Z_i - \hat{Y}_i(\hat{\theta}) \qquad (3\text{-}49)$$

After substituting $\hat{R}$ into equation (3-47), the final form of the cost function, as used in this study, is obtained. This cost function is given as

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{N} \left(Z_i - \hat{Y}_i\right)^T \hat{R}^{-1} \left(Z_i - \hat{Y}_i\right) + \text{constant} \qquad (3\text{-}50)$$

The cost function given by (3-50) is the same as that used in an output error technique except the measurement noise covariance matrix is used as a weighting matrix. The problem is now in the form of an unconstrained optimization problem where the cost function given in equation (3-50) must be minimized with respect to the unknown parameter vector, θ. The unknown parameters determined by this method, for this study, are the airplane stability and control derivatives and trim coefficients. In addition, measurement noise statistics (weighting matrix) and parameter standard errors are determined.

The standard errors determined are the Cramer-Rao lower bounds providing a measure of the maximum achievable accuracy for the parameters. These are defined by the Cramer-Rao inequality

$$E\left\{(\theta - \hat{\theta})(\theta - \hat{\theta})^T\right\} \geq -E\left\{\frac{\partial^2 L(\theta)}{\partial\theta\partial\theta^T}\right\}^{-1} \qquad (3-51)$$

where

$$-E\left\{\frac{\partial^2 L(\theta)}{\partial\theta\partial\theta^T}\right\} = M \qquad (3-52)$$

and M is usually referred to as the Fisher Information matrix. It is assumed in this study that the approximated Hessian matrix, H, from the optimization procedure is a good approximation of the Fisher Information matrix. The solution using a gradient optimization scheme generally has the following form for the kth iteration

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \hat{H}_k^{-1} g_k \qquad (3-53)$$

where

$$g_k = \frac{\partial J(\theta)}{\partial \theta}\bigg|_{\theta_0} \quad \text{and} \quad \hat{H} \approx M$$

B-2 STATISTICAL PROPERTIES OF ML ESTIMATES

The maximum likelihood method is popular, especially in flight test data analysis, because of the excellent large sample properties of its estimates. Although ML estimates do not possess optimal properties for small samples, sampling experiments (ref. 27) have shown that the ML method produces acceptable estimates in many situations. The ML estimate is robust and sufficient when a sufficient statistic exists. The large sample properties of an ML estimate are summarized here (derivation of these properties is given by Cramer, ref. 28):

1. Asymptotically unbiased: $\lim_{N \to \infty} E\{\hat{\theta}\} = \theta$

2. Asymptotically efficient: $E\{(\theta - \hat{\theta})(\theta - \hat{\theta})^T\} \geq -E\{\frac{\partial^2 L(\theta)}{\partial \theta \partial \theta^T}\}^{-1}$

3. Consistent: $\lim_{N \to \infty} Pr\{(\theta - \hat{\theta}) \leq \varepsilon\} = 1$, with ε arbitrarily small

4. Asymptotically normal: $\hat{\theta} = N(\theta, M^{-1})$

Asymptotic unbiasedness and consistency are very similar. However, consistency implies that if an estimator is consistent for θ, it is also consistent for any well behaved function of θ. Thus, consistency is more significant than unbiasedness. Asymptotic efficiency is a statement of the Cramer-Rao inequality; therefore, for large samples the Cramer-Rao lower bounds are obtained.

Chapter IV

OPTIMIZATION TECHNIQUES

The ML parameter estimates are obtained by solving an unconstrained, nonlinear optimization problem; that is, find $\theta*$ which minimizes the cost function $J(\theta)$. The necessary and sufficient conditions for this problem are as follows:

1. $J(\theta)$ is differentiable at $\theta*$.
2. $\nabla J(\theta*) = 0$.
3. $\nabla^2 J(\theta*) > 0$.

The theory for solving unconstrained, nonlinear optimization problems is often based on the assumption that the cost function $J(\theta)$ is a quadratic function of θ. This approximation provides a more tractable theory and allows basic theorems and properties of the optimization methods to be readily established. Corresponding theorems for solving general nonlinear functions of θ are very difficult to prove. However, techniques developed using the quadratic assumption are still very effective for nonlinear functions. Many techniques for solving nonlinear minimization problems are developed from practical experience.

Optimization techniques for unconstrained problems can be divided into two categories: derivative methods and search methods. Derivative methods may be further classified by the order of the derivatives used; search techniques can be divided into direct search and random search. Some techniques combine search and derivative methods; however, these hybrid methods are not considered in this study.

The choice between optimization categories depends on the particular problem. Search methods determine the optimization path solely from cost function evaluations and, therefore, do not require as much algorithm preparation as needed when using sensitivity equations. Search methods also do not need the regularity and continuity conditions that derivative methods need for the cost functions. In many unconstrained, nonlinear programming problems, however, derivative methods will converge faster (ref. 13), particularly for estimation problems involving dynamic systems.

This was demonstrated for airplane systems both in this study and in reference 29.

Various derivative techniques are available for a variety of nonlinear programming problems; however, no one technique is best for all problems. For example, the steepest descent method works better away from the minimum whereas Newton's method works better near a minimum. A compromise between these two techniques is the modified Newton-Raphson method (MNR). MNR belongs to a class of methods known as quasi-Newton or large step gradient methods; these methods approximate the Hessian matrix or its inverse while only using first derivative information.

The derivative information can be computed in a variety of ways. For dynamic systems, integrating analytically-derived expressions (sensitivity equations) for the derivatives is probably the most accurate as well as the most time-consuming. One alternative is a numerical approximation scheme. Finite difference (f.d.) methods are often used since they eliminate the additional burden of deriving and incorporating sensitivity equations into the algorithm. However, the f.d. methods require about the same level of computational effort as integrating the sensitivity equations. Another option is the proposed surface fitting method of the MNRES algorithm presented in section IV-B.

A. COMMONLY USED METHODS

Two optimization schemes, representing the two main categories of methods, are selected in this study primarily to provide a benchmark comparison with MNRES. They are commonly used in aircraft estimation and control problems and, therefore, are a good indicator of the relative merit of MNRES. The two optimization methods are the flexible polyhedron search (FPS) and the modified Newton-Raphson (MNR) method. More details are provided in references 13 and 30 for the FPS and in references 12 and 31 for the MNR. A variation of the MNR will be used in which the derivative information is computed by using finite differences (refs. 15 and 16). Both the f.d. form and the sensitivity equation form of MNR are used in this study.

A-1 FLEXIBLE POLYHEDRON SEARCH

Since FPS has been found to be advantageous in some aircraft design and control applications (ref.32), it may be a good candidate for reducing computational demands in aircraft estimation problems. FPS avoids derivative calculations, where the quasi-Newton methods spend most of the computational time. The algorithm is independent of model form and, thus, is readily applicable to any aerodynamic model.

Consider the unconstrained optimization problem of minimizing a scalar function of n_p variables $J(\theta)$. The FPS method uses a flexible polyhedron surface with $n_p + 1$ vertices where each vertex is defined by a vector θ. The vertex θ_H, producing the highest value of $J(\theta)$, is projected through the centroid of the remaining vertices to define a new vertex. This new vertex, and the remaining ones without θ_H, form a new polyhedron. This operation is called a "reflection." Figure 2 shows two steps in this process for the case with two unknown parameters. If the new vertex produces a lower cost than θ_L (the vertex producing the smallest $J(\theta)$), then an expansion takes place and a new vertex is located farther out along the same projection. Similarly, if higher costs are found, a contraction takes place. The minimum of the cost function is found by repeatedly deleting the point having the highest value of $J(\theta)$ and adding new projected points that produce lower $J(\theta)$. The flexible polyhedron is able to adapt to the shape of $J(\theta)$ by stretching down slopes, contracting near minima, and changing direction in curved valleys.

A-2 MODIFIED NEWTON-RAPHSON METHOD

This report is primarily concerned with nonlinear aircraft estimation problems. Since the MNR approach is commonly used for these problems, it is included as a benchmark algorithm. Although it is computationally burdensome to estimate derivatives, this information enables relatively fast convergence of the optimization process. In fact, Newton's method converges in one pass for cost functions which are quadratic. Hence, Newton and quasi-Newton techniques used for estimation problems of dynamic systems are expected to converge faster when the quadratic approximations

for the cost functions are valid. Also, these methods provide both step size and direction for each iteration. In some problems, however, additional control of step size is needed to ensure convergence. Since removing the requirement to solve sensitivity equations is desirable, the MNR algorithm in this report will use a simple finite-difference method except when otherwise noted. This is not too costly in terms of computational time (refs. 15 and 16); however, care must be taken to obtain the derivatives as accurately as possible.

The MNR and the MNRES algorithms are the derivative methods of interest to this study. As discussed in an earlier section the problem is to minimize the weighted square of the errors between the computed model outputs and the actual measured outputs. It is assumed that only the measured outputs are corrupted by noise and that the noise is zero mean and uncorrelated. This leads to a nonlinear estimation of unknown parameters. Consider the system equations (repeating (2-3) and (2-4)) and the measurement equations,

$$\dot{X}_s = f(X_s, U, \theta) \qquad\qquad X_s(0) = X_{s0} \qquad\qquad (2\text{-}3)$$

$$Y = h(X_s, U, \theta) \qquad\qquad\qquad (2\text{-}4)$$

$$Z_i = Y_i + v_i \qquad\qquad i = 1, 2, \ldots, N \qquad\qquad (4\text{-}1)$$

with

$$E(v_i) = 0 \qquad\text{and}\qquad E(v_i v_j^T) = R\delta_{ij} \qquad\qquad (4\text{-}2)$$

where X_s, U, and Y are the state, input and output vectors, respectively. θ is the unknown parameter vector. Z_i and v_i are the measurement vector and measurement noise vector, respectively, at $t = t_i$. R is a diagonal measurement noise covariance matrix which is, under the above assumptions, equal to the covariance matrix of the residuals. Without process noise the ML cost function to be minimized is given by equation (3-50), where the added constant and multiplicative factor of 1/2 are dropped without affecting the solution.

$$J(\theta) = \sum_{i=1}^{N} (Z_i - \hat{Y}_i)^T \hat{R}^{-1} (Z_i - \hat{Y}_i) \qquad (3\text{-}50)$$

The matrix R is given by equation (3-48)

$$\hat{R} = \frac{1}{N} \sum_{i=1}^{N} \nu_i \nu_i^T \qquad (3\text{-}48)$$

where

$$\nu_i = Z_i - \hat{Y}_i(\theta_0) \qquad (3\text{-}49)$$

and θ_0 is the initial estimate of the unknown parameter vector. The MNR method accomplishes the minimization by expanding Y, the computed output vector, about θ_0, the initial unknown parameter vector. A Taylor Series expansion of Y truncated to first order is

$$\hat{Y}(\theta) = \hat{Y}(\theta_0) + \frac{\partial Y}{\partial \theta}\Big|_{\theta_0} \Delta\theta \qquad (4\text{-}3)$$

where $\Delta\theta = \theta - \theta_0$. Then by substituting into J, a quadratic approximation of J is obtained. The increment $\Delta\theta$ is the unknown. Differentiating J with respect to θ and equating the derivative to zero to find a minimum results in

$$\frac{\partial J}{\partial \theta} = -\sum_{i=1}^{N} G_i^T \hat{R}^{-1} \nu_i + \sum_{i=1}^{N} G_i \hat{R}^{-1} G_i \Delta\theta = 0 \qquad (4\text{-}4)$$

where

$$G_i = \left\{ \frac{\partial y_k}{\partial \theta_j} \right\}_i \qquad (4\text{-}5)$$

y_k and θ_j are the kth and jth elements of the Y and θ vectors, respectively. Solving for $\Delta\theta$,

$$\Delta\hat{\theta} = \left[\sum_{i=1}^{N} G_i^T \hat{R}^{-1} G_i \right]^{-1} \sum_{i=1}^{N} G_i^T \hat{R}^{-1} \nu_i \qquad (4\text{-}6)$$

This is often written as

$$\Delta\hat{\theta} = -M^{-1} \left. \frac{\partial J}{\partial \theta} \right|_{\theta_0} \qquad (4\text{-}7)$$

emphasizing the Fisher Information matrix, M, and gradient terms. For the kth iteration the estimate $\hat{\theta}_{k+1}$ is given as $\hat{\theta}_{k+1} = \hat{\theta}_k + \Delta\hat{\theta}_{k+1}$. In this study convergence is achieved when $\Delta J_k/J_k$ and $\Delta\theta_k/\theta_k$ are less than .001. The sensitivities, G_i, are determined separately from the above steps. This may be done by integrating the sensitivity equations or using a finite difference approximation or by using MNRES.

B. MNRES METHOD

The MNRES method developed in this paper is essentially an MNR optimization algorithm with an efficient method for estimating sensitivities. As in the ML/MNR algorithm previously described, the same equations (eqs. (4-1) through (4-7) and (3-48) through (3-50)) apply for ML/MNRES; however, the sensitivities, G_i, are computed by using slope information from local surface approximations of $Y(\theta)$. The approximations are made near the series expansion point of equation (4-3). The sensitivities obtained from the fitted surface are determined with less computational effort than that obtained by either a finite-difference method or integrating analytically-determined sensitivity equations.

The MNRES algorithm is readily optimized for a particular application in that the user can select both the type of surface and the method of fitting the surface. Two types of surfaces which are very practical in

aeronautical applications are nth-order polynomials and splines. Two effi-
cient methods of fitting the surface are by solving n_p simultaneous equa-
tions for n_p unknowns (algebraic solution) and by solving a redundant set
of equations for n_p unknowns (least squares solution). The tradeoffs in
choosing a surface and a surface fitting method involve the choice between
accuracy of the sensitivities and computational effort.

B-1 ALGEBRAIC SOLUTION

The MNRES algorithm is best described by considering the computation-
ally least demanding approach of using a linear-surface approximation.
Expanding $Y(\theta)$ in a first-degree polynomial in θ for each point in time
and at n_p+1 different points in the n_p parameter space gives

$$y_{ki}^j (\theta^j) = s_{k0} + s_{k1} \theta_1^j + \ldots + s_{kp} \theta_p^j \qquad (4\text{-}8)$$

where i indicates the ith point in time; k indicates the kth element of the
output vector $Y(\theta)$; and j indicates one of the n_p+1 sample points used
to fit equation (4-8) to $Y(\theta)$. Note that

$$y_k^j(\theta^j) = y_k(\theta) \qquad (4\text{-}9)$$

at each of the n_p+1 points. The sample points are chosen by allowing a
small perturbation of each parameter around the point where the sensitivi-
ties are desired. Alternatively, the perturbation size can be selected to
reflect the relative significance of each parameter to the model. This
allows for larger perturbations of the less sensitive parameters and
smaller perturbations for the very sensitive parameters, thus providing
higher quality derivative calculations. This alternative is discussed
further at the end of this section. The slopes s_{k1} to s_{kn} are the
desired sensitivities, $(\partial y_k / \partial \theta_j)_i$. s_{k0} is the value of $y_k(\theta)$
evaluated at the series expansion point of equation (4-3). Note that
because this is a linear surface, the slopes are constant over the surface

and need not be evaluated specifically at s_{k0}. If a higher degree polynomial is fit to $y_k(\theta)$, the slopes will vary across the fitted surface and, therefore, must be evaluated specifically at s_{k0}. Consider the matrix representation of (4-8) for the first element of Y and for the n_p+1 sample points at time "i":

$$Y_{1i} = X \, S_{1i} \tag{4-10}$$

Note that y_k is the kth element of the output vector, Y, and y^j_k is the jth element of the surface fitting vector, Y_k. Matrix X contains n_p+1 rows defining the n_p+1 sample points. Expanding equation (4-10) to show the vector and matrix elements gives:

$$
\begin{bmatrix}
y^0_{1i} \\
y^1_{1i} \\
\cdot \\
\cdot \\
\cdot \\
y^n_{1i}
\end{bmatrix}
=
\begin{bmatrix}
1 & \theta^0_1 & \theta^0_2 & \cdots & \theta^0_n \\
1 & \theta^1_1 & \theta^1_2 & \cdots & \theta^1_n \\
\cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdot & & \cdot \\
1 & \theta^n_1 & \theta^n_2 & \cdots & \theta^n_n
\end{bmatrix}
\begin{bmatrix}
s_{10} \\
s_{11} \\
\cdot \\
\cdot \\
\cdot \\
s_{1n}
\end{bmatrix}_i
\tag{4-11}
$$

Since s_{10} is a known point, equation (4-11) can be simplified. The first line in equation (4-11) can be eliminated by subtracting it from the other n_p equations. Thus,

$$\Delta Y_{1i} = \Delta X \, S_{1i} \tag{4-12}$$

$$
\begin{bmatrix} y_{1i}^1 - y_{1i}^0 \\[2ex] y_{1i}^2 - y_{1i}^0 \\[1ex] \cdot \\ \cdot \\ \cdot \\[1ex] y_{1i}^n - y_{1i}^0 \end{bmatrix} = \begin{bmatrix} \Delta\theta_1^1 & \Delta\theta_2^1 & \cdots & \Delta\theta_n^1 \\[2ex] \Delta\theta_1^2 & \Delta\theta_2^2 & \cdots & \Delta\theta_n^1 \\[1ex] \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\[1ex] \Delta\theta_1^n & \Delta\theta_2^n & \cdots & \Delta\theta_n^n \end{bmatrix} \begin{bmatrix} s_{11} \\[2ex] s_{12} \\[1ex] \cdot \\ \cdot \\ \cdot \\[1ex] s_{1n} \end{bmatrix} \qquad (4\text{-}13)
$$

where

$$
\Delta\theta_k^j = \theta_k^j - \theta_k^0
$$

Thus, at time "i" the sensitivities for the first element in Y are given by

$$
S_{1i} = [\Delta X]^{-1} \Delta Y_{1i} \qquad (4\text{-}14)
$$

Note that the ΔX matrix is independent of time. This enables the sensitivities to be calculated rapidly during each iteration of the algorithm. This is a key factor in reducing the computational effort of the algorithm; in effect, the integration of the $n_s n_p$ sensitivity equations has been replaced by a set of n_0 matrix multiplications.

Figure 3 shows, geometrically, two iterations for the case where θ is of dimension two and a linear surface is used to fit a scalar y. The expansion at time "i" is

$$
\begin{bmatrix} y_i^1 - y_i^0 \\[2ex] y_i^2 - y_i^0 \end{bmatrix} = \begin{bmatrix} \Delta\theta_1^1 & \Delta\theta_2^1 \\[2ex] \Delta\theta_1^2 & \Delta\theta_2^2 \end{bmatrix} \begin{bmatrix} s_1 \\[2ex] s_2 \end{bmatrix}_i \qquad (4\text{-}15)
$$

During the first iteration, this expansion requires that $y(\theta)$ be evaluated at $n_p+1=3$ points: y^0, y^1, y^2. Computationally, the first iteration is the most costly phase of the MNRES algorithm. Each evaluation of Y requires that the equations of motion be integrated. The linear surface (indicated by the solid-line triangle in fig. 3) is fit and the slopes (sensitivities) are thereby determined. The algorithm proceeds, as in the ordinary MNR method, using equation (4-7) to obtain

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \Delta\hat{\theta}_{k+1} \qquad\qquad (4-16)$$

The estimated sensitivity values, s_{kj}, are used to define the elements of matrix G in equation (4-6). The new Y is evaluated (by integration of equations of motion) at θ_{k+1} to get $y^3(\theta)$. At this point the MNRES algorithm has reduced the sensitivity problem to solving a set of simultaneous equations. This is done by eliminating the θ^j in X which produced the greatest contribution to the cost in $J(\theta)$ and replacing that information with the newest estimate of θ. The new surface (indicated by the dotted triangle) in figure 3 assumes y^0 was the high cost point and so eliminated it from the fitted surface. The slopes of the new surface provide the sensitivities for the MNRES algorithm to proceed. In this scheme, a check should be made to ensure the new $Y^j(\theta_{k+1})$ produces a smaller value of $J(\theta)$. In some cases step-size control or complete restarting may be needed. Note that initialization of the algorithm requires that n_p+1 integrations be performed for the n_p+1 trajectories, Y^j. After this initializing pass only one integration of the system equations is needed to evaluate the cost, $J(\theta)$; outputs, $Y(\theta)$; and to update parameter estimates for each iteration.

As mentioned previously, in practice it is beneficial to choose the perturbation size in a different fashion from that used in a simple finite-difference method. Simply using a 1 percent perturbation on each element of θ to obtain the corresponding perturbation in each element of $Y(\theta)$ is

not optimum for derivative calculations. Experience has shown that it is beneficial to use perturbation sizes which reflect the importance of the parameter to the model. By computing the sensitivities of $\theta_j^2 M_{jj}$ for each parameter and then letting the perturbation sizes be scaled inversely proportional to the normalized ratios of sensitivities, more accurate derivative information can be obtained. Of course, this applies only when an initialization or "restart" is needed. The fundamental issue is that the less sensitive a parameter, the larger the perturbation necessary to obtain an appropriate size response in the outputs. This approach could also be applied to an MNR method. Theoretically, the same derivative should be obtained for any sufficiently small perturbation in θ; however, because of both the sometimes widely varying sensitivities of the parameters and the numerical-precision limitations, it is beneficial to vary the perturbation size according to the aforementioned rule. The sensitivity defined as $\theta_j^2 M_{jj}$ was introduced in reference 33 and used again in reference 34 as a means of quantifying the significance of a parameter to the model.

B-2 LEAST SQUARES SOLUTION

The least-squares approach to fitting the surface $Y(\theta)$ offers another advantage if a recursive least-squares method is used. The recursive method provides a memory device reducing the storage requirements from n_p+1 sets of output time histories to just two time histories. One of the two corresponds to the new response predicted by the most recent estimate of θ, and the other corresponds to the outgoing θ that produced the highest cost. The penalty for this advantage is the need to integrate equations of motion twice per iteration; this result still requires substantially less computational effort than that required with the usual MNR method.

When using the recursive least-squares approach, only two changes are made to the MNRES algorithm just described. The first change is in the calculation of the ΔX matrix, and the second change is in the sensitivity calculation. The development of this formulation will begin with equation (4-8) in condensed form (shown below). Everything discussed up to this equation in the previous development applies here.

$$Y_{ki}^{j} = X^{j}S_{ki} \qquad (4\text{-}17)$$

Simplifying the notation by dropping the i subscript and writing the matrix form of the equation (which removes the j superscript) gives

$$Y_{k} = X\, S_{k}. \qquad (4\text{-}18)$$

The least squares solution for the sensitivity vector is

$$S_{k} = \left[X^{T}X\right]^{-1} X^{T}Y_{k} \qquad (4\text{-}19)$$

Now, dropping the k subscript and letting the following apply to the kth element of the output vector Y, a recursive relation can be defined for the r+1 iteration

$$P_{r+1} = \left[X_{r}^{T}\, X_{r}\right]^{-1} \qquad (4\text{-}20)$$

and the updating equation defined as

$$P_{r+1} = \left[P_{r}^{-1} + a_{r}^{T}\, a_{r} - a_{r}^{oT}\, a_{r}^{o}\right]^{-1} \qquad (4\text{-}21)$$

where the row vector a_r is the new set of θ to be included and a_{r}^{o} is the outgoing set of θ which produced the high cost in J. The recursive relation for S, which has already been derived in chapter III-A, is now

$$S_{r+1} = S_{r} - P_{r+1}\left[a_{r}^{oT}\, y_{r}^{o} - a_{r}^{T}\, y_{r} + \left(a_{r}^{T}\, a_{r} - a_{r}^{oT}\, a_{r}^{o}\right)S_{r}\right] \qquad (4\text{-}22)$$

where y_r^o and y_r are the outgoing and incoming elements of Y_k, respectively, at the rth iteration. With the new sensitivities determined, the algorithm proceeds as before.

B-3 PROPERTIES OF MNRES

Properties of MNRES are discussed in comparison with the commonly used MNR algorithm. This approach allows comparison of convergence characteristics and computational advantages and disadvantages against a well-known benchmark.

Convergence of NR or MNR algorithms, both with and without finite-difference derivatives, has been well documented (ref. 13). Convergence of MNRES can be shown, at least heuristically, by considering several details. First, the MNRES method is still fundamentally an NR method or, for this study, an MNR method. The only critical difference is that the derivatives are approximate which makes MNRES closer to MNR with numerically determined derivatives. Second, note that fitting a first-degree, n_p-term polynomial to n_p+1 data points is equivalent to a simple finite-difference method. In effect, as $\Delta\theta^j$ (the distance between points on the fitted surface for MNRES) becomes small enough, the sensitivities become identical to that given by a simple finite-difference method, regardless of the actual functional representation of $Y(\theta)$. The MNRES algorithm simply relaxes the accuracy of the sensitivities in order to reduce substantially the integration requirements; the degree of relaxation varies during the optimization process but can be controlled by limiting step size.

The relaxation of sensitivity accuracy generally appears to be a very beneficial trade-off for Newton-Raphson algorithms. Before considering this relaxation of sensitivity accuracy, note that during an MNRES optimization there are two times that the MNRES scheme computes sensitivities which are very close to that computed by a finite-difference scheme. These times are, first, during the initialization or first pass of the algorithm and, second, toward the end of the optimization process. During

initialization of the algorithm, n_p different θ are chosen (a perturbation on each element of θ is sufficient) giving n_p different response time histories, $Y(\theta)$. The surface given by $Y(\theta)$ is fitted. The initial θ^j can be chosen such that

$$\left| (\theta^j - \theta^0)/\theta^0 \right| \ll 1 \qquad (4\text{-}23)$$

for each j. For the algebraic solution form of MNRES this is completely equivalent to a simple finite-difference scheme and for the least squares form it is a very close approximation. In this study, the same $\Delta\theta^j$ was used in the MNR with finite-difference derivatives as that used in the initialization of MNRES. This was done for comparison purposes; in practice, the choice of perturbation size for θ may be very different, as discussed later. Towards the end of optimization the MNRES scheme again becomes equivalent to a simple finite-difference scheme since the $\Delta\theta$ become very small. This forces the surface that is fit to $Y(\theta)$ to become very small; thus, the slope information is computed for a surface fit to a very small area.

The relaxation of sensitivity accuracy occurs between the two stages discussed above, i.e., after initialization and before convergence. During this part of the optimization large $\Delta\theta$ may occur; this is characteristic of NR, MNR, or MNRES algorithms. For MNRES, unlike NR or MNR, these large steps cause the surface fit area to expand which means the slopes or sensitivities no longer approximate the slope of the $Y(\theta)$ surface at a "point," i.e., no longer approximate the limit requirements of a derivative but rather average the slope over a larger area. This is a critical time period for the MNRES optimization.

Three factors aid in preventing divergence during the critical time period. The first factor is that as the optimization process advances, MNRES continually eliminates values of θ which are far from θ^*, the optimal solution. This, in effect, contracts the expanding surface which is fitting $Y(\theta)$ balancing the expansion process. As the updated estimates of θ get close to θ^* the contraction process will dominate and slope (sensi-

tivity) information will approach that given by a finite-difference scheme. The second factor is that Newton's algorithm and variations like NR, MNR, and MNRES advance more quickly as the quadratic approximation of the cost function improves; moreover, the Newton algorithm converges in one step for a quadratic cost function. Since the quadratic approximation of the cost function will generally improve the closer θ gets to θ^*, and since initial estimates of θ^* are often given by a least-squares procedure or knowledgeable user, θ_0 tends to be "close" to θ^*. Thus, for aircraft estimation problems, MNRES will generally start in a region conducive for convergence. The third factor is that step-size control logic can always be incorporated. Carried to the extreme, MNRES could always be forced to approximate the derivatives the same as a finite-difference method. Of course, convergence would be very slow because of the very small steps. In practice, one would let the algorithm take steps determined by the NR logic (as done in this study); and then if a convergence problem develops, step-size control would be incorporated.

The computational advantage of MNRES is tied to two primary factors. The first factor is the number of unknown parameters, n_p. Both MNR and MNRES must integrate $n_s + n_s n_p$ differential equations on the first iteration; after that, however, MNRES integrates only n_s state equations each pass and MNR continues to integrate n_s state plus $n_s n_p$ sensitivity equations. It appears that as n_p gets larger so does the advantage for MNRES. A limiting factor in MNRES is in equation (4-14) where ΔX must be inverted. This $n_p \times n_p$ matrix becomes more difficult to invert as n_p gets larger and, unfortunately, is made up of very small numbers as the optimization process converges. Also, note that the information gained each pass is not equivalent between the two methods. MNRES performs less computation each pass and, consequently, has less information to update the estimates. However, when sufficient passes are performed to make the work done by MNRES equal to MNR, MNRES will have already performed $n_p + 1$ parameter updates. This allows MNRES to step more quickly towards the final solution. MNRES will achieve convergence faster than MNR as the cost function becomes more quadratic in nature.

The second factor and primary reason for the success of MNRES has to do with the degree to which the cost function can be approximated by a quadratic function. The quadratic approximation is inherent to the Newton type of optimization scheme and, therefore, both MNR and MNRES improve performance as the quality of this approximation improves. However, convergence occurs more quickly with MNRES. This makes sense in light of the way convergence takes place.

Convergence takes place through an iterative process where estimates of the unknowns are updated each iteration. The updates are estimated by equation (5-9) which is the product of the information matrix and the gradient of the cost function. It is well known that convergence can be speeded up by holding the information matrix constant (see e.g. ref. 14) for a limited number of iterations. This eliminates the need to integrate the sensitivity equations for a limited number of iterations; note that integrating the sensitivity equations accounts for the majority of the computational effort. There are two choices, each representing one extreme, for optimization: (1) integrate the sensitivity equations to obtain the most accurate derivative information for each iteration (this is the most costly in terms of computational effort); or (2) hold the information matrix constant for a limited number of iterations (this is the least costly in terms of computational effort and the least desirable since there is no way to know what number of iterations to hold the information matrix constant without causing divergence). A compromise between these extremes is preferred.

MNRES provides this compromise in a very efficient manner. A trade-off between computational effort and sensitivity accuracy is made automatically by MNRES. By using the surface fitting technique only the state equations need be integrated each pass and this information is incorporated in the solution with relatively little computation. The sensitivities are only approximated in this process, however, their accuracy is controlled sufficiently to allow convergence.

The primary disadvantage of using MNRES comes from the computer memory required. n_p+1 sets of output variable time histories must be retained

in this procedure. The recursive least-squares method discussed earlier reduces this problem to 2 sets of time histories; however, the computational effort increases from n_s to $2n_s$ equations to be integrated each pass. The user's computer system would dictate which approach is more appropriate.

Chapter V

CONFIDENCE INTERVAL ESTIMATION

Confidence interval estimation (CIE) is an integral part of the parameter estimation problem. Point estimates of parameters without any qualifications to indicate their accuracy are of little value. An interval estimate which incorporates both variance and confidence level information provides a complete statement of the estimate quality. Although the Cramer-Rao lower bound is commonly used to qualify the ML parameter estimates, it is well known that in aircraft applications these bounds do not accurately reflect the true parameter variance. They are usually too optimistic (ref. 18). The difference between the lower bound and the actual parameter variance can be due to incorrect assumptions about measurement and process noise, bias errors in the estimates, or modelling error. However, the nonlinearity of the estimation problem appears to contribute significantly. In this chapter a method is discussed for determining confidence intervals by analysis of the confidence region contours using a search scheme. In addition, a measure of nonlinearity is developed to further characterize the problem.

A. CONFIDENCE REGION DESCRIPTION

Confidence regions are described by a surface in parameter space representing a certain confidence level. The surface is defined by a statistic which reflects the distribution of error in θ. From the distribution of the statistic, a statement can be made about the probability of the statistic being in a certain interval, I. Assuming the relationship between the statistic and the parameters can be described, a further statement can be made that the parameters are contained in region, R_c, with the same probability. Region R_c reflects the variation in θ as the statistic varies in interval I. The above procedure is the general process by which any confidence interval or region is defined. This definition will obviously vary according to the definition of the statistic. A useful

statistic for composite hypothesis tests is created using the ratio of likelihood functions.

Let Z_1, Z_2, ..., Z_N be N independent random variables with probability density functions

$$f(Z_i, \theta) \quad , \quad i = 1,2,...,N \tag{5-1}$$

The testing hypothesis is formulated as $H_0 : \theta \epsilon \omega$, where ω is a subset of parameter space Ω. Defining the likelihood functions as

$$L(\Omega) = \prod_{i=1}^{N} f(Z_i, \theta) \quad , \quad \theta \epsilon \Omega \tag{5-2}$$

$$L(\omega) = \prod_{i=1}^{N} f(Z_i, \theta) \quad , \quad \theta \epsilon \omega \tag{5-3}$$

and denoting $L(\Omega)$ and $L(\omega)$ as the maxima of the likelihood functions, the likelihood ratio is defined as

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} \tag{5-4}$$

This ratio forms a statistic which has a value between 0 and 1 since the numerator is limited by the H_0 hypothesis. The value of λ reflects the degree to which the H_0 hypothesis is accepted and, therefore, can be used as a statistic to test the hypothesis. If a probability density function can be defined for λ and the relationship between λ and θ can be solved, then a confidence region, R_c, can be defined. With the confidence region determined the confidence intervals (extrema of parameters within the confidence region) can be defined.

The confidence region, R_c, provides an exact description of the parameter error bounds. However, for the general nonlinear estimation problem, an approximation may be involved in defining the confidence level associated with R_c. To resolve this problem, Beale (ref. 21) recommended the statistic for the linear estimation problem be used along with a correction factor to account for moderate nonlinearity in the model. Since this approach for solving the nonlinear problem is based upon a correction to the linear problem, the development will continue with the linear case first.

A-1 CIE FOR THE LINEAR ESTIMATION PROBLEM

The estimation problem is defined to be linear if the model equations are linear in the unknown parameters; the state, input and response variables may or may not appear linearly in the model equaions. The form of the linear estimation model (single output) is given by equation (3-5), repeated here using θ as the vector of unknown parameters (the number of measurements is taken to be N for this discussion)

$$Y = X\theta + \varepsilon \tag{3-5}$$

In this linear regression problem, if Y is $N(X\theta, I\sigma^2)$ and the testing hypotheses are considered as

$$H_0: \quad \theta_t = \theta \tag{5-5}$$

$$H_1: \quad \theta_t \neq \theta \tag{5-6}$$

where the subscript t indicates the true value. It can be shown that the likelihood ratio has the form

$$\lambda = \exp\left\{ -\frac{1}{2\sigma^2} \left[J(\theta) - J(\hat{\theta}) \right] \right\} \tag{5-7}$$

where

$$J(\theta) = (Y - X\theta)^T(Y - X\theta) \qquad (5-8)$$

and

$$J(\hat{\theta}) = (Y - X\hat{\theta})^T(Y - X\hat{\theta}) \qquad (5-9)$$

The statistic λ can be equivalently replaced by

$$\mu = J(\theta) - J(\hat{\theta}) \qquad (5-10)$$

or in practice by the statistic

$$F = \frac{N-n_p}{n_p} \frac{J(\theta)-J(\hat{\theta})}{J(\hat{\theta})} \qquad (5-11)$$

where F is the $1-\alpha_p$ point of the $F(n_p, N-n_p)$ distribution and α_p is the confidence level. This is possible when the model is correct and $J(\theta)-J(\hat{\theta})$ is independent of $J(\theta)$ (ref. 26). In addition, for the linear estimation problem, it is known that (ref. 23):

1. $\hat{\theta}$ is an unbiased estimate of θ.

2. Cramer-Rao bound is reached.

Once the data is determined, $J(\theta)$ is a function of the n_p-dimensional parameter space. In parameter space the function $J(\theta)-J(\hat{\theta})$ can be represented by the contours of a surface. The contours are defined by $J(\theta)$=constant. Again considering the general linear problem (single output)

$$Y = X\theta + \epsilon \qquad (3-5)$$

the cost, $J(\theta)$, can be expanded as

$$J(\theta) = (Y - X\theta)^T (Y - X\theta) \qquad (5\text{-}12)$$

$$= Y^T Y - 2\theta^T X^T Y + \theta^T X^T X\theta \qquad (5\text{-}13)$$

Differentiating and setting to zero, the normal equations are obtained

$$0 = -X^T Y + X^T X\theta \qquad (5\text{-}14)$$

and the solution for θ is

$$\hat{\theta} = \left[X^T X\right]^{-1} X^T Y \qquad (5\text{-}15)$$

The ellipsoidal surface with center at $\hat{\theta}$ is expressed in terms of θ as

$$J(\theta) - J(\hat{\theta}) = \theta^T X^T X\theta - 2\theta^T X^T Y + 2\hat{\theta}^T X^T Y - \hat{\theta}^T X^T X\hat{\theta} \qquad (5\text{-}16)$$

Substituting for $X^T Y$ from the normal equations gives

$$J(\theta) - J(\hat{\theta}) = (\theta - \hat{\theta})^T X^T X (\theta - \hat{\theta}) \qquad (5\text{-}17)$$

which defines an ellipsoidal surface in the n_p-dimensional parameter space. For the linear estimation problem, the contours form an ellipsoidal surface with a single global minimum. The slopes and orientation of the contour depends on the model and data; in addition, they give an indication of parameter correlation and conditioning of the problem. If the contours are greatly elongated (indicating many values of θ give the same cost), an ill-conditioned problem may exist. Inadequate data or possibly overparameterization may be the problem.

With the relationship in equation (5-11), a confidence ellipsoid in n_p-dimensional parameter space with center $\hat{\theta}$ is defined such that the probability is $100(1-\alpha_p)\%$ that the true parameter point θ is contained within the ellipsoid. This can be expressed by substituting equation (5-17) into (5-11)

$$(\theta - \hat{\theta})^T X^T X (\theta - \hat{\theta}) \leq n_p s^2 F_{\alpha_p}(n_p, \text{ N- } n_p) \tag{5-18}$$

where

$$s^2 = J(\hat{\theta})/(N - n_p) \tag{5-19}$$

The confidence limits are determined by realizing the true value of θ lies inside the ellipsoid if and only if it lies between all points of parallel tangent planes to the ellipsoid. Therefore, the true value lies between the two tangent planes orthogonal to vector b (b≠0) if and only if (see ref. 35)

$$\left| b^T(\theta - \hat{\theta}) \right| \leq (b^T H_\alpha^{-1} b)^{1/2} \tag{5-20}$$

where

$$H_\alpha = \left\{ n_p s^2 F_{\alpha_p}(n_p, \text{ N } - n_p) \right\} H \tag{5-21}$$

and H is the Hessian matrix of J. Therefore, the probability is $1-\alpha_p$ that for all b

$$\left| b^T \theta - b^T \hat{\theta} \right| \leq n_p s^2 F_{\alpha_p}(n_p, \text{ N } - n_p) (b^T H^{-1} b)^{1/2} \tag{5-22}$$

This states that the probability is $1-\alpha_p$ that for all θ_i, $i=1,2,\ldots n_p$

$$\left| \theta_i - \hat{\theta}_i \right| \leq \sqrt{k} \, s \, \sqrt{d_{ii}} \tag{5-23}$$

or, expressed in terms of confidence limits, the probability is $1-\alpha_p$ that simultaneously for all θ_i

$$\hat{\theta} - \sqrt{k} \, \sigma_{\hat{\theta}} < \theta < \hat{\theta} + \sqrt{k} \, \sigma_{\hat{\theta}} \qquad (5-24)$$

where

$$k = n_p \, F_{\alpha_p} (n_p, \, N - n_p) \qquad (5-25)$$

$$\hat{\sigma}_{\theta_i} = s \sqrt{d_{ii}} \qquad (5-26)$$

$$H^{-1} = \{d_{ij}\} \qquad (5-27)$$

A-2 CIE FOR THE NONLINEAR ESTIMATION PROBLEM

The nonlinear estimation problem occurs when the unknown parameters appear nonlinearly in the model equations. In the nonlinear problem, several results will change from that found in the linear case (ref. 26):

1. Assuming ε is normally distributed will not imply θ is normally distributed.

2. $s^2 = J(\theta)/(N - n_p)$ is no longer an unbiased estimate of σ^2.

3. There is no covariance matrix in general.

4. F tests and lack of fit tests are not valid in general.

There are some results which remain true, however. These are:

1. The sum of squares, $J(\theta)$, still represents the square of the distance from $(Z_1, Z_2, Z_3, \ldots, Z_N)$ to a point in the estimation space.

2. Minimization of $J(\theta)$ still corresponds to finding a point in estimation space closest to $(Z_1, Z_2, \ldots, Z_N)$.

3. Confidence regions can still be defined; however, the confidence level will be an approximation.

In reference 20 Beale recommends using a correction factor, N_ϕ, as a means to extend the confidence level definition of λ to moderately nonlinear problems. For this case, equation (5-11) becomes

$$J(\theta) - J(\hat{\theta}) = n_p s^2 F_{\alpha_p}(n_p, N - n_p) \left(1 + \frac{N(n_p+2)}{(N-n_p)n_p} N_\phi\right) \qquad (5\text{-}28)$$

and for the multi-output case,

$$J(\theta) - J(\hat{\theta}) = n_p s^2 F_{\alpha_p}(n_p, Nn_o-n_p) \left(1 + \frac{Nn_o(n_p+2)}{(Nn_o-n_p)n_p} N_\phi\right) \qquad (5\text{-}29)$$

where s is now given as

$$s^2 = \frac{J(\hat{\theta})}{(Nn_o-n_p)} \qquad (5\text{-}30)$$

N_ϕ represents a measure of nonlinearity (normalized curvature) of the solution locus near $J(\hat{\theta})$. The method of computing N_ϕ for the multivariable aircraft estimation problem is discussed in the next section.

Determining the confidence contours, defined by equation (5-29), cannot be accomplished analytically as done in the linear case since the contours are not necessarily ellipsoidal. The contours may be very irregular and possibly with several local and global minima. Figure 4 shows the construction of a confidence interval for a one-dimensional problem. The solid and dashed curves in the figure represent nonlinear and linear cases, respectively. In parameter space, the dashed curve would form a symmetric ellipsoidal surface, whereas the solid curve would vary from this shape depending on the degree of nonlinearity. The confidence interval for the nonlinear case is indicated by θ_{min} and θ_{max}; for the linear case, the confidence interval is given by the dashed vertical lines, equidistance from $\hat{\theta}$. The search algorithm used in this study for finding the contour boundaries was presented in reference 24. This method tests a series of randomly selected points in parameter space to determine the position of the confidence region. Through many iterations, the limits of the region

are determined by retaining and updating the points on the boundary which maximize or minimize the unknown parameters. This turns out to be computationally very demanding even for problems with relatively few parameters.

B. NONLINEARITY MEASURE FOR AIRCRAFT APPLICATIONS

The following is a generalization and adaptation of Beale's development (ref. 21) of an intrinsic nonlinearity measure, N_ϕ, to the multivariable problem of airplane parameter estimation. This is an empirical measure of nonlinearity which, in this study, has demonstrated some utility in CIE problems.

If $P(\theta)$ represents the estimation space (or solution locus) in sample space, then $P(\hat{\theta})$ is the point on the solution locus closest to the measurement Z. $\hat{\theta}$ is the point in parameter space which minimizes the cost function. If $T(\theta)$ is defined as a point on a tangent plane at $P(\hat{\theta})$ and W different values are chosen for the vector θ near $\hat{\theta}$ (i.e. θ_w, $w=1,\ldots,W$), then a crude measure of nonlinearity can be written as

$$Q_\psi = \sum_{w=1}^{W} \left| P(\theta_w) - T(\theta_w) \right|^2 \qquad (5\text{-}31)$$

A graphical representation of these quantities for a two dimensional sample space is shown in figure 5. Q_ψ is the sum of squares of the distances from the points $P(\theta_w)$ to the associated points $T(\theta_w)$ on the tangent plane at $P(\hat{\theta})$. Clearly, Q_ψ depends on the number of points, $P(\theta_w)$, and on their distances from $P(\hat{\theta})$. Beale suggests that these distances are proportional to the square of the distance of $P(\theta_w)$ to $P(\hat{\theta})$. If D is defined as the sum of squares of the squared distances then

$$D = \sum_{w=1}^{W} \left| P(\theta_w) - P(\hat{\theta}) \right|^4 \qquad (5\text{-}32)$$

and, thus, Q_ψ is normalized as

$$N_\psi = n_p s^2 Q_\psi / D \qquad (5-33)$$

The values of this measure may still depend on the configuration or orientation of the points $P(\theta_w)$ relative to $P(\hat{\theta})$, but it should not depend significantly on the number of points, $P(\theta_w)$, chosen or on their distances from $P(\hat{\theta})$ (if not too large).

To obtain the "intrinsic nonlinearity", N_ϕ, that Beale recommends, N_ψ must be further restricted. N_ϕ is the value of N_ψ when the parameters are chosen such that $T(\theta)$ is always at the foot of a perpendicular from $P(\theta)$ on to a tangent plane at $P(\hat{\theta})$. In other words, N_ϕ is the minimum value of N_ψ if the model and the experimental design are fixed (see fig. 5).

The practical computation of the intrinsic nonlinearity measure is described in the following development. The sensitivities determined during the estimation process are needed in advance of the following calculations. According to Beale, an empirical estimate of N_ϕ is

$$\hat{N}_\phi = n_p s^2 \hat{Q}_\phi / D \qquad (5-34)$$

where n_p is the number of unknown parameters, s^2 is the sum of squares of residuals. For the multiple output case, s^2 is given by equation (5-30).

The denominator, D, in equation (5-34) can be formulated as

$$D = \sum_{w=1}^{W} \left\{ \sum_{i=1}^{N} \left(\hat{Y}_i(\theta_w) - \hat{Y}_i(\hat{\theta}) \right)^T \hat{R}^{-1} \left(\hat{Y}_i(\theta_w) - \hat{Y}_i(\hat{\theta}) \right) \right\}^2 \qquad (5-35)$$

where i is the time variable up to N data points.

By letting $T(\theta)$ be expressed as a first order Taylor expansion while using the sensitivity information from the estimation, Q_ψ can be computed as

$$Q_\psi = \sum_{w=1}^{W} \sum_{i=1}^{N} \left(\hat{Y}_i(\theta_w) - \hat{Y}_i(\hat{\theta}) - G_i \psi_w \right)^T \hat{R}^{-1} \left(\hat{Y}_i(\theta_w) - \hat{Y}_i(\hat{\theta}) - G_i \psi_w \right) \quad (5\text{-}36)$$

where

$$G_i = \left\{ \frac{\partial Y_k}{\partial \theta_j} \right\}_i \quad (5\text{-}37)$$

$$\psi_w = \theta_w - \hat{\theta} \quad (5\text{-}38)$$

Rewriting Q_ψ and letting

$$\delta \hat{Y}_i = \hat{Y}_i(\theta_w) - \hat{Y}_i(\hat{\theta}) \quad (5\text{-}39)$$

then

$$Q_\psi = \sum_{w=1}^{W} \sum_{i=1}^{N} \left(\delta Y_i - G_i \psi_w \right)^T \hat{R}^{-1} \left(\delta Y_i - G_i \psi_w \right) \quad (5\text{-}40)$$

This is now in the form of a standard least squares problem. The problem is to find the value of ψ which minimizes Q_ψ; that is, minimize the distance (see fig. 5) given as

$$d = \left| P(\theta_w) - T(\theta) \right|^2 \quad (5\text{-}42)$$

Therefore, the value of ψ which minimizes Q_ψ is Φ given by W sets of least squares minimizations:

$$\text{for } w=1,2,\ldots,W: \quad \Phi_w = \left| \sum_{i=1}^{N} G_i^T \hat{R}^{-1} G_i \right|^{-1} \sum_{i=1}^{N} G_i^T \hat{R}^{-1} \delta Y_i \quad (5\text{-}43)$$

Thus

$$\hat{Q}_\phi = \sum_{w=1}^{W} \sum_{i=1}^{N} \left(\delta Y_i - G_i \Phi_w\right)^T \hat{R}^{-1} \left(\delta Y_i - G_i \Phi_w\right) \tag{5-44}$$

For application to the parameter error bound problem for aircraft, the following assumptions were made:

1. Selecting $W = 2n_p$ is sufficient to adequately sample the local surface of $\hat{J}(\theta)$ near $\hat{\theta}$.

2. Selecting $\Delta\theta$ as

$$\left|\Delta\theta_i\right| = \hat{\sigma}_{\theta_i} \tag{5-45}$$

   provides a reasonable distance from θ to sample the surface $J(\theta)$. The goal is to detect the nonlinearity from the tangent plane without going too far into the nonlinear range where the curvature (based on sensitivity information at $\hat{\theta}$) no longer applies.

These assumptions represent a proposal for a unified approach in computing N_ϕ. Using this approach, results of various studies can be compared and the differences between confidence limits based on Cramer-Rao bounds and random search can be examined. N_ϕ and the error bounds determined by random search can also indicate the effect of experimental design, especially input form and model error, on identifiability.

Chapter VI

APPLICATION TO SIMULATED AND REAL DATA

The examples considered in this chapter demonstrate the ML/MNRES algorithm for estimating parameters and the search technique for estimating confidence intervals of the parameters. These methods are compared against commonly used techniques to provide a benchmark for comparison. The commonly used techniques are the ML/MNR for parameter estimation and the Cramer-Rao bounds for confidence interval determination. ML/MNR is used with both analytically- and numerically-determined derivatives. The Cramer-Rao bounds, taken from the information matrix, are adjusted to the 95 percent confidence level.

Only dynamical systems or airplane estimation problems are used in this study rather than classical test problems such as Rosenbrock's function (ref. 13). Using classical optimization problems, which usually require very little computational time to evaluate the cost function, could lead to different conclusions about the algorithms. For aircraft estimation problems, the bulk of computer time is spent performing integrations of the state and sensitivity equations. To prevent any bias in the results due to variations in programming efficiency or integration techniques, only estimation algorithms using the same integration method are compared.

The performance of the methods used in this report will be evaluated with the following criteria:

1. Accuracy of estimates
2. CPU time to termination

Termination is obtained when parameter and cost function fractional changes are computed to be within a specified precision. Both cost function change, $\Delta J/J$, and parameter change, $\Delta\theta/\theta$, are required to be satisfied simultaneously to prevent premature termination on a plateau where $\Delta J \ll 1$ and $\Delta\theta$ is relatively large or on a steep slope where $\Delta\theta \ll 1$ and ΔJ is relatively large.

Besides estimate accuracy and CPU time to termination, an additional observation provided is the number of "equivalent evaluations." One unit of this measure is the amount of calculation required to integrate the system equations and to evaluate time histories of the output variables. Each method described in this report requires a different number of equivalent evaluations to make one update in the parameter estimates. This measure provides an indication of how efficiently information gained from system integrations is utilized. System integrations are the primary computational burden for any estimation method applied to dynamical systems.

The examples in this study use both simulated data (examples I-III) and flight data (examples IV-VI). Except for the first two examples, the parameters estimated are the nondimensional aircraft stability and control derivatives conforming to standard NASA notation. For examples I and II a simple linear system is integrated with an Euler integration method. Examples III-VI involve the airplane problem and use the general equations of motion given by equations (2-5) to (2-15). These equations are integrated with a fourth-order Runge-Kutta integration scheme. For comparison purposes, the same integration step size and the same computer (CYBER 175) are used in each example.

For the airplane examples, the ML/MNRES algorithm is applied through program MAX. MAX is a very modular FORTRAN 5 code with dynamic memory. The modular format allows aerodynamic models or entire system models to be changed easily. The dynamic memory capability adjusts core memory automatically to the dimensions required. A block diagram of the general computing scheme for ML/MNRES is given in figure 6. A block diagram for program MAX is given in figure 7 followed by table 1 defining the elements in figure 7.

A. SIMULATED DATA STUDIES

Simulated data offers three advantages for testing a new estimation algorithm. The first and most important advantage is that the true values

of the parameters are known; the second advantage is that the problem can be well posed and defined without modelling error; and third, the degree of complexity can be selected. In this study, the first two of three simulation examples use a single-input/double-output, linear second-order system,

$$\dot{X}_s = AX_s + BU, \quad X_s(0) = 0 \qquad (6\text{-}1)$$

$$Y = X_s \qquad (6\text{-}2)$$

These examples are used to demonstrate the relative speeds and accuracy of the sample estimation algorithms and to initially indicate the preferred methods. The third simulation, uses a nonlinear aircraft model and simulates varying levels of measurement noise found in real flight data. This example demonstrates the accuracy of program MAX.

A-1 EXAMPLE I

Example I demonstrates and compares the FPS, MNR, and MNRES optimization schemes in a simple ML estimation problem. The MNR method uses a finite difference method to compute derivatives. This satisfies one requirement of this study, which is to eliminate the need to derive analytical gradients. MNR generally performs with about the same speed using either numerically-determined derivatives or integrated sensitivity equations (ref. 15). MNRES uses the same finite difference method as MNR to determine sensitivities during initialization; however, MNRES uses the recursive least squares form of the algorithm during optimization. Because of the small memory requirements to store time histories in this example, only one integration per pass is performed. The purpose of this example is to compare the relative performance of each method on a problem involving a simple dynamic system.

Example I has six unknown parameters. The six unknown parameters in equation (6-1) are the four elements of the 2x2 system matrix, A, and two elements of the control input matrix, B. The input form was chosen as

$$u = \begin{cases} \sin t \; , & 0 < t < 2\pi \\ \\ 0 \quad , & t > 2\pi \end{cases}$$

and data points were generated every .25 seconds. Process and measurement noise were excluded for this example.

Table II shows the true values, initial values, and final estimated values of the six unknown parameters. Figure 8 shows the input and response time histories. All three algorithms accurately converged to the true parameter values using only the first 5 seconds of data. The MNR method was 30 times faster than FPS and MNRES was twice as fast as MNR or 60 times faster than FPS. The number of equivalent evaluations had similar ratios, i.e. 715:28:12.

Table II shows clearly that optimization problems having reasonable starting values and involving time consuming cost function evaluations should not be solved with direct search schemes, such as FPS. This result is supported by an independent study using aircraft estimation problems in reference 29. Reasonable initial values tend to provide a more quadratic-like cost function for which Newton's method is most effective. If reasonable initial values are not available, the FPS may be more attractive. In light of the results of Example I, further study concentrated only on the gradient methods.

A-2 EXAMPLE II

Example II is provided to demonstrate the robustness of MNRES compared with the commonly used MNR. The more common form of MNR with analytically-derived sensitivity equations is used to prevent any deterioration of the algorithm due to approximating the sensitivities. The system from example I is analyzed again except measurement noise is added and a pulse input is used to excite the system. Two cases are considered, each with different levels of measurement noise. The noise is zero mean and Gaussian with standard errors of 0.0001 and 0.001 for cases 1 and 2, respectively. Figure 9 shows time histories of the input and response variables for the two cases. Table III shows the estimation results.

In case 1 both methods produce equally degraded results; however, MNRES still converged to the same precision level more quickly. In case 2, with a severe noise level and limiting the information to 3 seconds of data, MNRES was unable to converge. The results showed that it was oscillating about a solution, unable to find a new parameter vector which would produce a lower cost. The MNRES used on this example had no special step-size control logic. The solution that was obtained, however, was as accurate as that obtained by MNR, which did converge.

Meeting convergence requirements does not guarantee accurate results; the error in the estimates ranged from 5 percent error to 130 percent error. MNR had both the most accurate and the least accurate estimate. The importance (sensitivity) of a parameter to the model will significantly affect the accuracy of the estimate, particularly under these adverse conditions. Based on these examples, it appears that MNRES is computationally more efficient than MNR while providing the same level of accuracy.

A-3 EXAMPLE III

In this example, the accuracy and robustness of ML/MNRES are demonstrated by application to a nonlinear aircraft simulation with known measurement noise levels. In addition, program MAX is validated. For this example and all other aircraft examples, the computationally least demanding form of MNRES is used to compute sensitivities. This form uses the linear surface fit with equation (4-14) instead of the recursive least squares form with equation (4-22). The simulation involves a nonlinear lateral model of a general aviation aircraft.

Three cases are considered: case 1 without any measurement noise; case 2 with a representative noise level typical of flight data for the aircraft; and, case 3 with twice the noise level of case 2. The standard errors of the simulated measurement noise are shown in table IV. In each case, the noise is zero mean and Gaussian. The simulated data was created by a fourth-order Runge-Kutta integration with a step size of .05 sec. Table V shows the terms used in the nonlinear aerodynamic model to create the simulation and the parameter estimates obtained through analysis of the

simulated data. Time histories are provided for the three cases in figure 10. The control inputs were the same for all three cases and are shown in figure 11.

Program MAX was applied using two convergence criteria: $\Delta J/J \leq 1.E-03$ and $\Delta\theta/\theta \leq 1.E-03$. As expected, the estimates of the less easily identified nonlinear terms, such as $C_{n_{\alpha_r}}$ and $C_{\ell_{\alpha_p}}$, are more quickly corrupted as the noise levels increase; however, the estimates are still very reasonable and the time histories are accurately predicted. Table V shows that the MNRES method can be used effectively in estimating parameters for nonlinear aircraft systems.

B. REAL FLIGHT DATA STUDIES

In this section three examples are considered. In each example the model structure and initial parameter estimates were determined using the MSR program of reference 11. For the first two examples, the parameter estimation problem is solved by using two different ML programs. The first is program MAX which uses the ML/MNRES algorithm as described in example III. The second is program MAXLIK which uses an ML/MNR algorithm. This MNR algorithm integrates analytically derived sensitivity equations to obtain sensitivities. MAXLIK is a proven code for aircraft-parameter estimation documented in reference 31. In the last example, program MAX is used to compute parameter estimates and Cramer-Rao bounds. These bounds are adjusted to the 95 percent confidence level and compared with that obtained using the search method.

For comparison purposes, both program MAX and MAXLIK use a fourth-order Runge-Kutta integration method with the same integration step size (.05 sec in example IV and .04 sec in examples V and VI). A convergence criterion is set at $\Delta J/J = .001$ for both codes. Program MAX normally uses an additional criterion restricting the parameter change $\Delta\theta/\theta$; however, it is removed in these examples to ensure that both programs converge for the same criterion. Both programs use the same bias and scale-factor corrections to the flight data for each example. These corrections were determined by using a compatibility program developed in reference 34. The same initial parameter values are used by both MAX and MAXLIK.

B-1 EXAMPLE IV

Example IV uses flight data from a general aviation aircraft operating at an angle of attack of 8°. The estimation problem involves the nonlinear lateral model. Table VI presents a comparison between parameter estimates and Cramer-Rao bounds from both MAX and MAXLIK. Initial values and sensitivities computed as $\theta^2_j M_{jj}$ are also given. Again, there is reasonable agreement between the two approaches. Cramer-Rao bound estimates tend to be a little higher for program MAX. This is probably due to their sensitivity to the derivative information.

Repeating the calculations with program MAX, by allowing the sensitivity ratios to be incorporated into the initializing derivative calculations, provided a small improvement in the overall speed of the algorithm. This occurred because only one restart was required during the optimization process. More improvement would be realized in problems where restarting occurs several times. Time histories of the measured flight data and predicted response using the estimated model are shown in figure 12. Execution times for example IV are 793 seconds for program MAX and 1036 seconds for program MAXLIK.

B-2 EXAMPLE V

This example uses flight data from an advanced twin engine fighter operating at an angle of attack of 6°. A nonlinear longitudinal model is used. Table VII presents a comparison of parameter estimates and their CR bounds for both MAXLIK and MAX (case 1). Also shown for each program is the time to reach convergence expressed in seconds. Program MAX converged very close to the same values as program MAXLIK except processing was done in one fourth the time. This example reflects the effectiveness of MNRES under fortunate conditions (that is, where the cost is well approximated by a quadratic and a moderate number of unknowns (11 parameters) are determined). The quadratic nature of the cost is indicated by a very small value of N_ϕ. The value of N_ϕ was .003 for this example. The mean value estimates of MAX are very good and the CR bounds are good but tend to indicate a slightly larger error bound than MAXLIK.

Although the Fisher information matrix is updated each iteration by both programs, program MAX delays updating the weighting matrix, R^{-1}, until convergence is achieved. The example is solved once more by program MAX and the weighting matrix is updated twice. The results are shown in table VII, MAX (case 2). The mean values are essentially the same since they are independent of the weighting matrix used, except possibly through some numerical errors. The standard errors are slightly closer to the MAXLIK results and further updating brings only very small improvements. These estimates were obtained by MAX in about 40 percent of the MAXLIK processing time.

B-3 EXAMPLE VI

The sixth example uses flight data from an advanced single engine fighter operating at an angle of attack of 4°. This example involves a nonlinear lateral model. In this example, 95 percent confidence intervals are estimated using two approaches. One approach is based on the CR bound using program MAX and the other on a random search technique. The 95 percent confidence intervals determined by each approach are presented in table VIII. In this example N_ϕ was found to be .02; however, it was set to zero for the interval computations. Even with this correction, only a very small increase in interval size would be obtained. The confidence intervals determined by the search method are significantly larger than the corresponding CR estimates and indicate an asymmetric confidence interval.

C. DISCUSSION OF RESULTS

The general experience with ML/MNRES and the examples chosen for this study indicate that ML/MNRES will perform better than ML/MNR for estimation problems involving dynamic systems such as aircraft systems. In general, MNRES should perform well in any problems for which the Newton-Raphson family of methods is appropriate (that is, where the cost can be reasonably approximated by a quadratic). The results of this study also indicate that a search technique is needed to accurately assess the parameter error bounds in the nonlinear estimation problem; for the linear problem or problems with very little nonlinearity, the Cramer-Rao bounds should agree with values determined by the search technique.

- 63 -

C-1 PARAMETER ESTIMATION

Two important goals in this study were to develop an estimation algorithm which, first, eliminated the requirement to reformulate the algorithm for each new model and, second, provided a more efficient method to deal with the computationally more burdensome nonlinear problems. The first goal has been surpassed through the ML/MNRES algorithm coded in program MAX in two respects: (1) it does not require the derivation of sensitivity equations to complete the formulation of the algorithm and the modular form of MAX allows easy application to any system; and (2) it provides many computationally efficient options to the user as to how the sensitivities will be approximated (accuracy and order of derivatives, also options for memory requirements) and these options are readily incorporated because of the modular format. The second goal has been demonstrated in the examples but further discussion will clarify the conditions under which this goal has been met.

The first two examples use a simulated linear system with and without noise. This system is readily identifiable except when severe noise and nonoptimal inputs are included. The initial values were relatively close to the final solution and so allowed a good quadratic approximation of the cost function thus providing a condition conducive to convergence. The two Newton-Raphson methods, MNR and MNRES, were substantially faster than the search method as expected under these conditions. MNRES, however, capitalized more efficiently than MNR on the information obtained from each integration of the system equations. Each integration of the system equations provides information which is immediately incorporated into the numerical process when using MNRES. When using MNR, n_p+1 system integrations (equivalent evaluations) are required before each updating operation, MNRES requires only one; for example I, the ratio of equivalent evaluations was 28:12. MNRES made much more efficient use of the system integrations. The results indicated both MNR and MNRES to be very fast relative to the search technique, thus search methods were eliminated from any further study. The MNRES approach was a little more than twice as fast as MNR.

The third example demonstrated that the ML/MNRES algorithm was a viable method for a nonlinear aircraft estimation problem with both realistic and heavy noise levels. This example provides confidence in the ML/MNRES approach. However, since it is a simulated example, it cannot be accepted as conclusive. Simulations always provide optimal conditions for estimation algorithms since problems such as modelling error, bias errors, unknown noise spectra and general data incompatibility are not present.

Unlike simulated data, real flight data often present problems (as just mentioned) for any estimation method; these problems may slow the convergence process or even stop it. The last three examples consider real flight data; these examples were specially selected to reflect differing levels of difficulty for the estimation algorithms. Examples IV and VI compared with example V demonstrate a representative range in the degree of difficulty (measured by computational effort) for the algorithms and, unsurprisingly, the degree of nonlinearity (measured by N_ϕ) is also largely varying. N_ϕ differed by an order of magnitude between example V and VI (VI being more nonlinear). ML/MNRES was again faster than the benchmark program, ML/MNR. For the more nonlinear examples, convergence time for ML/MNRES was 70 to 80 percent of the time required for ML/MNR; in the less difficult problem, ML/MNRES required only 40 percent of the benchmark time. These examples give some credence to the superiority of ML/MNRES. However, they also indicate a large variability in the superiority.

The variability appears to be that as the degree of nonlinearity increases, the methods approach the same speed of convergence. The computational advantage of ML/MNRES tends to be reduced as the nonlinearity increases. A moderate number of unknowns are used in each case (examples V and VI) so the advantage due to the difference in $n_s + n_s n_p$ integrations per pass and n_s integrations per pass is probably a small factor (see IV-B). The main factor, however, is the quality of the quadratic approximation of the cost function which, of course, is directly related to the degree of nonlinearity of the cost. Both methods are slowed as the nonlinearity increases or the quadratic approximation becomes poorer, but the MNRES method is much more dependent on the quality of the quadratic

approximation since in effect it is an approximation of the MNR method. As the nonlinearity increases, MNRES looses its advantage over MNR.

Figure 13 offers a graphical view of the performance of MNR and MNRES in example V, where the quadratic approximation is fairly good. The graphs show the value of the convergence criterion versus CPU time. Program MAXLIK using MNR follows a typical convergence pattern; the small oscillations before convergence are due to the updating of the weighting matrix, R, each pass after the criterion falls below .01 in value. This approach with MNR has been found to be beneficial in these problems. Program MAX, on the other hand, updates the information matrix each pass but holds the weighting matrix constant until the first convergence is achieved. At this point the final parameter estimates are obtained, however, the Cramer-Rao bounds (determined from the information matrix) are not converged. The information matrix, M, will not converge until the weighting matrix, R^{-1}, is updated sufficiently. This is understandable since the parameter estimates are asymptotically independent of the weighting matrix whereas the Cramer-Rao bounds depend on the information matrix which in turn depends on the value of R. Therefore, two more cycles are made to convergence ensuring that the weighting matrix has stabilized. Note that the first step in each method takes the same amount of time and achieves the same reduction in cost; this is expected since initializing MNRES requires the same computations as the first pass in MNR.

A key feature of ML/MNRES is in the method of updating the information matrix. Although both ML/MNRES and ML/MNR update the information matrix each pass, they each do it quite differently. One well known way to improve the speed of techniques which involve the Hessian matrix, or approximations to it, is to hold the information matrix, M, constant for one or more iterations. This reduces the amount of integration required per pass to be the same as in MNRES since no sensitivity equations are integrated. How many iterations M can be held constant is unknown and unknowable in advance. So there are two alternatives generally known and used. One is to integrate the sensitivity equations each pass requiring maximum computational effort but giving maximum accuracy to the optimization process. The other alternative is to skip integrating the sensitivity

equations for a necessarily infrequent number of iterations to hopefully increase convergence speed. In methods like the ones considered in this study where large steps in the optimization process may occur, there is a lot of danger in not updating the information matrix. ML/MNRES finds an effective compromise between the alternatives discussed above.

The compromise is achieved by updating the information matrix each pass. However, only the information obtained from integrating the equations of motion once each pass is incorporated. Thus, a compromise is achieved where updating is occurring at minimal computational cost. Because of the limited information to update the information matrix, a suboptimal but computationally more efficient path is followed to convergence. The result is that ML/MNRES requires many more passes to reach convergence, but each pass requires much less computational effort than ML/MNR. The net result is much faster convergence depending on the degree of nonlinearity of the cost and the quality of the quadratic approximation used by the method.

C-2 CONFIDENCE INTERVAL ESTIMATION

Confidence intervals obtained in example V were found to be very close to the CR bounds adjusted to the 95 percent confidence level. In addition, the value of N_ϕ was very small and convergence occurred relatively quickly. This indicates that the cost function was very well approximated by a quadratic function. In analogy to figure 4, the construction of the confidence intervals for example V would place the dashed and solid lines virtually on top of each other at the error level selected. The quality of the quadratic approximation is confirmed by the very fast convergence of MNRES relative to MNR.

Confidence intervals obtained in the sixth example were found to be 5 to 10 times the size of the CR bound adjusted to the 95 percent confidence level (Table VIII). This is in agreement with references 18 and 23 on the values generally found in analyzing actual flight data. Reference 23 used the search technique and reference 18 estimated the confidence intervals by computing an ensemble average. This also indicates that this cost function

is much more nonlinear than that for the first example, and it is not well approximated by a quadratic function at the error level considered. This is confirmed by the relative speed of convergence between MNRES and MNR. MNRES, with two updates of R, required 85 percent of the time MNR required.

It appears from the examples considered that a primary factor in determining the confidence interval size for airplane stability and control derivatives is the degree of nonlinearity of the cost function. Other factors, such as bias errors, modelling errors, noise level and noise spectra, may contribute directly to confidence interval size or may manifest themselves as part of the nonlinearity of the cost function. In this study, the other factors were not tested to determine their contribution.

At present, the random search technique is the only method to determine confidence bounds accurately. Clearly, the CR bounds, which are tied to the quadratic approximation inherent in the information matrix, will always be different from the parameter variance. This difference will mainly depend on the quality of the quadratic approximation. The only disadvantage of the search technique is its relatively poor convergence rate combined with the large computational effort required in this type of problem. Although Beale's measure of nonlinearity, N_ϕ, was designed to correct the confidence level in the CIE problem, there seems to be more utility in considering N_ϕ (or some similar measure) as a way to discern if the lengthy computations of the random search are needed. If very little nonlinearity exists, the user can be reasonably confident that the Cramer-Rao bounds provide accurate error bound information.

Chapter VII

SUMMARY AND CONCLUDING REMARKS

A. SUMMARY

An algorithm for maximum likelihood (ML) estimation is developed with an efficient method for approximating the sensitivities. The algorithm is applicable to most parameter estimation problems and is particularly suited for nonlinear, multivariable, dynamic systems. The ML algorithm relies on a new optimization method closely related to a modified Newton-Raphson (MNR) technique; the new method is referred to as a modified Newton-Raphson with estimated sensitivities (MNRES).

MNRES determines sensitivities by using slope information from local surface approximations of each output variable in parameter space. The fitted surface allows sensitivity information to be updated at each iteration with a significant reduction in computational effort. MNRES determines the sensitivities with less computational effort than using either a finite-difference method or integrating the analytically-determined sensitivity equations. The choice of the type of surface (for example, nth-order polynomial or spline, etc.) and the method of fitting the surface (for example, least squares or simply solving simultaneous equations) is made by the user to suit the particular need. MNRES eliminates the need to derive sensitivity equations for each new model, thus eliminating algorithm reformulation with each new model and providing flexibility to use model equations in any format that is convenient.

Two surface-fitting methods are discussed and demonstrated, while other possibilities are indicated. Comparisons are made between MNRES and other commonly used optimization methods such as a search method called the flexible polyhedron search (FPS) and a gradient method called the modified Newton-Raphson method. Several sample problems are solved to compare the techniques. Simple linear systems are used at first, and then nonlinear aircraft estimation problems are solved by using both real and simulated data. MNRES is found to be equally accurate and substantially faster than

the commonly used techniques. The reduction in computational effort provided by MNRES is dependent on several factors: the choice of surface-fitting method; the number of unknown parameters; data quality; accuracy of the sensitivity calculations; and, particularly, the degree of nonlinearity of the cost function.

A search technique for determining the confidence limits of ML parameter estimates is applied to nonlinear estimation problems for airplanes. The confidence intervals obtained by the search are compared with Cramer-Rao (CR) bounds at the same confidence level. It is observed that the degree of nonlinearity of the cost function is an important factor in the relationship between CR bounds and the error bounds determined by the search technique. The CR bounds were found to be close to the bounds determined by the search when the degree of nonlinearity was small. The CR bounds were 5 to 10 times too conservative (too small) when the nonlinearity was significant. Beale's measure of nonlinearity is developed in this study for airplane identification problems; it is used to empirically correct confidence levels for the parameter confidence limits. The primary utility of the measure, however, was found to be in predicting the degree of agreement between Cramer-Rao bounds and search estimates.

B. CONCLUDING REMARKS

The primary contribution in this study is an efficient maximum likelihood estimation algorithm. However, inherent in this study is a suggested methodology for solving the nonlinear airplane identification problem. The use of a modified stepwise regression in conjunction with several testing criteria is suggested to determine the airplane aerodynamic model structure. This very efficient scheme was developed in a prior study to handle the widely varying model structure in nonlinear flight regimes. A maximum likelihood scheme (ML/MNRES), developed in this study, is then recommended to obtain optimal parameter estimates. This method is more efficient than other commonly used techniques in airplane estimation problems and provides some practical computing options. Finally, a random search procedure is

required to determine parameter confidence limits for the nonlinear case. This is used in conjunction with Beale's measure of nonlinearity (adapted to the airplane problem) to make an empirical correction to the confidence level. It is also used to determine if the extensive calculations of the random search are needed to estimate confidence limits.

The new optimization algorithm, MNRES, has three advantages over other commonly used techniques. The first advantage is that the algorithm removes the need to derive sensitivity equations for each new model; this eliminates the computational burden of integrating the sensitivity equations during each iteration of the algorithm. This also provides much flexibility, allowing the model equations to be in any format that is convenient - such as splines, polynomials, or a nonanalytic form. Also the quickly varying model structure sometimes found in the nonlinear regimes is readily handled. The second advantage is that the algorithm is effective for a variety of surface fitting methods chosen to fit the output vector surface in parameter space (needed for sensitivity estimation); this allows the user to choose a surface-fitting method best suited to the problem. An approach is discussed which reduces storage requirements with little additional computation. The third advantage of the algorithm is that it reduces the computational effort in comparison with the commonly used modified Newton-Raphson (MNR) method. For small problems (fewer than 15 parameters to be estimated), the reduction can be substantial. For larger nonlinear problems, the reduction may be more modest; however, improvements may still be significant if data quality, signal compatibility, and sensitivity calculations are good. Even though the application of interest for this study was an aircraft operating in nonlinear flight regimes, the approach should be effective for many other nonlinear, dynamic systems. Based on this study, the ML/MNRES algorithm generally performs better and offers more versatility than the commonly used ML/MNR algorithm.

The suggested methodology recommends a random search technique to obtain parameter confidence limits for the maximum likelihood estimates. Since the nonlinear problem does not lend itself to an explicit analytical solution, the search uses a random sampling algorithm to find the confidence limits; unfortunately, this method is computationally demanding,

particularly for cases with a large number of unknowns. Unless sufficient repeated measurements are available, it is the only method to accurately determine confidence region boundaries in the nonlinear case. Beale's measure of nonlinearity is used to provide an empirical correction to the confidence level used by the search. Although this was Beale's intended use, it has little affect on the confidence limits for airplane applications. However, it was shown that the degree of nonlinearity and the degree to which the Cramer-Rao bounds and the random search estimates agree is closely related. Therefore, it is recommended to use this or some similar measure to determine the necessity of the search calculations.

If further studies are made with MNRES, it should prove beneficial to use more efficient inversion schemes than the standard Gaussian elimination used in this study. This may improve the algorithm for larger numbers of unknowns. Also, further consideration should be given to defining the relationship confidence intervals and the nonlinearity of the cost function have with other factors such as bias errors, modelling errors, input form, and noise spectra. In addition, more investigation into measures of nonlinearity and their best computing schemes would be advantageous. Nonlinearity measures may be useful for reflecting the quality of the experiment since parameter error bounds will vary with model error and optimality of the input form. Finally, significant computational savings would be achieved if the confidence limits for the nonlinear estimation problem could be determined using gradient techniques rather than the computationally demanding search scheme used in this study.

REFERENCES

1. Aström, K. J.; and Ekhoff, P.: System Identification - A Survey. Automatica, Vol. 7, 1971.

2. Greenberg, H.: A Survey of Methods for Determining Stability Parameters of an Airplane from Dynamic Flight Measurements. NACA TN-2340, 1951.

3. Shinbrot, M.: On the Analysis of Linear and Nonlinear Dynamical Systems from Transient-Response Data. NACA TN-3288, 1954.

4. Shinbrot, M.: A Least Squares Curve Fitting Method with Applications to the Calculation of Stability Coefficients from Transient Response Data. NACA TN-2341, 1951.

5. Rediess, H. A.: An Overview of Parameter Estimation Techniques and Applications in Aircraft Flight Testing. NASA TN D-7647, April 1974.

6. Larson, D. B.: Identification of Parameters by Method of Quasi-linearization. CAL Rep. No. 164, Cornell Aero. Lab., Inc., May 14, 1968.

7. Taylor, L. W., Jr.; and Iliff, K. W.: A Modified Newton-Raphson Method for Obtaining Stability Derivatives from Flight Data. Computing Methods in Optimization Problems-2, L.A. Ladeh, L. W. Newstadt, A. V. Balakrishnan, eds., Academic Press, 1969.

8. Mehra, R. K.: Maximum Likelihood Identification of Aircraft Parameters. 1970 JACC, Atlanta, GA, 1970.

9. Stephen, D. E.; Mehra, R. K.: Maximum-Likelihood Identification and Optimal Input Design for Identifying Aircraft Stability and Control Derivatives. NASA CR-2200, March 1973.

10. Rault, A.: Identification Applications to Aeronautics. Proceedings of the 3rd IFAC Symposium, June 1973, pp. 49-65.

11. Klein, V.; Batterson, J. G.; and Murphy, P. C.: Determination of Airplane Model Structure from Flight Data by Using Modified Stepwise Regression. NASA TP-1916, October 1981.

12. Klein, V: Identification Evaluation Methods. Parameter Identification, AGARD-LS-104, November 1979, pp. 2-1 through 2-21.

13. Himmelblau, D. M.: Applied Nonlinear Programming. McGraw Hill Book Co., c.1972.

14. Gupta, N. K.; and Mehra, R. K.: Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculation. IEEE Transactions on Automatic Control, Vol. AC-19, No. 6, December 1974.

15. Trankle, T. L.; Vincent, J. H.; and Franklin, S. N.: Recent Methods for Nonlinear System Identification. AIAA Paper No. 81-1865, August 1981.

16. Trankle, T. L; Vincent, J. H.; and Franklin, S. N.: System Identification of Nonlinear Aerodynamic Models. Advances in the Techniques and Technology of the Application of Nonlinear Filters and Kalman Filters, AGARD-AG-256, March 1982, pp. 7-1 through 7-26.

17. Ralston, Mary L.; and Jennrich, Robert I.: Dud, A Derivative-Free Algorithm for Nonlinear Least Squares. Technometrics, Vol. 20, No. 1, February 1978, pp. 7-13.

18. Klein, V.: Determination of Stability and Control Parameters of a Light Airplane from Flight Data Using Two Estimation Methods. NASA TP-1306, March 1979.

19. Balakrishnan, A. V.; and Main, R. E.: Improvements in Aircraft Extraction Programs. NASA CR-145090, 1976.

20. Maine, R. E.; and Iliff, K. W.: The Theory and Practice of Extracting the Accuracy of Dynamic Flight-Determined Coefficients. NASA RP-1077, July 1981.

21. Beale, E. M.: Confidence Regions in Nonlinear Estimation. Research Methods Meeting, Admiralty Research Laboratory, December 1979.

22. Bates, D. M.; and Watts, D. G.: Relative Corrective Measures of Nonlinearity. J. R. Student Soc. B., Vol. 42, No. 1, 1980.

23. Mereau, P.; and Prevost, G.: Calculation of Uncertainty Intervals in Nonlinear Estimation. 5th IFAC Symposium, Darmstadt, Germany, September 1979, pp. 365-372.

24. Mereau, P.; and Raymond, J.: Computation of the Intervals of Uncertainties About the Parameters Found for Identification. NASA TM-76978, August 1980.

25. Klein, V.; Batterson, J. G.; and Smith, P. L.: On the Determination of Airplane Model Structure from Flight Data. Identification and System Parameter Estimation - Sixth IFAC Symposium, Volume 2, George A. Bekey and George N. Saridis, eds., Pergamon Press, Inc., c.1982, pp. 1034-1039.

26. Draper, N. R.; and Smith, H.: Applied Regression Analysis. John Wiley & Sons, Inc., c.1981.

27. Bard, Jonathan: Nonlinear Parameter Estimation. Academic Press, c.1974.

28. Carmer, H.: Mathematical Methods of Statistics. Princeton University Press, New Jersey, 1946.

29. Tategronkar, R.; and Plaetschke, E.: Nonlinear Parameter Estimation from Flight Test Data Using Minimum Search Methods. DFVLR, Forschungsbericht 83-15, 1983.

30. Olsson, D. M.; and Nelson, L. S.: The Nelder-Mead Simplex Procedure for Function Minimization. Technometrics, Vol. 17, No. 1, February 1975.

31. Grove, R. D; Bowles, R. L; and Mayhew, S. C.: A Procedure for Estimating Stability and Control Parameters from Flight Test Data by Using Maximum Likelihood Methods Employing A Real-Time Digital System. NASA TN D-6735, 1972.

32. Adams, W. M.; and Tiffany, S. H.: Design of a Candidate Flutter Suppression Control Law for DAST ARW-2. AIAA Guidance and Control Conference, August 1983.

33. Klein, V.: On the Adequate Model for Aircraft Parameter Estimation. Report Aero No. 28, Cranfield Institute of Technology, March 1975.

34. Klein, V.; and Schiess, J. R.: Compatibility Check of Measured Aircraft Responses Using Kinematic Equations and Extended Kalman Filter. NASA TN D-8514, August 1977.

35. Schweppe, F. C.: Uncertain Dynamic Systems. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1973.
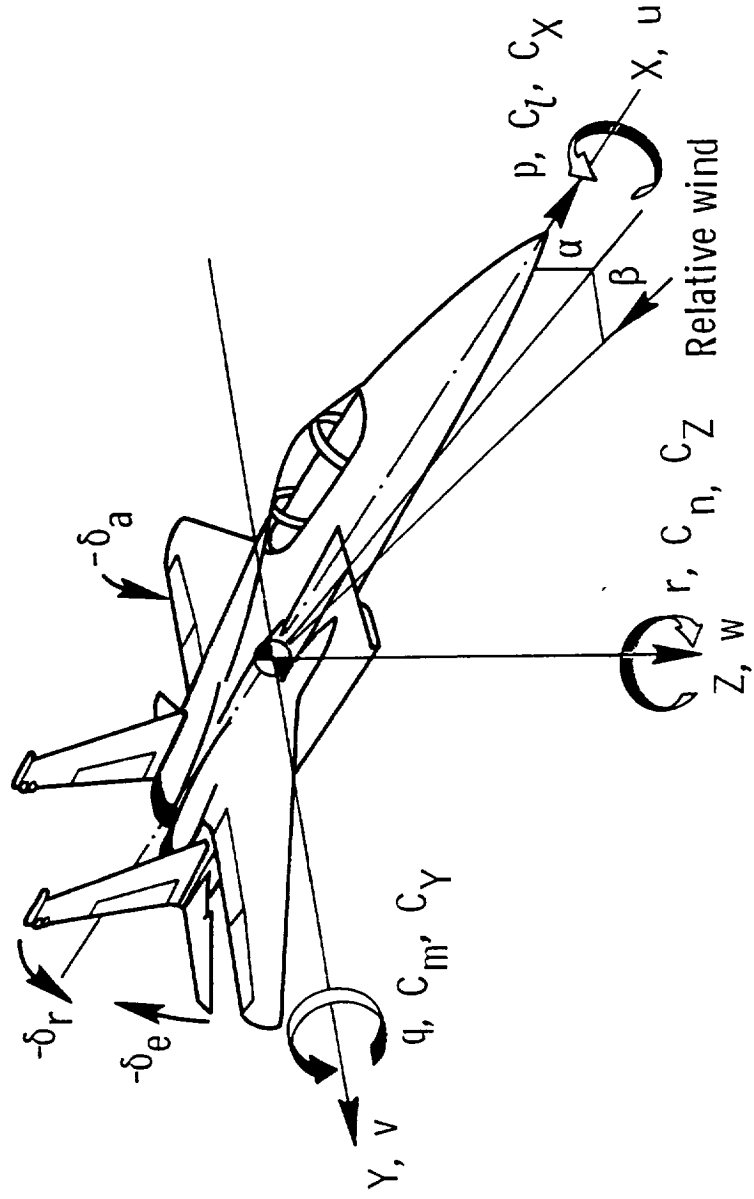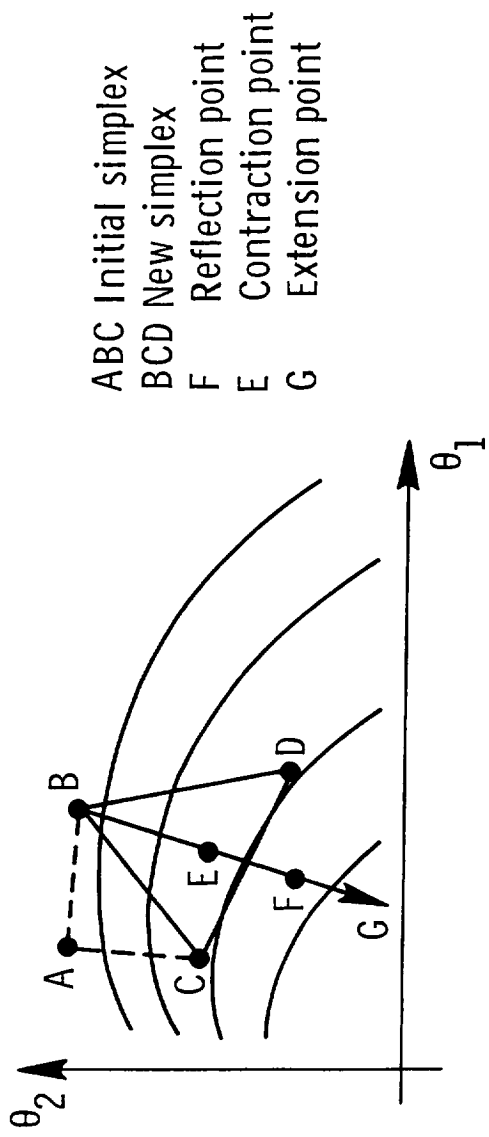
Figure 1.- Airplane body axis system with forces and moments.

- 76 -

ABC Initial simplex
BCD New simplex
F Reflection point
E Contraction point
G Extension point

Figure 2.- Two steps of a two-dimensional flexible polyhedron search.
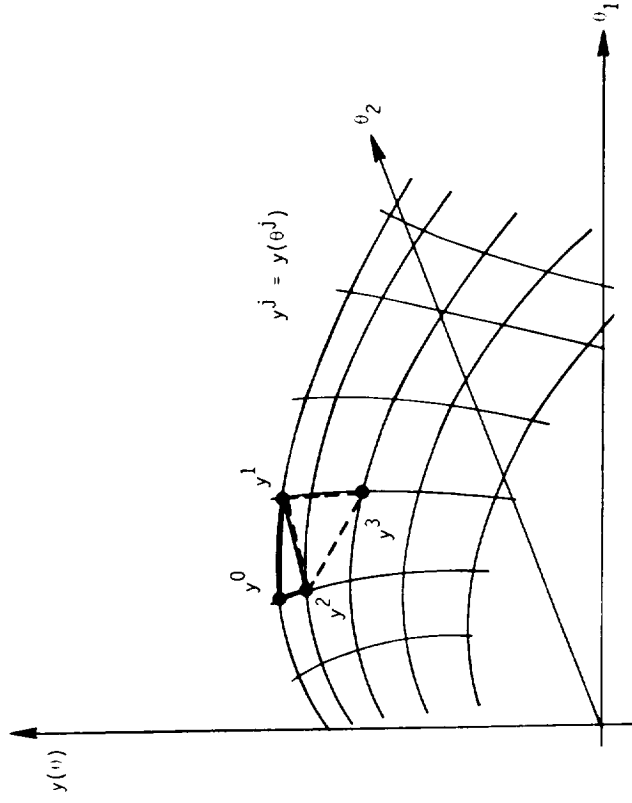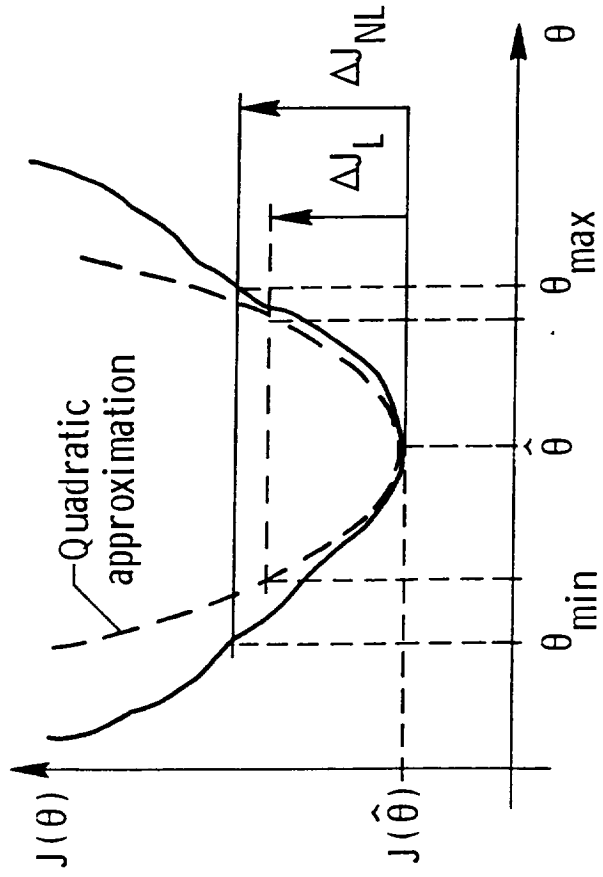
Figure 3.- Linear-surface fit for two iterations of MNRES.

$$\Delta J_L = n_p s^2 F_{\alpha_p}(n_p, N-n_p) \quad \Delta J_{NL} = n_p s^2 F_{\alpha_p}(n_p, N-n_p)\left(1 + \frac{N(n_p+2)}{(N-n_p)n_p}N_\phi\right)$$

Figure 4.- Construction of confidence intervals for one-dimensional case.
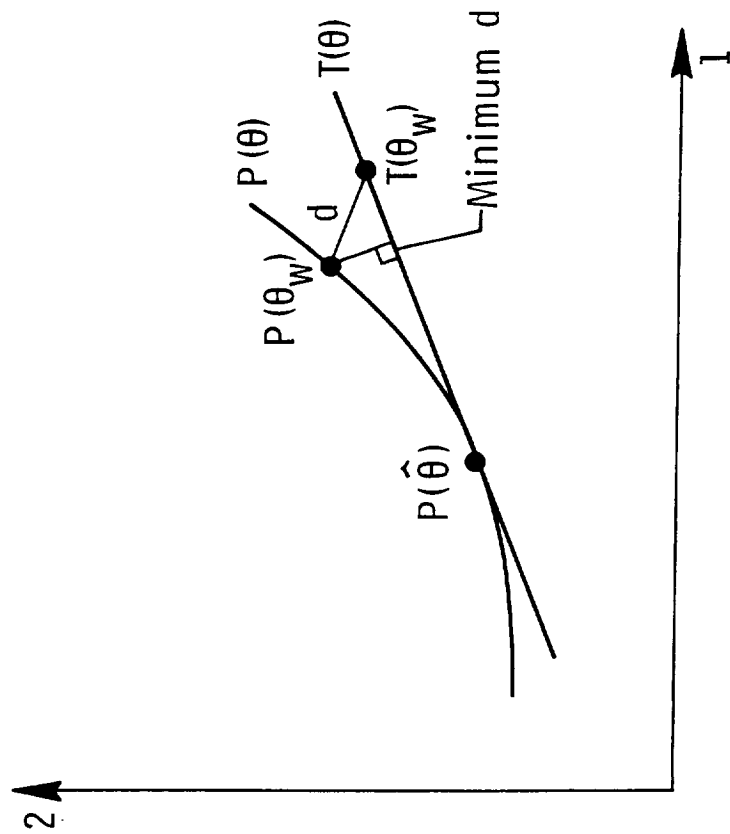
Figure 5.— Two-dimensional sample space with solution locus, $P(\theta)$, and tangent plane, $T(\theta)$, at $P(\hat{\theta})$.
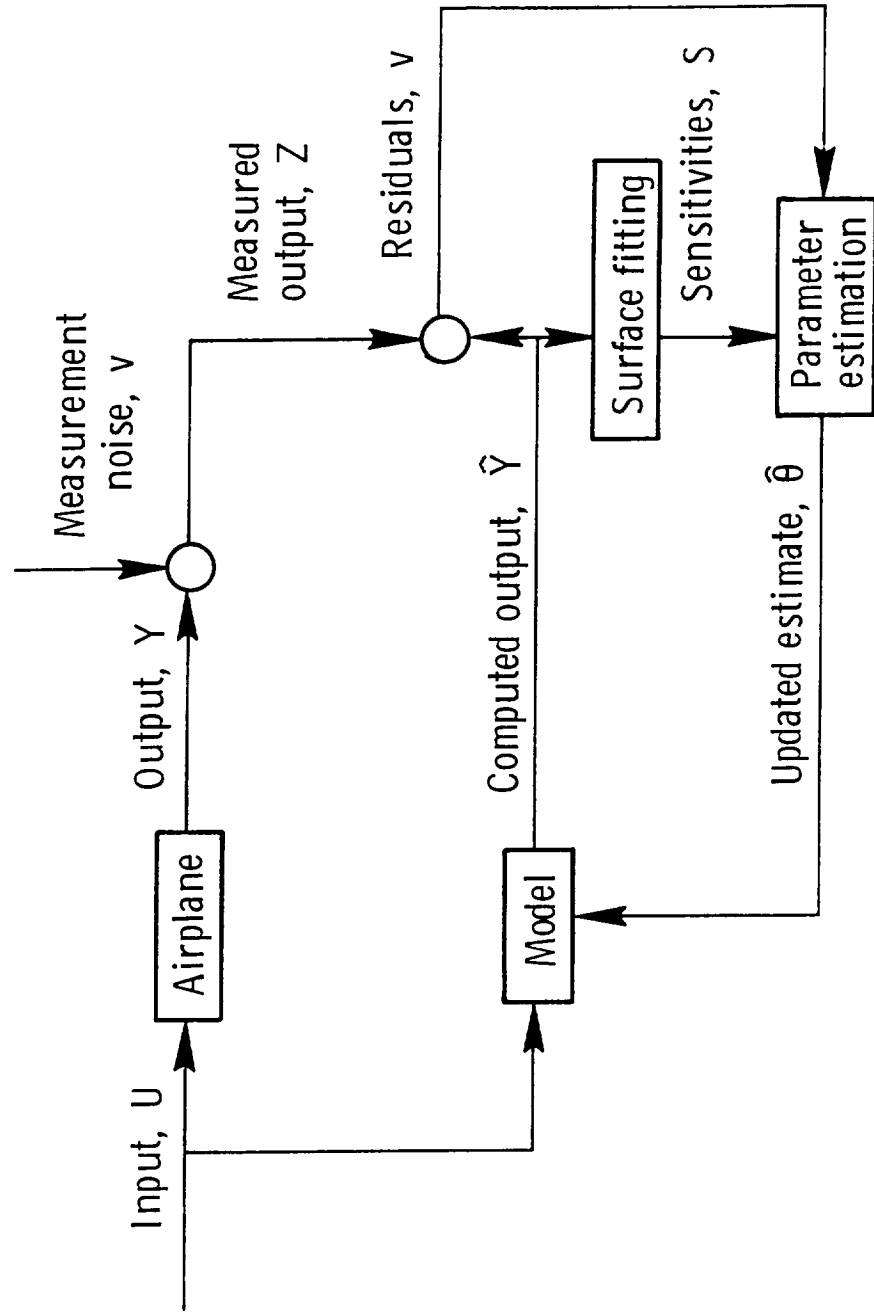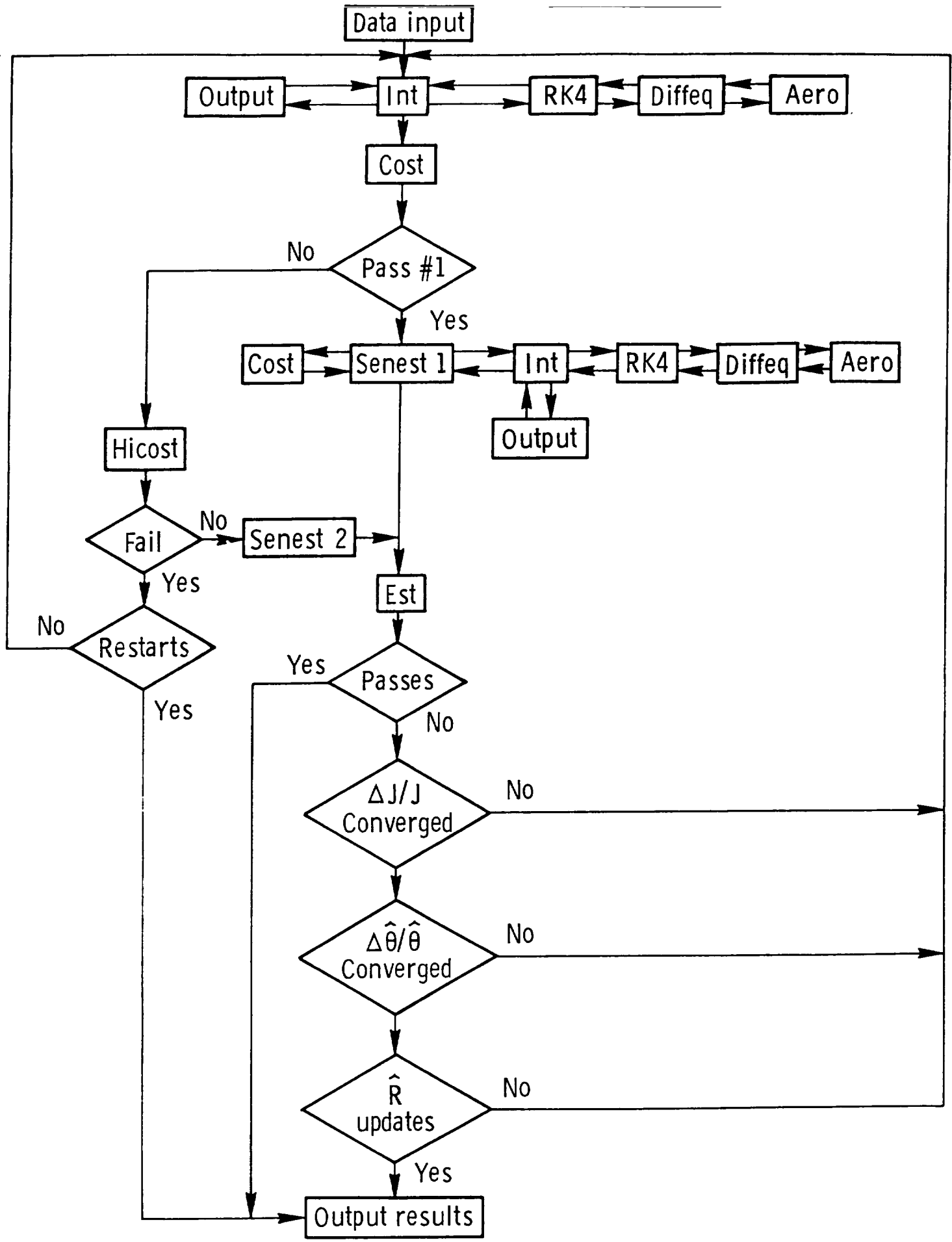
Figure 6.— Computing scheme for ML/MNRES.

Figure 7.- Flowchart for program MAX.

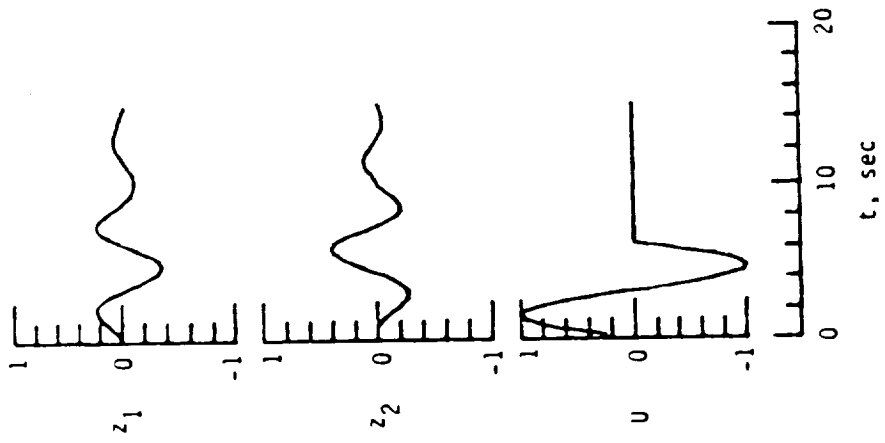Figure 8.- Time histories of input and response variables for Example I.
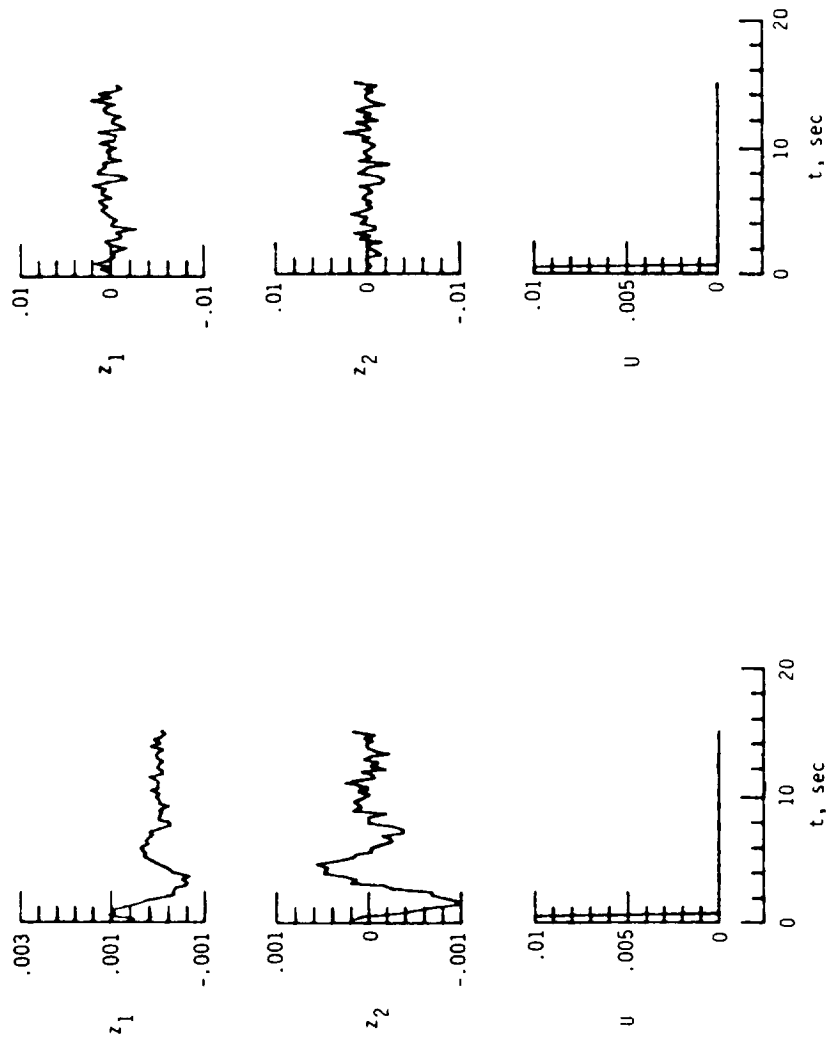
Figure 9.- Time histories of input and response variables for Example II.
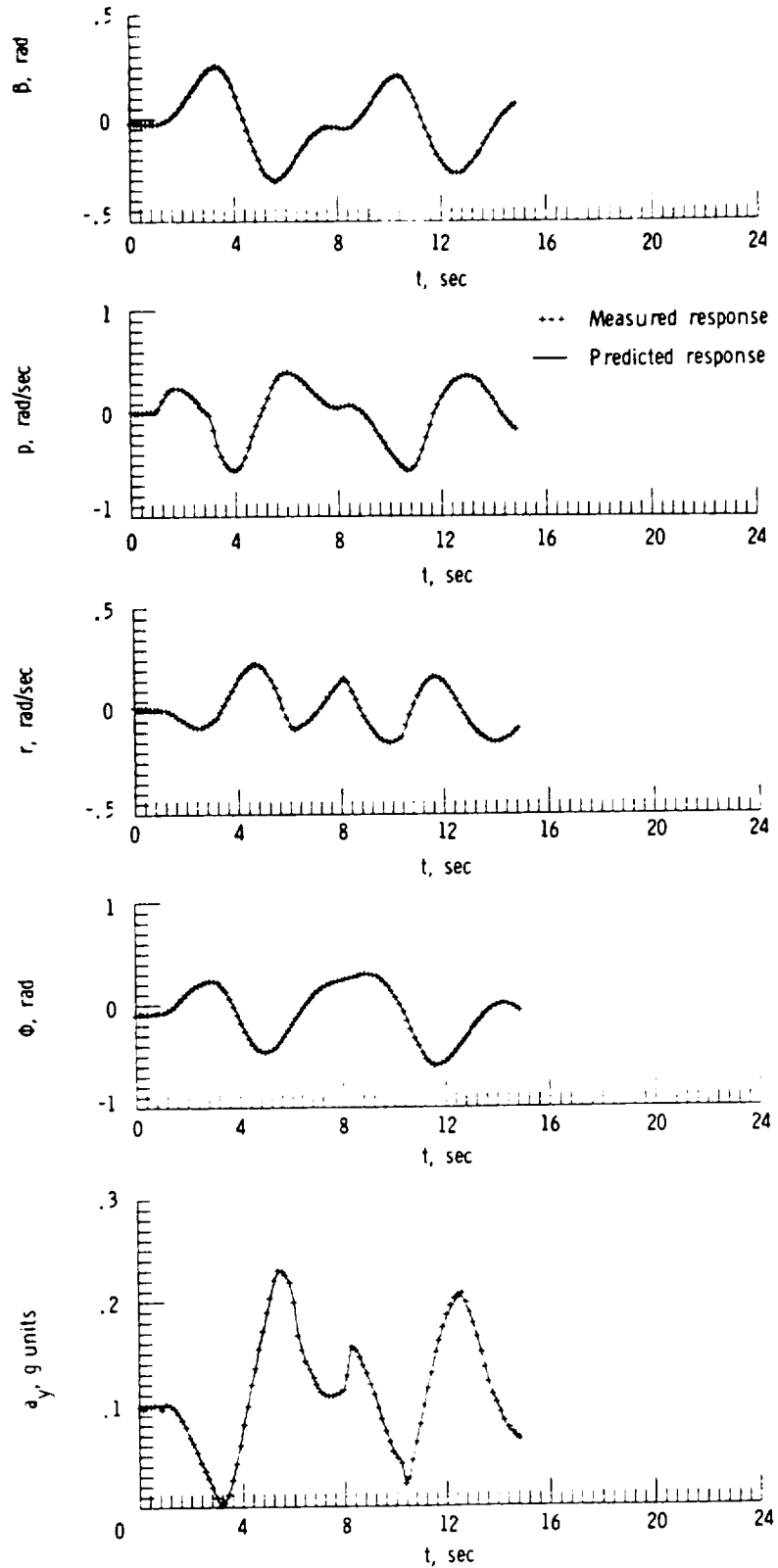
(a) Case 1.

Figure 10.- Measured and predicted responses for lateral simulation.
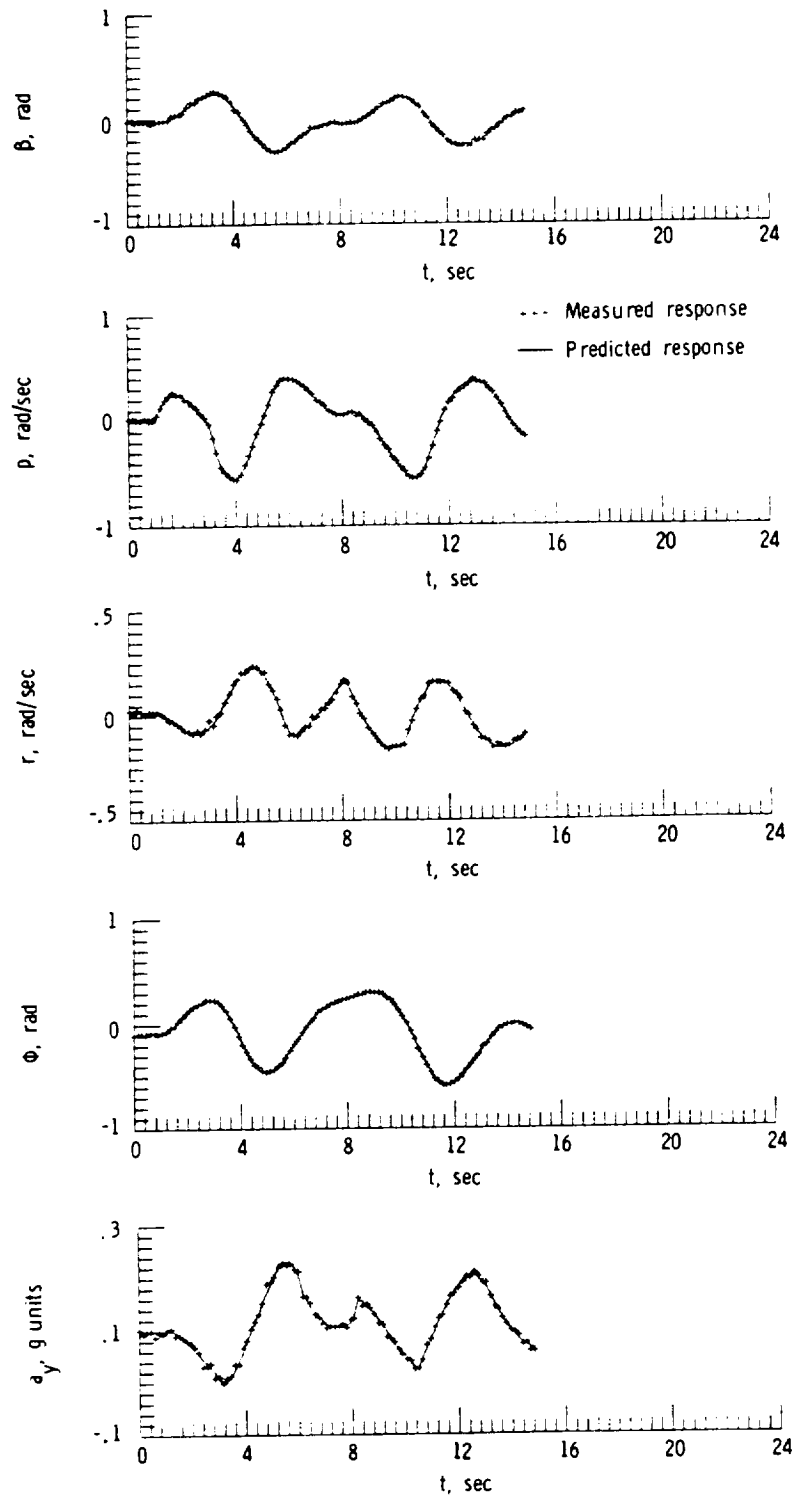
(b) Case 2.

Figure 10.- Continued.

(c) Case 3.

Figure 10.- Concluded.

Figure 11.- Control inputs for lateral simulation.
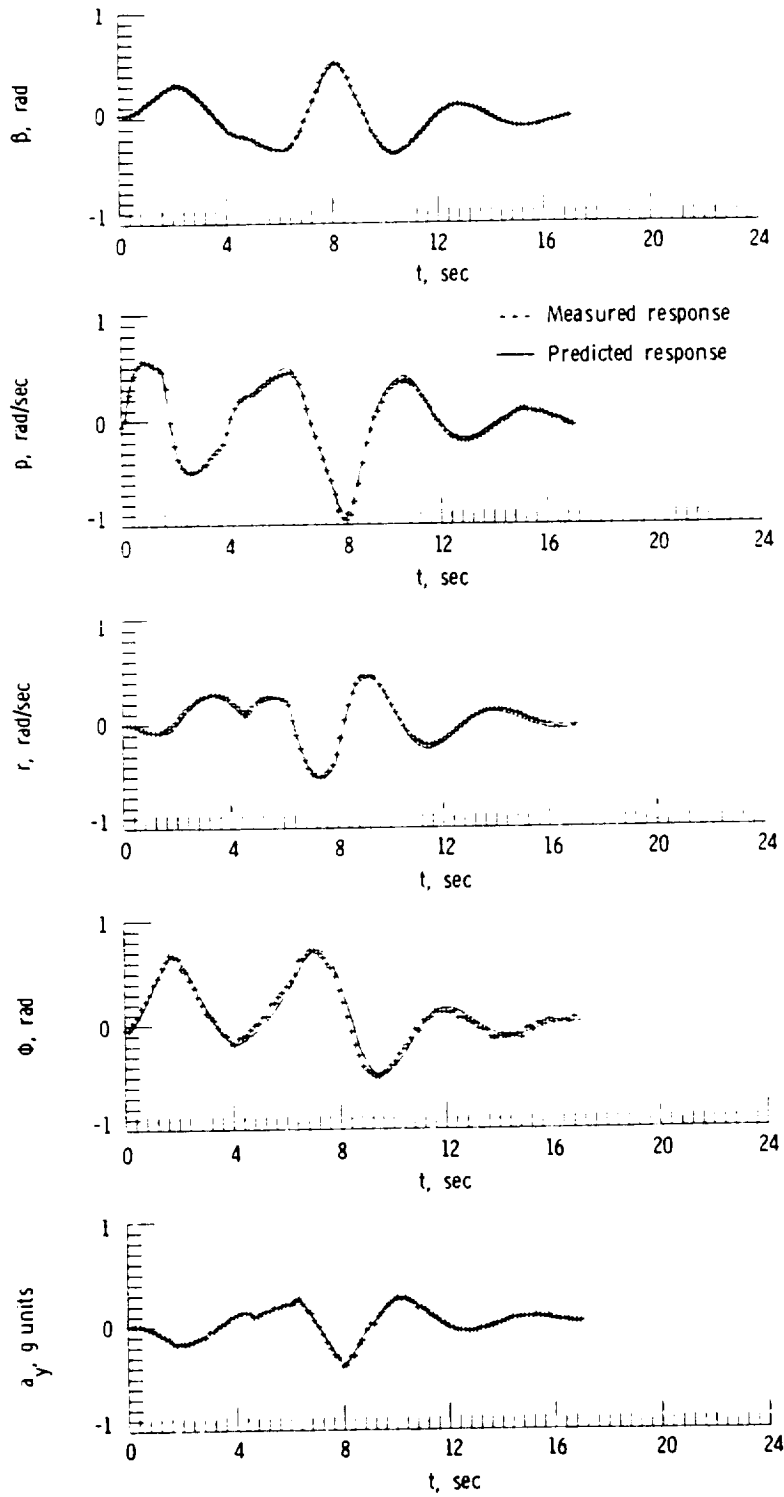
Figure 12.- Measured and predicted responses for lateral maneuvering using real flight data.
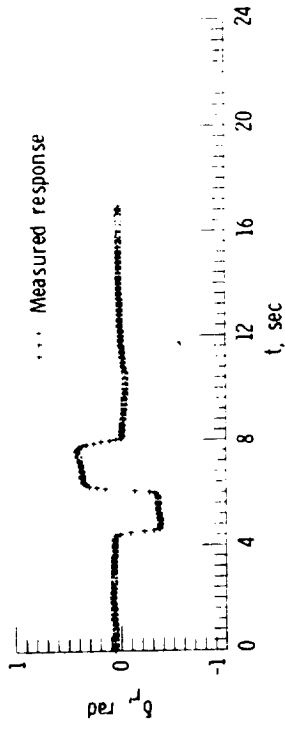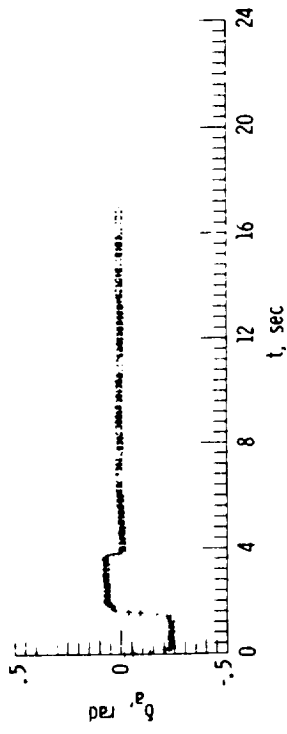
Figure 12.- Concluded.
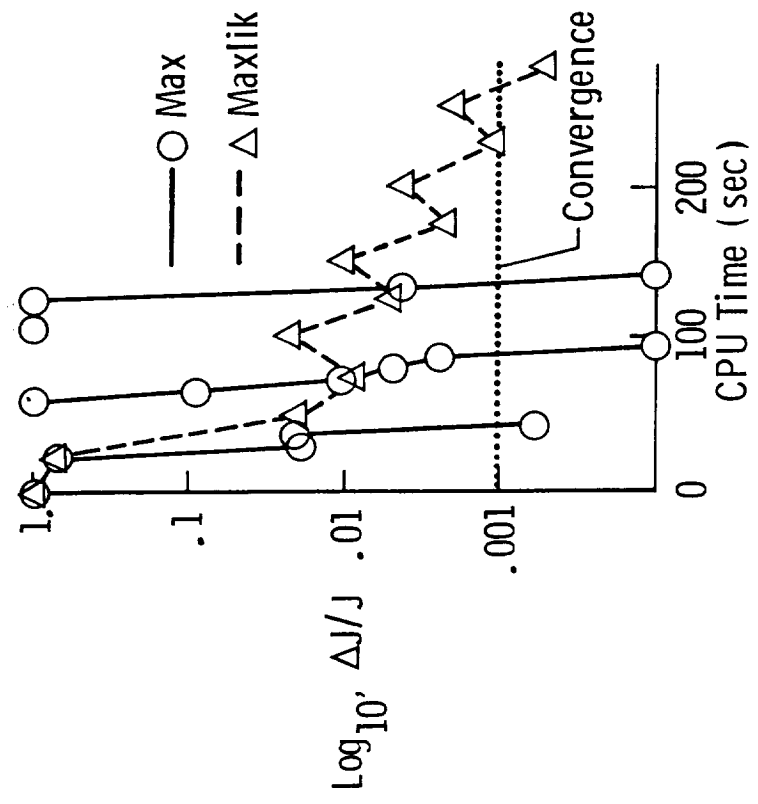
Figure 13.- Convergence criterion versus CPU time for programs MAX and MAXLIK.

Table 1. Primary subroutines for program MAX and flowchart block definitions

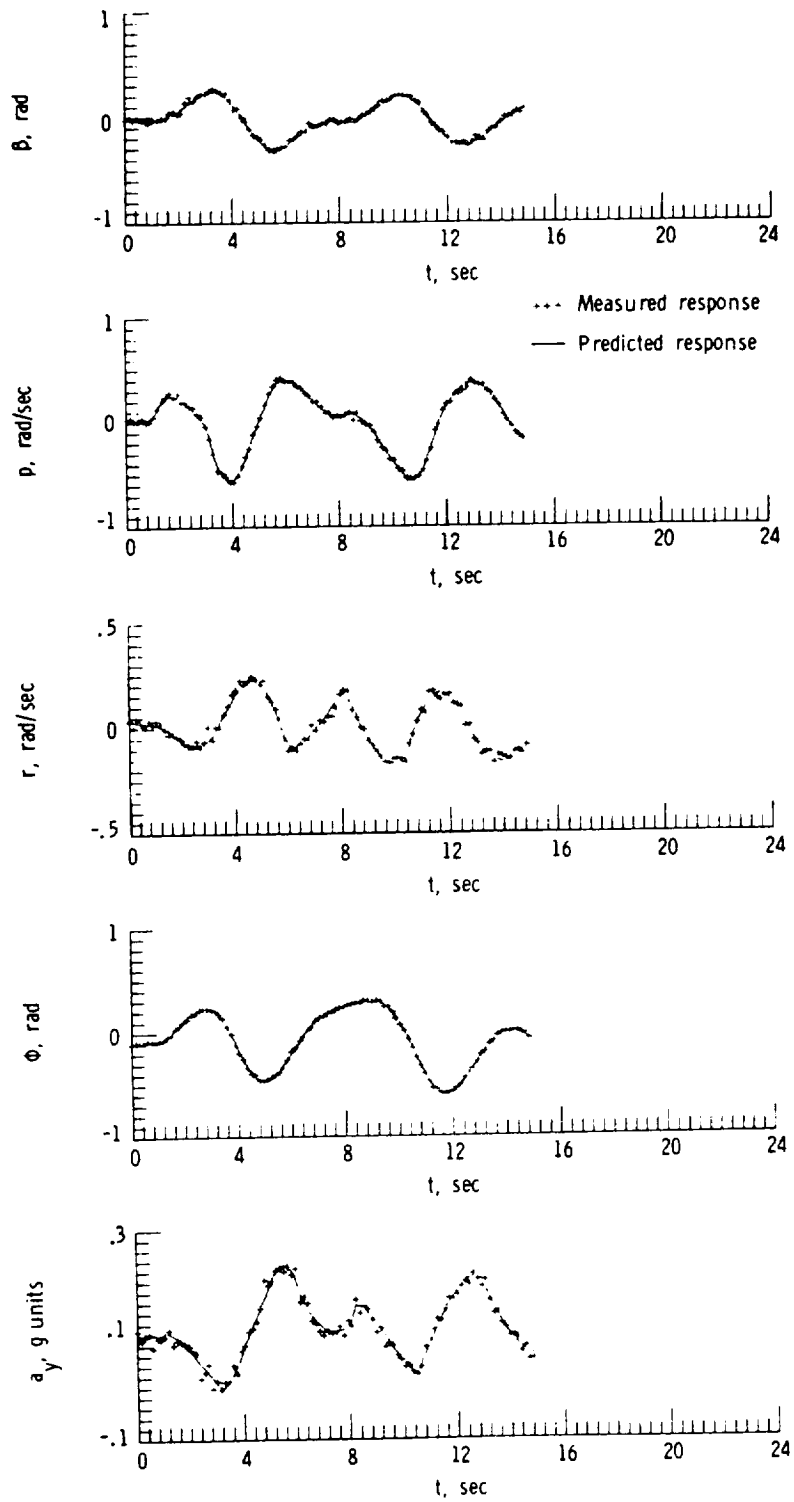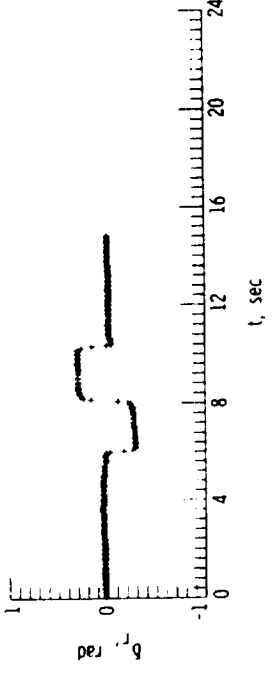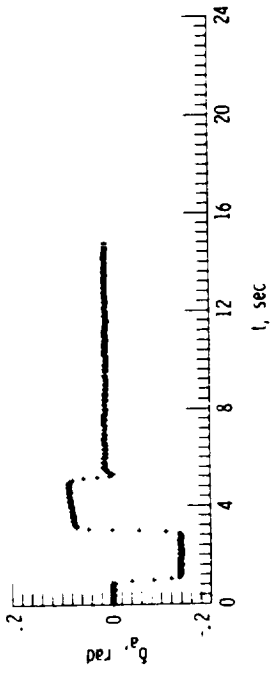| Subroutine	Description
AERO	Computes aerodynamic forces and moments with selected aerodynamic model
COST	Computes residuals, fit error, $\hat{R}_r^{-1}$ and cost
DIFFEQ	Computes state derivatives from selected equations of motion
EST	Computes updated parameter estimates
HICOST	Determines if new estimates reduce cost and updates storage of outgoing and incoming parameters and response time histories
INT	Main integration subroutine, computes initial conditions and input arrays
MASTER	Primary subroutine represented by flowchart; handles initializations, I/O operations and memory management
OUTPUT	Computes selected output time histories for cost function and plot routines
RK4	Fourth-order Runge-Kutta integration scheme
SENEST1	Computes sensitivities using a finite difference method
SENEST2	computes sensitivities using a selected surface-fitting method
Decision Blocks	**Definitions**
FAIL	test if new estimates reduce cost
PASSES	test for maximum number of allowed passes
PASS #1	test for first pass
RESTARTS	test for maximum number of restarts
$\hat{R}$ UPDATES	test for maximum number of weighting matrix updates

Table II. Comparison of estimates and computation time for FPS, MNR, and MNRES using a linear system without measurement noise (Example I)

Unknown parameters, θ	True Value of θ	Initial Value of θ	Final estimated values using method –		
			FPS	MNR	MNRES
θ_1	0.	0.01	-0.12 E-03	0.89 E-07	0.73 E-06
θ_2	-1.5	-1.6	-1.5	-1.5	-1.5
θ_3	1.0	1.1	1.0	1.0	1.0
θ_4	- .5	- .6	-.5	-.5	-.5
θ_5	.2	.25	.2	.2	.2
θ_6	.1	.15	.1	.1	.1
Cost			0.14 E-08	0.61 E-10	0.11 E-07
Equivalent evaluation ...			715	28	12
Δt, sec			2948	106	47

Table III. Comparison of estimates and computation time for MNR, and MNRES using a linear system with two noise levels (Example II)

(a) Case 1

Unknown parameters, θ	True Value of θ	Initial Value of θ	Final estimated values using method –	
			MNR	MNRES
θ_1	0.	0.01	−0.0675	−0.0684
θ_2	−1.5	−1.6	−1.471	−1.471
θ_3	1.0	1.1	1.009	1.010
θ_4	− .5	− .6	− .449	− .448
θ_5	.2	.25	.202	.202
θ_6	.1	.15	.098	.098
Cost			0.105 E−06	0.105 E−06
Equivalent evaluation ...			42	12
Δt, sec			77.54	24.08

(b) Case 2

Unknown parameters, θ	True Value of θ	Initial Value of θ	Final estimated values using method –	
			MNR	MNRES[1]
θ_1	0.	0.01	− .705	− .410
θ_2	−1.5	−1.6	−1.228	−1.549
θ_3	1.0	1.1	1.757	.799
θ_4	− .5	− .6	.159	− .238
θ_5	.2	.25	.210	.251
θ_6	.1	.15	.087	.037
Cost			0.104 E−04	0.122 E−04
Equivalent evaluation ...			70	27
Δt, sec			134.44	56.05

[1]MNRES did not converge.

Table IV. Standard errors of simulated measurement noise

Output Variable	Standard deviations for –		
	Case 1	Case 2	Case 3
β, rad	0	0.010	0.02
p, rad/sec ...	0	0.010	0.02
r, rad/sec ...	0	0.010	0.02
ϕ, rad	0	0.005	0.01
a_y, g units ..	0	0.005	0.01

Table V. Simulated-data analysis of a nonlinear aircraft system using ML/MNRES

Unknown parameter, θ	Simulation Values	Parameter estimates for –		
		Case 1	Case 2	Case 3
$C_{Y,0}$	0.13	0.1299	0.1298	0.1295
C_{Y_β}	-.411	-.4136	-.4261	-.4401
C_{Y_p}	-.146	-.1524	-.1874	-.2379
C_{Y_r}	.63	.6686	.6070	.5412
$C_{Y_{\delta a}}$	-.053	-.0618	-.0733	-.0872
$C_{Y_{\delta r}}$	.075	.0794	.0775	.0751
$C_{\ell,0}$	0	.0001	-.0003	-.0005
C_{ℓ_β}	-.123	-.1223	-.1228	-.1240
C_{ℓ_p}	-.397	-.3988	-.4026	-.4094
C_{ℓ_r}	.257	.2573	.2409	.2239
$C_{\ell_{\delta a}}$	-.182	-.1815	-.1778	-.1755
$C_{\ell_{\approx r}}$	.077	.0067	.0059	.00497
$C_{\cong_{\tau p}}$	2.63	2.6254	2.519	2.4359
$C_{n,0}$	0	-.00005	-.00008	-.0001
C_{n_β}	0	.000003	.0001	.0005
C_{n_p}	-.15	-.1488	-.1524	-.1558
C_{n_r}	-.083	-.0828	-.0861	-.0911
$C_{n_{\delta r}}$	-.0431	-.0425	-.0434	-.0445
$C_{n_{\alpha r}}$	1.7	1.7343	1.4419	1.0118

Table VI. Real-data analysis of a nonlinear aircraft system using programs MAX and MAXLIK

Unknown parameter, θ	Initial Value of θ	Program MAX $\hat{\theta}$	Program MAX $\hat{\sigma}$	Program MAXLIK $\hat{\theta}$	Program MAXLIK $\hat{\sigma}$	$\theta_j^2 \, M_{jj}$
$C_{Y,0}$	.036	0.0061	0.0006	0.0213	.0005	.7533 E+03
C_{Y_β}	− .479	−.4603	.0075	−.4608	.0067	.3756 E+05
C_{Y_p}	− .186	−.1378	.0485	−.0604	.0439	.9373 E+03
C_{Y_r}	.522	.6677	.0289	.6209	.0256	.1915 E−04
$C_{Y_{\delta a}}$	− .08	−.0504	.0166	−.0375	.0150	.8129 E+03
$C_{Y_{\delta r}}$	.083	.0814	.0043	.0763	.0037	.9922 E+03
$C_{Y_{\alpha\beta}}$	.45	.4300	.0592	.4512	.0504	.9399 E+03
$C_{\ell,0}$	0.	.0002	.00005	−.0001	.00005	.9618 E+03
C_{ℓ_β}	− .079	−.0872	.0015	−.08	.0013	.1493 E+06
C_{ℓ_p}	− .47	−.5320	.0102	−.4823	.0085	.4529 E+07
C_{ℓ_r}	.187	.1700	.0043	.1543	.0045	.6838 E+04
$C_{\ell_{\delta a}}$	− .19	−.2035	.0036	−.1852	.0031	.7969 E+05
$C_{\ell_{\delta r}}$	.01	.00055	.00024	−.0012	.00072	.1340 E+04
$C_{\ell_{\alpha\beta}}$	− .26	−.2707	.0116	−.2105	.0091	.6327 E+04
$C_{n,0}$	0.	−.00063	.00003	−.002	.00002	.1909 E+04
C_{n_β}	.04	.0323	.00045	.0329	.0004	.1515 E+06
C_{n_p}	− .056	−.1043	.0026	−.0916	.0022	.1400 E+06
C_{n_r}	− .15	−.1462	.002	−.1534	.0017	.5039 E+05
$C_{n_{\delta a}}$	0.	−.0044	.001	−.0037	.0009	.1780 E+04
$C_{n_{\delta r}}$	− .053	−.0550	.0003	−.0532	.0003	.9048 E+07
$[1]C_{Y_\beta 3}$	− .39	−.39		−.39		
$[1]C_{n_\beta 3}$	.08	.08		.08		

[1]Parameter held fixed

Table VII. Comparison of parameter estimates, their
standard errors and time to reach convergence
for programs MAX and MAXLIK

	MAXLIK		MAX			
	CPUT = 342*		CASE 1		CASE 2	
			CPUT = 105		CPUT = 157	
	$\hat{\theta}$	$\hat{\sigma}(\hat{\theta})$**	$\hat{\theta}$	$\hat{\sigma}(\hat{\theta})$	$\hat{\theta}$	$\hat{\sigma}(\hat{\theta})$
C_{x_0}	.17017	.208E-03	.17078	.249E-03	.17070	.212E-03
C_{x_α}	.9464	.616E-02	.9245	.703E-02	.92511	.555E-02
$C_{x_{\delta e}}$	.2789	.404E-02	.2754	.116E-01	.2755	.876E-02
C_{z_0}	-.83946	.896E-03	-.84335	.106E-02	-.84252	.103E-02
C_{z_α}	-5.311	.230E-01	-5.197	.223E-01	-5.194	.215E-01
C_{z_q}	-18.7	.162+01	-20.3	.189E+01	-20.5	.177E+01
$C_{z_{\delta e}}$	-.618	.264E-01	-.566	.350E-01	-.570	.324E-01
C_{m_0}	-.001251	.828E-04	-.001610	.942E-04	-.001555	.937E-04
C_{m_α}	-.5129	.102E-02	-.5186	.115E-02	-.5183	.110E-02
C_{m_q}	-16.95	.157E+00	-19.06	.144E+00	-18.83	.144E+00
$C_{m_{\delta e}}$	-1.3409	.576E-02	-1.4150	.447E-02	-1.4075	.461E-02

* Central processing unit time in sec.
** Cramer-Rao bound.

Table VIII. Parameter estimates and their confidence limits using Cramer-Rao boards and a random search technique

Para-meter	$\hat{\theta}$	95% Confidence Intervals		
		Cramer-Rao bound	Random Search	
c_{Y_β}	-.77	± .27	.92	-.89
$c_{Y_{\delta r}}$	.18	± .27	.87	-.89
c_{ℓ_β}	-.228	± .021	.042	-.10
c_{ℓ_p}	-.88	± .13	.30	-.99
c_{ℓ_r}	-4.20	± .98	2.1	-5.5
$c_{\ell_{\delta a}}$	-.152	± .020	.036	-.11
c_{n_β}	.0826	± .0040	.011	-.013
c_{n_p}	-.078	± .019	.050	-.10
c_{n_r}	-1.24	± .19	.49	-.76
$c_{n_{\delta r}}$	-.0860	± .0064	.020	-.021