NASA-RP-1153 19860013108

NASA
Reference
Publication
1153

April 1986

A Methodology for Airplane Parameter Estimation and Confidence Interval Determination in Nonlinear Estimation Problems

Patrick C. Murphy

NASA

NASA Reference Publication 1153

1986

A Methodology for Airplane Parameter Estimation and Confidence Interval Determination in Nonlinear Estimation Problems

Patrick C. Murphy

Langley Research Center
Hampton, Virginia

National Aeronautics and Space Administration

Scientific and Technical Information Branch

SUMMARY

A methodology is presented for airplane parameter estimation and confidence interval determination in nonlinear estimation problems. In addition, as part of the methodology, an efficient scheme to determine aerodynamic model structure is suggested and briefly described. The algorithms described provide a unified approach to solving nonlinear airplane identification problems. The nonlinear estimation problems of interest to this study are further complicated by nonlinear system dynamics and in particular nonlinear aerodynamic models.

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 modified Newton-Raphson with estimated sensitivities (MNRES).

MNRES determines sensitivities by using slope information from local surface approximations of each output variable in the parameter space. The fitted surface allows sensitivity information to be updated at each iteration with a significant reduction in computational effort. With MNRES, the sensitivities can be determined with less computational effort than with either a finite-difference method or integration of the analytically determined sensitivity equations. The type of surface (for example, nth-order polynomial or spline) and the method of fitting the surface (for example, least squares or solution of simultaneous equations) are chosen 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 convenient format.

Two surface-fitting methods are discussed and demonstrated, while other possibilities are indicated. MNRES is compared with other commonly used optimization methods, 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 depends 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 3 to 8 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 CR bounds and search estimates.

CONTENTS

TABLES

FIGURES

SYMBOLS

$\mathbf{a}_k$	incoming parameter vector for MNRES algorithm at kth iteration
$\mathbf{a}_k^o$	outgoing parameter vector for MNRES algorithm at kth iteration
a_x, a_y, a_z	acceleration in x-, y-, and z-directions, g units
$\mathbf{A}$	system matrix for state equation
$\mathbf{b}$	arbitrary vector through confidence region
b_w	wing span, m
$\mathbf{B}$	control-distribution matrix for state equation
$\mathbf{B}_{i/i-1}$	covariance matrix for the measurement $(\mathbf{Z}_i/\mathbf{Z}_{i-1}, \boldsymbol{\theta})$
$\bar{c}$	mean aerodynamic chord, m
C_D	drag coefficient, $F_D/\bar{q}S_w$
C_l	rolling-moment coefficient, $M_X/\bar{q}S_w b_w$
$C_{l,o}, C_{m,o}, C_{n,o}$	aerodynamic moments for trimmed flight
C_L	lift coefficient, $F_L/\bar{q}S_w$
C_m	pitching-moment coefficient, $M_Y/\bar{q}S_w\bar{c}$
C_n	yawing-moment coefficient, $M_Z/\bar{q}S_w b_w$
C_T	thrust coefficient, $F_T/\bar{q}S_w$
C_X	longitudinal-force coefficient, $F_X/\bar{q}S_w$
$C_{X,o}, C_{Y,o}, C_{Z,o}$	aerodynamic force for trimmed flight
C_Y	lateral-force coefficient, $F_Y/\bar{q}S_w$
C_Z	vertical-force coefficient, $F_Z/\bar{q}S_w$
d	squared distance between $P(\boldsymbol{\theta}_w)$ and $T(\boldsymbol{\theta})$
d_{ij}	element ij of the inverse of $\mathbf{H}$ or $\mathbf{M}$
D	sum of squares of the squared distances from $P(\boldsymbol{\theta}_w)$ to $P(\hat{\boldsymbol{\theta}})$
$\mathbf{e}$	vector of equation error
e_i	scalar equation error at ith data point
$E\{\ \}$	expectation operator
f	general function; also probability density function
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
F_{α_p}	F-statistic at confidence level $1 - \alpha_p$
g	acceleration due to gravity, 9.81 m/sec^2
$\mathbf{G}_i$	sensitivity matrix at ith data point
h	general function
$\mathbf{H}$	Hessian matrix
$\mathbf{H}_\alpha$	$= \left\{ n_p s^2 F_{\alpha_p}(n_p, Nn_o - n_p) \right\}^{-1} \mathbf{H}$

H_0, H_1	null and alternative hypotheses
I_{ij}	moments of intertia, where i and 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 and j are x, y, or z
L_1	likelihood function
L	$= -\ln L_1$
m	mass, kg
$\mathbf{M}$	Fisher information matrix
$M_{\ell\ell}$	element $\ell\ell$ on diagonal of $\mathbf{M}$
M_X, M_Y, M_Z	rolling, pitching, and yawing moments, N-m
n	number of surface-fitting points
n_o	number of outputs
n_p	number of parameters
n_s	number of states
N	number of data points
N_ϕ	nondimensional intrinsic nonlinearity measure
N_ψ	nondimensional nonlinearity measure
p	roll rate, rad/sec
$\mathbf{P}_k$	least-squares-parameter covariance matrix at kth iteration divided by σ
$P(\boldsymbol{\theta})$	estimation space
Pr{ }	probability of { }
q	pitch rate, rad/sec
q	dynamic pressure, $\rho V^2/2$, Pa
Q_ϕ	dimensional intrinsic nonlinearity measure
Q_ψ	dimensional nonlinearity measure
r	yaw rate, rad/sec
$\mathbf{R}$	measurement noise covariance matrix
R_c	confidence region
s^2	$= J(\hat{\boldsymbol{\theta}})/(Nn_o - n_p)$
$\mathbf{S}$	vector of sensitivities
$\mathbf{S}_{1i}$	vector of sensitivities for first element of $\mathbf{Y}$ at the ith data point
$\mathbf{S}_k$	vector of sensitivities for kth element of $\mathbf{Y}$
S_w	wing area, m^2
$s_{k\ell}$	sensitivity of kth element of $\mathbf{Y}$ to ℓth element of $\boldsymbol{\theta}$
t	time, sec

$T(\boldsymbol{\theta})$	tangent plane to $P(\boldsymbol{\theta})$ at $P(\hat{\boldsymbol{\theta}})$
u	input function or velocity along X-body axis, m/sec
$\mathbf{U}$	input vector
v	velocity along Y-body axis, m/sec
$\mathbf{v}_i$	vector of measurement noise at ith data point
$\mathbf{v}_x, \mathbf{v}_y$	measurement noise vectors for predictor variable $\mathbf{X}_i$ and response variable $\mathbf{Y}$
V	airspeed, m/sec
w	velocity along Z-body axis, m/sec
$\mathbf{w}_p$	process noise vector
W	number of different $\boldsymbol{\theta}$ chosen for N_ϕ computation
x_i	scalar element of $\mathbf{X}_i$
$\mathbf{X}_i$	vector of predictor variables at ith data point
$\mathbf{X}$	matrix of predictor variables: parameters in MNRES and state variables in LR
$\mathbf{X}_s$	vector of states
X, Y, Z	longitudinal, lateral, and vertical body axes
y	scalar output
y_k	kth element of output vector $\mathbf{Y}$
y_k^o	outgoing kth element of output vector $\mathbf{Y}$
y_{ki}^j	kth element of $\mathbf{Y}$, at the ith data point, evaluated at the jth surface-fitting point, $\boldsymbol{\theta}^j$
$\mathbf{Y}$	vector of n_o outputs or N values of response variable
$\hat{\mathbf{Y}}_{i/i-1}$	best estimate of the measurement at time t_i given measurements up to and including the previous point
$\mathbf{Y}_{1i}$	vector of j values of y_{ki}^j where $k = 1$
z_j	jth element of $\mathbf{Z}_i$
$\mathbf{Z}_i$	vector of measured outputs at ith data point
$\mathbf{Z}_i^I$	$= \mathbf{Z}_i, \mathbf{Z}_{i-1}, \ldots, \mathbf{Z}_1$
α	angle of attack, rad
α_p	probability of type I error; $1 - \alpha_p$ is the confidence level associated with F-statistic
α_T	angle of thrust, rad
β	sideslip angle, rad
δ_a	aileron deflection, rad
δ_e	elevator deflection, rad
δ_{ij}	Kronecker delta
δ_r	rudder deflection, rad

ε	vector sum of process noise $\mathbf{w}_p$ and measurement noise $\mathbf{v}_y$
$\boldsymbol{\theta}$	vector of unknown parameters
θ_i	ith element of $\boldsymbol{\theta}$
$\boldsymbol{\theta}^*$	vector of optimal parameters
Θ	pitch angle, rad
λ	ratio of likelihood functions, L
μ	general statistic equal to $J(\boldsymbol{\theta}) - J(\hat{\boldsymbol{\theta}})$
$\boldsymbol{\nu}_i$	vector of residuals at ith data point
ρ	air density, kg/m^3
σ	standard error
ϕ	roll angle, rad
Φ	value of ψ which minimizes Q_ψ
Ψ_w	$= \boldsymbol{\theta}_w - \hat{\boldsymbol{\theta}}$
ω	subset of Ω
Ω	parameter space
Subscripts:	
E	measured quantity
H, L	highest and lowest costs
i, j, k, ℓ	general indices
r, k	iteration number (except as δ_r, which is rudder deflection)
t	true value
w	index for different values of $\boldsymbol{\theta}$ during N_ϕ computation
0	initial condition
Superscript:	
j	index of surface-fitting points
Abbreviations:	
CIE	confidence interval estimation
CPU	central processing unit
CR	Cramer-Rao
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 method
MNRES	modified Newton-Raphson method with estimated sensitivities
MSR	modified stepwise regression

NR	Newton-Raphson
LS	least squares

Matrix exponents:

T	transpose of matrix
-1	inverse of matrix

Mathematical notation:

$\hat{}$	estimated quantity when over symbol
$\cdot$	derivative with respect to time when over symbol
∇	gradient operator
Δ	incremental value
$\max_{\theta}( )$	maximum of ( ) with respect to θ
$N(\bar{x}, \sigma^2)$	normal distribution with mean $\bar{x}$ and variance σ^2
$\in$	is an element of
$\ni$	such that
$\|\ \|$	determinant of matrix

Aerodynamic derivatives:

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

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

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

$$C_{l_\beta} = \frac{\partial C_l}{\partial \beta} \qquad C_{l_p} = \frac{\partial C_l}{\partial \frac{pb_w}{2V}} \qquad C_{l_r} = \frac{\partial C_l}{\partial \frac{rb_w}{2V}}$$

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

$$C_{l_{\alpha p}} = \frac{\partial^2 C_l}{\partial \alpha \partial \frac{pb_w}{2V}} \qquad C_{n_\beta} = \frac{\partial C_n}{\partial \beta} \qquad C_{n_p} = \frac{\partial C_n}{\partial \frac{pb_w}{2V}}$$

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

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

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

$$C_{Z_{\delta e}} = \frac{\partial C_Z}{\partial \delta_e} \qquad C_{Z_q} = \frac{\partial C_Z}{\partial \frac{q\bar{c}}{2V}} \qquad C_{m_\alpha} = \frac{\partial C_m}{\partial \alpha}$$

$$C_{m_{\delta e}} = \frac{\partial C_m}{\partial \delta_e} \qquad C_{m_q} = \frac{\partial C_m}{\partial \frac{q\bar{c}}{2V}}$$

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 **U** and **f** are given and **Y** is to be determined, (2) the controls problem where **f** and the desired **Y** are given and **U** is to be determined, and (3) the identification problem where **U** and **Y** have been measured and **f** is to be modeled.

The theory for system identification provides a way for modeling an unknown system on the basis of input and output information. 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 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 identification theory usually follows four basic stages. The first stage requires the design of an experiment, that is, specification of the identification objectives, statement of system configuration and conditions, and selection of an input form. Determining an optimal input for identification can be critical; all the modes of a system must be excited in order to identify the system correctly and completely. The second stage is 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, with or without measurement noise, etc. The unknown parameters in the model may include system parameters as well as initial conditions, bias terms, and characteristics of measurement and process noise. The third stage involves parameter and state estimation, which provides 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. The fourth stage is verification, accomplished by comparing estimates from experimental data sets and other estimation techniques. In addition, other sources provide comparisons; for 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.

System identification theory has become important to aircraft technology for several reasons. 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. By offering an opportunity to observe actual vehicle performance, flight testing results in better calibration and understanding of wind tunnel results and more accurate modeling for ground-based simulators.

The development of aircraft parameter estimation paralleled 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 analyzing simple transient maneuvers (refs. 2, 3, and 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. Large-capacity, high-speed digital computers and highly automated data acquisition systems were introduced. In 1968, when Larson (ref. 6) applied the method of quasi-linearization 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), Stepner and Mehra (ref. 9), and Rault (ref. 10).

During the past decade application of estimation theory to nonlinear systems has become an increasing concern, stimulated primarily by aerospace applications. Today, nonlinear dynamics is commonly incorporated with linear aerodynamic models 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, modeling the combination of nonlinear dynamics and 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, which suggests a useful technique for model structure determination when nonlinear aerodynamic effects are present. 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 that 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, 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 integration of 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 unknown parameters are not uncommon for one flight condition. If a model is desired throughout the entire flight envelope, the computational requirements become overwhelming, particularly when analysis of various flight conditions requires 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 high computational cost associated with the ML/MNR algorithm, an additional difficulty 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 modeling aircraft in nonlinear flight regimes, since model structure may change significantly from one flight condition to another. Therefore, an algorithm that is independent of sensitivity equations is very advantageous.

Reducing computational requirements of the ML method requires careful examination of the optimization methods used 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 (ref. 14) considered the numerical aspects of computing ML estimates for linear dynamic systems in state-vector form and methods for speeding up convergence. Bowles and Straeter (ref. 15) considered computational aspects for several methods used in aircraft identification. Trankle, Vincent, and Franklin (ref. 16) considered the difficulties associated with use of a nonlinear dynamic model in ML parameter estimation and parameter covariance estimation; sensitivity calculation methods were also considered. More recently, Trankle, Vincent, and Franklin (ref. 17) considered the overall methodology of system identification for nonlinear aerodynamic models including computational aspects of the problem. In reference 18, 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 tested on problems that do not involve dynamic systems. Presented in the current study is a significantly improved maximum likelihood algorithm, initially developed in reference 19, 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 that 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 computational demand is significantly reduced.

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 (ref. 20). Attempts have been made, therefore, either to modify the CR bounds by considering a band-limited measurement noise (refs. 21 and 22) or to estimate the parameter variance directly from measured data (ref. 10). Advances in statistical methods also came about with the availability of high-speed computers. In 1960, Beale (ref. 23) considered the problem in nonlinear estimation of determining the approximate parameter confidence regions using likelihood ratios.

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. 24). In 1979, Mereau and Prevost (ref. 25) used the likelihood ratio approach to determine confidence regions for aircraft systems. In 1980, Mereau and Raymond (ref. 26) developed a search procedure to find the "iso-distances" defining the confidence regions.

The goal of the current study is to provide a unified methodology for solving nonlinear airplane identification problems. Improved techniques for estimating parameters and their confidence limits in nonlinear, multivariable, dynamic systems, in particular, aircraft systems, are presented. The improved techniques 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 those obtained using CR bounds, and (4) an adaptation of Beale's approach to the airplane estimation problem.

The development of this work begins in section II with a description of the airplane model and the regression method used to determine it. Section III describes the parameter estimation techniques used and their statistical properties. The primary estimation method used in this study to determine airplane parameters 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 section II. In section 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 compared with MNRES. Section V develops the theory for confidence interval estimation and the adaptation of Beale's work to the airplane problem. Finally, section VI presents the results and discussion for the application of these methods to simulated and real data.

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, 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

$$\dot{\mathbf{X}}_s = f(\mathbf{X}_s, \mathbf{U}, \boldsymbol{\theta}, t) \quad \mathbf{X}_s(0) = \mathbf{X}_{s0} \tag{2.1}$$

$$\mathbf{Y} = h(\mathbf{X}_s, \mathbf{U}, \boldsymbol{\theta}, t) \tag{2.2}$$

where $\mathbf{X}_s$ is a vector of state variables, $\mathbf{Y}$ is a vector of output variables, $\mathbf{U}$ is a vector of input variables, and $\boldsymbol{\theta}$ 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 significant 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 vary 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 may depend on the initial guesses for the parameters because failure may occur when a parameter is outside a stability boundary. Unfortunately, obtaining stability boundaries is really practical only for linear, time-invariant systems. Finally, numerical difficulties with truncation and rounding errors are always present when numerical differentiation and integration are performed.

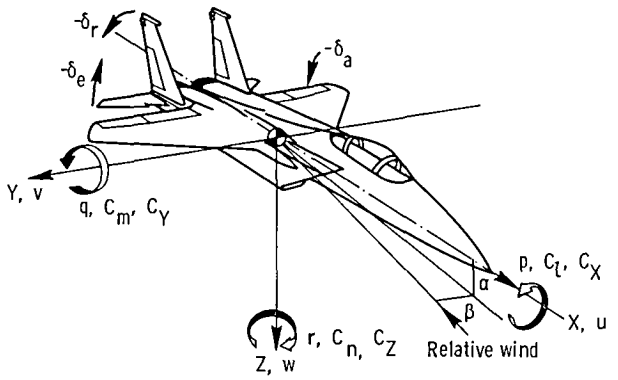


Figure 1. Airplane body-axis system with forces and moments.

A. Airplane Equations of Motion

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

$$\dot{\mathbf{X}}_s = f(\mathbf{X}_s, \mathbf{U}, \boldsymbol{\theta}) \quad \mathbf{X}_s(0) = \mathbf{X}_{s0} \tag{2.3}$$

$$\mathbf{Y} = h(\mathbf{X}_s, \mathbf{U}, \boldsymbol{\theta}) \tag{2.4}$$

The equations of motion used are referred to a body-axis system (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 the 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. These equations of motion are given as follows:

$$\dot{u} = -qw + rv + g\ell_{xz} + \frac{\bar{q}S_w}{m}C_X \tag{2.5}$$

$$\dot{v} = -ru + pw + g\ell_{yz} + \frac{\bar{q}S_w}{m}C_Y \tag{2.6}$$

$$\dot{w} = -pv + qu + g\ell_{zz} + \frac{\bar{q}S_w}{m}C_Z \tag{2.7}$$

$$\dot{p} = \frac{I_z}{I_xI_z - I_{zx}^2}F_1 + \frac{I_{zx}}{I_xI_z - I_{zx}^2}F_2 \tag{2.8}$$

$$\dot{q} = \frac{I_{zx}}{I_y}\left(r^2 - p^2\right) + pr\frac{I_z - I_x}{I_y} + \frac{\bar{q}S_w\bar{c}}{I_y}C_m \tag{2.9}$$

$$\dot{r} = \frac{I_x}{I_xI_z - I_{zx}^2}F_2 + \frac{I_{zx}}{I_xI_z - I_{zx}^2}F_1 \tag{2.10}$$

$$\dot{\ell}_{xz} = r\ell_{yz} - q\ell_{zz} \tag{2.11}$$

$$\dot{\ell}_{yz} = -r\ell_{xz} + p\ell_{zz} \tag{2.12}$$

$$\dot{\ell}_{zz} = q\ell_{xz} - p\ell_{yz} \tag{2.13}$$

where

$$F_1 = (I_y - I_z)qr + I_{zx}pq + \bar{q}S_wb_wC_l \tag{2.14}$$

$$F_2 = (I_x - I_y)pq - I_{zx}qr + \bar{q}S_wb_wC_n \tag{2.15}$$

The nondimensional aerodynamic forces and moments, C_X, C_Y, C_Z, C_l, C_m, and 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. 27). The form of the aerodynamic model equations is

$$y(t) = \theta_0 + \theta_1x_1 + \theta_2x_2 + \cdots + \theta_{n_p-1}x_{n_p-1} \tag{2.16}$$

or in vector form

$$y_i = \mathbf{X}_i\boldsymbol{\theta} \tag{2.17}$$

where $y(t)$ or y_i represents 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 the trim forces or moments corresponding to an initial trimmed flight condition are represented by θ_0. The x_1 to x_{n_p-1} represent any functions of the state and control variables chosen for the model. The row vector $\mathbf{X}_i$ is given as

$$\mathbf{X}_i = \begin{bmatrix}1 & x_1 & x_2 & \cdots & x_{n_p-1}\end{bmatrix} \tag{2.18}$$

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

The following output equations, used in this study, reflect the measurements commonly available from flight tests:

$$V = \sqrt{u^2 + v^2 + w^2} \tag{2.19}$$

$$\beta = \sin^{-1}\left(\frac{v}{V}\right) \tag{2.20}$$

$$\alpha = \tan^{-1}\left(\frac{w}{u}\right) \tag{2.21}$$

$$\Theta = \sin^{-1}(-\ell_{xz}) \tag{2.22}$$

$$\phi = \tan^{-1}\left(\frac{\ell_{yz}}{\ell_{zz}}\right) \tag{2.23}$$

$$a_x = \frac{1}{g}(\dot{u} + qw - rv - g\ell_{xz}) \tag{2.24}$$

$$a_y = \frac{1}{g}(\dot{v} + ru - pw - g\ell_{yz}) \tag{2.25}$$

$$a_z = \frac{1}{g}(\dot{w} + pv - qu - g\ell_{zz}) \tag{2.26}$$

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 20. Thus, the states, outputs, and inputs for the two cases are given as follows:

For the longitudinal case,

$$\mathbf{X}_s = [u\ w\ q\ \ell_{xz}\ \ell_{yz}\ \ell_{zz}]^T \tag{2.27}$$

$$\mathbf{Y} = [V\ \alpha\ q\ \Theta\ a_x\ a_z]^T \tag{2.28}$$

$$\mathbf{U} = [\delta_e\ \beta_E\ v_E\ p_E\ r_E\ \phi_E]^T \tag{2.29}$$

For the lateral case,

$$\mathbf{X} = [v\ p\ r\ \ell_{xz}\ \ell_{yz}\ \ell_{zz}]^T \tag{2.30}$$

$$\mathbf{Y} = [\beta\ p\ r\ \phi\ a_y]^T \tag{2.31}$$

$$\mathbf{U} = [\delta_a\ \delta_r\ u_E\ w_E\ \Theta_E\ q_E\ \alpha_E]^T \tag{2.32}$$

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 that can be classified as adequate. 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 primary 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, significant terms among the candidate variables are determined and corresponding parameters are estimated. 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, a new variable entering the model and the variables incorporated into the model in previous stages are reexamined. Any variable providing 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 28. Two procedures utilizing the algorithm in reference 11 offer certain advantages by capitalizing on the use of splines for modeling (ref. 27) and large amplitude maneuvers for efficient data analysis (ref. 29).

The stepwise regression procedure in reference 11 provides a very efficient method of determining aerodynamic model structure and initial parameter estimates. This makes the computationally demanding problem of modeling the entire flight envelope more tractable. Once the aerodynamic model structure is determined, a more advanced method is needed to improve the biased parameter estimates obtained in the regression. More advanced methods are discussed in the next 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 (CR) lower bounds) from flight data. The ML algorithm is used with various optimization schemes, which are described in section IV. LR is used for three different applications in this study: (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 previous section. The third application was accomplished using an algorithm developed in this study.

A. Linear Regression

Linear regression analysis is a part of statistical theory that 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 is 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. A LS problem can be solved in a batch mode or recursive mode and both modes have application for determining the sensitivities in MNRES. A comprehensive discussion of regression analysis is given in reference 28.

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

$$y_i = \mathbf{X}_i\mathbf{S} + e_i \qquad (i = 1, 2, \cdots, N) \tag{3.1}$$

where y_i is the ith value of one response variable; $\mathbf{X}_i$ is the ith vector of predictor variables; $\mathbf{S}$ 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 modeling error. However, no assumptions are made about the statistical properties of e_i (see ref. 28). In application to MNRES, y_i represents one element of the output vector $\mathbf{Y}(\boldsymbol{\theta})$; $\mathbf{X}_i$ represents the ith set of values for the vector of unknown parameters $\boldsymbol{\theta}$; and $\mathbf{S}$ is the n_p-dimensional vector of coefficients to be computed. If a first-degree n_p-polynomial expansion is chosen for $\mathbf{X}_i$, then the elements of $\mathbf{S}$ are the desired sensitivities (slopes).

Applying the LS criterion, which requires minimization of the mean square error, gives the cost function as

$$J(\mathbf{S}) = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (y_i - \mathbf{X}_i\mathbf{S})^2 \tag{3.2}$$

and minimization requires that

$$\frac{\partial J(\mathbf{S})}{\partial \mathbf{S}} = 0 = \sum_{i=1}^{N} \mathbf{X}_i^T (y_i - \mathbf{X}_i\mathbf{S}) \tag{3.3}$$

Solving for $\mathbf{S}$ gives

$$\hat{\mathbf{S}} = \left(\sum_{i=1}^{N} \mathbf{X}_i^T \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i^T y_i \tag{3.4}$$

2. Recursive Processing

Recursive processing significantly reduces 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 LS 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 LS recursive algorithm is designed specifically for the MNRES algorithm. MNRES requires the parameters to be updated by including new information while removing 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 LS problem formulated as

$$\mathbf{Y} = \mathbf{X}\mathbf{S} + \mathbf{e} \tag{3.5}$$

where $\mathbf{Y}$ is a vector of n data points on a surface to be fit by the model given as $\mathbf{XS}$, $\mathbf{S}$ is a vector of n_p

unknown coefficients (slopes), and $\mathbf{X}$ is an n by n_p matrix defined as

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n_p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n_p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nn_p} \end{bmatrix} \tag{3.6}$$

LS estimation gives a solution as

$$\hat{\mathbf{S}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \qquad (n \geq n_p) \tag{3.7}$$

Now define a recursive relation for the $(k+1)$th iteration

$$\mathbf{P}_{k+1} = \left(\mathbf{X}_k^T\mathbf{X}_k\right)^{-1} \tag{3.8}$$

and the updating equation as

$$\mathbf{P}_{k+1} = \left(\mathbf{P}_k^{-1} + \mathbf{a}_k^T\mathbf{a}_k - \mathbf{a}_k^{oT}\mathbf{a}_k^o\right)^{-1} \tag{3.9}$$

where the row vector $\mathbf{a}_k$ is the new set of $\boldsymbol{\theta}$ to be included in $\mathbf{X}$ and $\mathbf{a}_k^o$ is the outgoing set of $\boldsymbol{\theta}$, which produced the highest value of the cost function. The recursive relation for $\mathbf{S}$ is

$$\mathbf{S}_{k+1} = \mathbf{S}_k - \mathbf{P}_{k+1}\left[\mathbf{a}_k^{oT}y_k^o - \mathbf{a}_k^T y_k + \left(\mathbf{a}_k^T\mathbf{a}_k - \mathbf{a}_k^{oT}\mathbf{a}_k^o\right)\mathbf{S}_k\right] \tag{3.10}$$

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

The derivation for equations (3.9) and (3.10) is as follows. Define $\mathbf{Z}$ to be the common elements of $\mathbf{X}$ between two iterations. Partitioning $\mathbf{X}$ for the $(k-1)$th and kth iterations results in

$$\mathbf{X}_{k-1} = \begin{bmatrix} \mathbf{a}_k^o \\ \mathbf{Z}_k \end{bmatrix} \tag{3.11}$$

$$\mathbf{X}_k = \begin{bmatrix} \mathbf{a}_k \\ \mathbf{Z}_k \end{bmatrix} \tag{3.12}$$

From equations (3.11) and (3.12),

$$\mathbf{X}_{k-1}^T\mathbf{X}_{k-1} = \mathbf{a}_k^{oT}\mathbf{a}_k^o + \mathbf{Z}_k^T\mathbf{Z}_k \tag{3.13}$$

$$\mathbf{X}_k^T\mathbf{X}_k = \mathbf{a}_k^T\mathbf{a}_k + \mathbf{Z}_k^T\mathbf{Z}_k \tag{3.14}$$

From equation (3.13),

$$\mathbf{Z}_k^T\mathbf{Z}_k = \mathbf{X}_{k-1}^T\mathbf{X}_{k-1} - \mathbf{a}_k^{oT}\mathbf{a}_k^o \tag{3.15}$$

Substituting equation (3.15) into (3.14) gives

$$\mathbf{X}_k^T\mathbf{X}_k = \mathbf{X}_{k-1}^T\mathbf{X}_{k-1} - \mathbf{a}_k^{oT}\mathbf{a}_k^o + \mathbf{a}_k^T\mathbf{a}_k \tag{3.16}$$

Substituting equation (3.16) into (3.8) gives

$$\mathbf{P}_{k+1} = \left(\mathbf{X}_{k-1}^T\mathbf{X}_{k-1} - \mathbf{a}_k^{oT}\mathbf{a}_k^o + \mathbf{a}_k^T\mathbf{a}_k\right)^{-1} \tag{3.17}$$

which can also be written as equation (3.9).

Applying the same development to equation (3.7) gives

$$\mathbf{S}_{k+1} = \mathbf{P}_{k+1}\left(\mathbf{X}_{k-1}^T\mathbf{Y}_{k-1} - \mathbf{a}_k^{oT}y_k^o + \mathbf{a}_k^T y_k\right) \tag{3.18}$$

and substituting equations (3.7) and (3.8) delayed a step into equation (3.18) gives

$$\mathbf{S}_{k+1} = \mathbf{P}_{k+1}\left(\mathbf{P}_k^{-1}\mathbf{S}_k - \mathbf{a}_k^{oT}y_k^o + \mathbf{a}_k^T y_k\right) \tag{3.19}$$

Expanding equation (3.19) gives

$$\mathbf{S}_{k+1} = \mathbf{P}_{k+1}\mathbf{P}_k^{-1}\mathbf{S}_k - \mathbf{P}_{k+1}\mathbf{a}_k^{oT}y_k^o + \mathbf{P}_{k+1}\mathbf{a}_k^T y_k \tag{3.20}$$

Note that

$$\mathbf{S}_k - \mathbf{P}_{k+1}\mathbf{P}_{k+1}^{-1}\mathbf{S}_k = 0 \tag{3.21}$$

and then add equation (3.20) to (3.21) to obtain

$$\mathbf{S}_{k+1} = \mathbf{S}_k - \mathbf{P}_{k+1}\mathbf{P}_{k+1}^{-1}\mathbf{S}_k + \mathbf{P}_{k+1}\mathbf{P}_k^{-1}\mathbf{S}_k - \mathbf{P}_{k+1}\mathbf{a}_k^{oT}y_k^o + \mathbf{P}_{k+1}\mathbf{a}_k^T y_k \tag{3.22}$$

Combining terms in equation (3.22) gives

$$\mathbf{S}_{k+1} = \mathbf{S}_k - \mathbf{P}_{k+1}\left[\left(\mathbf{P}_{k+1}^{-1} - \mathbf{P}_k^{-1}\right)\mathbf{S}_k + \mathbf{a}_k^{oT}y_k^o - \mathbf{a}_k^T y_k\right] \tag{3.23}$$

and substituting equation (3.9) into (3.23) yields equation (3.10).

3. Statistical Properties of LS Estimates

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

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

$$\mathbf{Y}_t = \mathbf{X}_t\boldsymbol{\theta} + \mathbf{w}_p \tag{3.24}$$

where $\mathbf{Y}_t$ is an N by 1 vector of the response variable for the N data measurement points, subscript t indicating that this is the true value of the variable without measurement noise, $\mathbf{X}_t$ is an N by n_p matrix of the state and input variables; $\boldsymbol{\theta}$ is the n_p by 1 vector of unknown parameters; and $\mathbf{w}_p$ is the N by 1 vector of process noise. The measurements provide

$$\mathbf{Y} = \mathbf{Y}_t + \mathbf{v}_y \tag{3.25}$$

$$\mathbf{X} = \mathbf{X}_t + \mathbf{v}_x \tag{3.26}$$

where $\mathbf{v}_y$ and $\mathbf{v}_x$ are the measurement noise in $\mathbf{Y}$ and $\mathbf{X}$. The process noise and measurement noise are typically assumed to be stationary, zero mean, and independent random processes. The solution to this LS problem (from the last section) is

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{3.27}$$

Premultiplying equation (3.24) by $[\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T$ and substituting equations (3.25), (3.26), and (3.27) results in

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{w}_p + \mathbf{v}_y - \mathbf{v}_x\boldsymbol{\theta}) \tag{3.28}$$

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

$$E\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\} = -E\{(\mathbf{X}^T X)^{-1}\mathbf{X}^T\mathbf{v}_x\}\boldsymbol{\theta} \tag{3.29}$$

and the covariance matrix is

$$\begin{aligned}\mathrm{cov}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = {} & E\{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\} \\ & + E\{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{v}_x \\ & \boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{v}_x\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\}\end{aligned} \tag{3.30}$$

where

$$\boldsymbol{\varepsilon} = \mathbf{w}_p + \mathbf{v}_y$$

From these equations it can be seen that the LS estimator is biased even if the measurement noise, $\mathbf{v}_x$ and $\mathbf{v}_y$, and the process noise, $\mathbf{w}_p$, are zero mean and independent. Only with the additional assumption that $\mathbf{X}$ is known without error (i.e., $\mathbf{v}_x = 0$), as might be the case in a curve-fitting problem, will the estimates be unbiased. Note that the covariance matrix is affected by all the measurement and process noise.

B. Maximum Likelihood

General parameter estimation for an airplane involves solving the nonlinear estimation problem in the presence of both process and measurement noise while modeling the airplane with the coupled, nonlinear equations of motion. One of the advanced techniques commonly used for this problem because of its superior statistical properties is maximum likelihood (ML).

1. Algorithm Development

Assume that the outcome of an experiment is N observations of the n_o by 1 output vector $\mathbf{Z}_i$ for $i = 1, 2, \ldots, N$, which depends on the unknown parameter vector $\boldsymbol{\theta}$. In general, the unknown parameters are the aerodynamic parameters, initial conditions, and measurement and process noise statistics. Let $f(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_N/\boldsymbol{\theta})$ be the conditional probability density function for the measurements given $\boldsymbol{\theta}$. The maximum likelihood estimate is the estimate for which the outcome of the experiment $\mathbf{Z}_i$ for $i = 1, 2, \ldots, N$ is most likely to occur; that is, the probability density is maximized with respect to $\boldsymbol{\theta}$. The ML estimate can be expressed as

$$\hat{\boldsymbol{\theta}} \ni \max_{\boldsymbol{\theta}} f(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_N/\boldsymbol{\theta}) = f(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_N/\hat{\boldsymbol{\theta}}) \tag{3.31}$$

Using the property of joint conditional probability density functions that

$$f(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3/\boldsymbol{\theta}) = f(\mathbf{Z}_3/\mathbf{Z}_2, \mathbf{Z}_1, \boldsymbol{\theta}) f(\mathbf{Z}_2/\mathbf{Z}_1, \boldsymbol{\theta}) f(\mathbf{Z}_1/\boldsymbol{\theta}) \tag{3.32}$$

allows the density function to be written as

$$\begin{aligned} f(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_N/\boldsymbol{\theta}) = {} & f(\mathbf{Z}_N/\mathbf{Z}'_{N-1}, \boldsymbol{\theta}) \\ & \times f(\mathbf{Z}_{N-1}/\mathbf{Z}'_{N-2}, \boldsymbol{\theta}) \\ & \times \cdots f(\mathbf{Z}_2/\mathbf{Z}_1, \boldsymbol{\theta}) f(\mathbf{Z}_1/\boldsymbol{\theta}) \end{aligned} \tag{3.33}$$

$$= \prod_{i=1}^{N} f(\mathbf{Z}_i/\mathbf{Z}'_{i-1}, \boldsymbol{\theta}) \tag{3.34}$$

where

$$\mathbf{Z}'_{i-1} = \mathbf{Z}_{i-1}, \mathbf{Z}_{i-2}, \ldots, \mathbf{Z}_1 \tag{3.35}$$

If the $\mathbf{Z}_i$ measurements are treated as fixed, the density function becomes a function of $\boldsymbol{\theta}$ only. This function is usually referred to as the likelihood function, that is,

$$L_1(\boldsymbol{\theta}) = \prod_{i=1}^{N} f(\mathbf{Z}_i/\mathbf{Z}'_{i-1}, \boldsymbol{\theta}) \tag{3.36}$$

Consequently, the problem of finding a maximum likelihod estimate becomes the problem of finding the $\boldsymbol{\theta}$ that maximizes the likelihood function.

To define the likelihood function, the distribution of the measurements given $\boldsymbol{\theta}$ must be defined. If the distribution of the measurement and process noise is Gaussian, then the distribution of the measurements

given θ is also Gaussian and can be uniquely determined by computing the mean and covariance. The mean is

$$E\{\mathbf{Z}_i/\mathbf{Z}'_{i-1},\theta\} = \hat{\mathbf{Y}}_{i/i-1}(\theta) \tag{3.37}$$

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

$$\begin{aligned}\text{cov}\{\mathbf{Z}_i/\mathbf{Z}'_{i-1},\theta\} &= E\{(\mathbf{Z}_i - \hat{\mathbf{Y}}_{i/i-1})(\mathbf{Z}_i - \hat{\mathbf{Y}}_{i/i-1})^T\} \\ &= E\{\boldsymbol{\nu}_i\boldsymbol{\nu}_i^T\} = \mathbf{B}_{i/i-1}(\theta)\end{aligned} \tag{3.38}$$

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

$$\boldsymbol{\nu}_i = \mathbf{Z}_i - \hat{\mathbf{Y}}_i(\hat{\theta}) \tag{3.39}$$

where $\hat{\mathbf{Y}}_i$ is the predicted value of the output vector at time t_i. Without process noise the conditional subscript is not needed since the Kalman filter update is zero. Also, assuming stationary statistics, the mean and covariance of the residuals may be written as

$$E\{\boldsymbol{\nu}_i\} = 0 \tag{3.40}$$

$$E\{\boldsymbol{\nu}_i\boldsymbol{\nu}_j^T\} = \mathbf{R}\delta_{ij} \tag{3.41}$$

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

$$f(\mathbf{Z}_i/\mathbf{Z}'_{i-1},\theta)(= (2\pi)^{-n_o/2}|\mathbf{R}|^{-1/2}\exp\left(-\frac{1}{2}\boldsymbol{\nu}_i^T\mathbf{R}^{-1}\boldsymbol{\nu}_i\right) \tag{3.42}$$

Hence, using equation (3.42), the likelihood function is

$$\begin{aligned}L_1(\theta) &= \prod_{i=1}^{N} f(\mathbf{Z}_i/\mathbf{Z}'_{i=1},\theta) \\ &= (2\pi)^{-Nn_o/2}\prod_{i=1}^{N}|\mathbf{R}|^{-1/2} \\ &\quad\times\exp\left(-\frac{1}{2}\sum_{i=1}^{N}\boldsymbol{\nu}_i^T\mathbf{R}^{-1}\boldsymbol{\nu}_i\right)\end{aligned} \tag{3.43}$$

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

$$\begin{aligned}L(\theta) &= \frac{1}{2}\sum_{i=1}^{N}(\mathbf{Z}_i - \hat{\mathbf{Y}}_i)^T\mathbf{R}^{-1}(\mathbf{Z}_i - \hat{\mathbf{Y}}_i) \\ &\quad + \frac{N}{2}\ln|\mathbf{R}| + \text{ Constant}\end{aligned} \tag{3.44}$$

The unknown $\mathbf{R}$ can be estimated by minimization of the likelihood function with respect to $\mathbf{R}$. This produces

$$\begin{aligned}\hat{\mathbf{R}} &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{Z}_i - \hat{\mathbf{Y}}_i)(\mathbf{Z}_i - \hat{\mathbf{Y}}_i)^T \\ &= \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\nu}_i\boldsymbol{\nu}_i^T\end{aligned} \tag{3.45}$$

After substituting the estimate $\hat{\mathbf{R}}$ into equation (3.44), the final form of the cost function, as used in this study, is obtained:

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{N}(\mathbf{Z}_i - \hat{\mathbf{Y}}_i)^T\hat{\mathbf{R}}^{-1}(\mathbf{Z}_i - \hat{\mathbf{Y}}_i) + \text{ Constant} \tag{3.46}$$

The cost function given by equation (3.46) is the same as that used in an output error technique (ref. 12) 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 (3.46) must be minimized with respect to the unknown parameter vector θ. The unknown parameters determined by this method are, for this study, 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 (CR) lower bounds, providing a measure of the maximum achievable accuracy for the parameters. These are defined by the CR inequality

$$E\{(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^T\} \geq -E\left\{\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right\}^{-1} \tag{3.47}$$

where

$$-E\left\{\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right\} = \mathbf{M} \tag{3.48}$$

and $\mathbf{M}$ is usually referred to as the Fisher information matrix. It is assumed in this study that the approximated Hessian matrix $\mathbf{H}$ 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 (ref. 13) for the kth iteration

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k - \hat{\mathbf{H}}_k^{-1}\mathbf{g}_k \tag{3.49}$$

where

$$\mathbf{g}_k = \left.\frac{\partial J(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_0} \quad \text{and} \quad \hat{\mathbf{H}} \approx \mathbf{M}$$

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. 30) have shown that the ML method produces acceptable estimates in many situations. The large sample properties of an ML estimate are summarized as follows (Cramér, ref. 31, derives these properties):

1. Asymptotically unbiased:

$$\lim_{N\to\infty} E\{\hat{\boldsymbol{\theta}}\} = \boldsymbol{\theta}$$

2. Asymptotically efficient:

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

3. Consistent:

$$\lim_{N\to\infty} \Pr\{(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) \leq \varepsilon\} = 1$$

   with ε arbitrarily small

4. Asymptotically normal:

$$\hat{\boldsymbol{\theta}} = N(\boldsymbol{\theta}, \mathbf{M}^{-1})$$

Asymptotic unbiasedness and consistency are very similar. However, consistency implies that if an estimator is consistent for $\boldsymbol{\theta}$, it is also consistent for any well-behaved function of $\boldsymbol{\theta}$. Thus, consistency is more significant than unbiasedness. Asymptotic efficiency is a statement involving the CR inequality; therefore, for large samples the CR lower bounds are obtained. Asymptotic normality states that as the sample size gets very large, the estimates $\boldsymbol{\theta}$ approach a normal distribution with mean and covariance given by $\boldsymbol{\theta}$ and $\mathbf{M}^{-1}$, respectively.

IV. OPTIMIZATION TECHNIQUES

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

1. $J(\theta)$ is differentiable at θ^*.
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 for the cost functions that derivative methods need. In many unconstrained, nonlinear programming problems, however, derivative methods converge faster (ref. 13), particularly for estimation problems involving dynamic systems. This was demonstrated for airplane systems both in this study and in reference 32.

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 using only 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 methods are often used because they eliminate the additional burden of deriving and incorporating sensitivity equations into the algorithm. However, the finite-difference methods require about the same 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 good indicators of the relative merit of MNRES. The two optimization methods are the flexible polyhedron search (FPS) and the modified Newton-Raphson method (MNR). More details are provided in references 13 and 33 for FPS and in references 12 and 34 for MNR. In a variation of MNR the derivative information is computed by using finite differences (refs. 16 and 17). Both the finite-difference form and the sensitivity equation form of MNR are used in this study.

1. Flexible Polyhedron Search

Since FPS has been found to be advantageous in some aircraft design and control applications (ref. 35), 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 $J(\theta)$, a scalar function of n_p variables. The FPS method uses a flexible polyhedron surface or simplex with $n_p + 1$ vertices, each 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

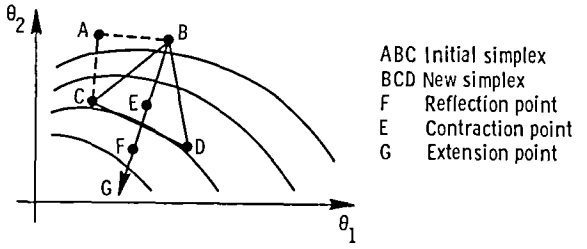


Figure 2. Two steps of a two-dimensional flexible polyhedron search.

with two unknown parameters. If the new vertex produces a lower cost than $\boldsymbol{\theta}_L$ (the vertex producing the smallest $J(\boldsymbol{\theta})$), 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 $J(\boldsymbol{\theta})$ and adding new projected points that produce lower $J(\boldsymbol{\theta})$. The flexible polyhedron is able to adapt to the shape of $J(\boldsymbol{\theta})$ by stretching down slopes, contracting near minima, and changing direction in curved valleys.

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 estimating derivatives is computationally burdensome, this information enables relatively fast convergence of the optimization process. In fact, Newton's method converges in one pass for cost functions that 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 of solving sensitivity equations is desirable, the MNR algorithm in this report uses a simple finite-difference method except when otherwise noted. This is not too costly in terms of computational time (refs. 16 and 17); 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 eqs. (2.3) and (2.4)) and the measurement equations,

$$\dot{\mathbf{X}}_s = f(\mathbf{X}_s, \mathbf{U}, \boldsymbol{\theta}) \qquad \mathbf{X}_s(0) = \mathbf{X}_{s0} \tag{2.3}$$

$$\mathbf{Y} = h(\mathbf{X}_s, \mathbf{U}, \boldsymbol{\theta}) \tag{2.4}$$

$$\mathbf{Z}_i = \mathbf{Y}_i + \mathbf{v}_i \qquad (i = 1, 2, \ldots, N) \tag{4.1}$$

with

$$E\{\mathbf{v}_i\} = 0 \quad \text{and} \quad E\{\mathbf{v}_i \mathbf{v}_j^T\} = \mathbf{R}\delta_{ij} \tag{4.2}$$

where $\mathbf{X}_s$, $\mathbf{U}$, and $\mathbf{Y}$ are the state, input, and output vectors, respectively; $\boldsymbol{\theta}$ is the unknown parameter vector; $\mathbf{Z}_i$ and $\mathbf{v}_i$ are the measurement vector and measurement noise vector, respectively, at $t = t_i$; $\mathbf{R}$ is a diagonal measurement noise covariance matrix, which under the above assumptions is approximated by the covariance matrix of the residuals. Without process noise the ML cost function to be minimized is given by equation (3.46) from which the added constant and multiplicative factor of 1/2 are dropped without affecting the solution:

$$J(\boldsymbol{\theta}) = \sum_{i=1}^{N} (\mathbf{Z}_i - \hat{\mathbf{Y}}_i)^T \hat{\mathbf{R}}^{-1} (\mathbf{Z}_i - \hat{\mathbf{Y}}_i) \tag{3.46}$$

The matrix $\hat{\mathbf{R}}$ is given by equation (3.45):

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\nu}_i \boldsymbol{\nu}_i^T \tag{3.45}$$

where

$$\boldsymbol{\nu}_i = \mathbf{Z}_i - \hat{\mathbf{Y}}_i(\boldsymbol{\theta}_0) \tag{4.3}$$

and $\boldsymbol{\theta}_0$ is the initial estimate of the unknown parameter vector. The MNR method accomplishes the minimization by expanding $\mathbf{Y}$, the computed output vector, about $\boldsymbol{\theta}_0$, the initial unknown parameter vector. A Taylor series expansion of $\mathbf{Y}$ truncated to first order is

$$\hat{\mathbf{Y}}(\boldsymbol{\theta}) = \hat{\mathbf{Y}}(\boldsymbol{\theta}_0) + \left.\frac{\partial \mathbf{Y}}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_0} \Delta\boldsymbol{\theta} \tag{4.4}$$

where $\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_0$. Then substituting this expression into equation (3.46) results in a quadratic approximation of J. The increment $\Delta\boldsymbol{\theta}$ is the unknown. Differentiating J with respect to $\boldsymbol{\theta}$ and equating the

derivative to zero to find a minimum results in

$$\frac{\partial J}{\partial \boldsymbol{\theta}} = -\sum_{i=1}^{N} \mathbf{G}_i^T \hat{\mathbf{R}}^{-1} \boldsymbol{\nu}_i + \sum_{i=1}^{N} \mathbf{G}_i \hat{\mathbf{R}}^{-1} \mathbf{G}_i \, \Delta\boldsymbol{\theta} = 0 \tag{4.5}$$

where

$$\mathbf{G}_i = \left[\frac{\partial y_k}{\partial \theta_\ell}\right]_i \tag{4.6}$$

and y_k and θ_ℓ are the kth and ℓth elements of the $\mathbf{Y}$ and $\boldsymbol{\theta}$ vectors, respectively. Solving for $\Delta\boldsymbol{\theta}$ gives

$$\Delta\hat{\boldsymbol{\theta}} = \left(\sum_{i=1}^{N} \mathbf{G}_i^T \hat{\mathbf{R}}^{-1} \mathbf{G}_i\right)^{-1} \sum_{i=1}^{N} \mathbf{G}_i^T \hat{\mathbf{R}}^{-1} \boldsymbol{\nu}_i \tag{4.7}$$

This is often written as

$$\Delta\hat{\boldsymbol{\theta}} = -\mathbf{M}^{-1} \frac{\partial J}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}_0} \tag{4.8}$$

emphasizing the Fisher information matrix $\mathbf{M}$ and gradient terms. For the kth iteration the estimate $\hat{\boldsymbol{\theta}}_{k+1}$ is given as $\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + \Delta\hat{\boldsymbol{\theta}}_{k+1}$. In this study, convergence is achieved when $\Delta J_k/J_k$ and $\Delta\boldsymbol{\theta}_k/\boldsymbol{\theta}_k$ are less than 0.001. The sensitivities $\mathbf{G}_i$ are determined separately from the above steps. This may be done by integrating the sensitivity equations, by using a finite-difference approximation, or by using MNRES.

B. MNRES Method

The MNRES method developed in this report 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)–(4.8) and (3.45)–(3.46)) apply for ML/MNRES; however, the sensitivities $\mathbf{G}_i$ are computed by using slope information from local surface approximations of $\mathbf{Y}(\boldsymbol{\theta})$. The approximations are made near the series expansion point of equation (4.4). The sensitivities obtained from the fitted surface are determined with less computational effort than that required by either a finite-difference method or integration of analytically determined sensitivity equations.

The MNRES algorithm is readily optimized for a particular application because the user can select both the type of surface and the method of fitting the surface. Two very practical types of surfaces in aeronautical applications are nth-order polynomials and splines. Two efficient methods of fitting the surface are by solving n_p simultaneous equations for n_p unknowns (algebraic solution) and by solving a redundant set of equations for n_p unknowns (least squares solution). The trade-offs in choosing a surface and a surface-fitting method involve the choice between accuracy of the sensitivities and computational effort.

1. Algebraic Solution

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

$$y_{ki}^j(\boldsymbol{\theta}^j) = s_{k0} + s_{k1}\theta_1^j + \cdots + s_{kn_p}\theta_{n_p}^j \tag{4.9}$$

where i indicates the ith point in time; k indicates the kth element of the output vector $\mathbf{Y}(\boldsymbol{\theta})$; and j indicates one of the $n_p + 1$ sample points used to fit equation (4.9) to $\mathbf{Y}(\boldsymbol{\theta})$. Note that

$$y_k^j(\boldsymbol{\theta}^j) = y_k(\boldsymbol{\theta}) \tag{4.10}$$

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 sensitivities 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 and thus provides higher quality derivative calculations. This alternative is discussed further at the end of this section. The slopes s_{k1} to s_{kn_p} are the desired sensitivities, $(\partial y_k/\partial\theta_\ell)_i$, and s_{k0} is the value of $y_k(\boldsymbol{\theta})$ at the series expansion point of equation (4.4). 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 $y_k(\boldsymbol{\theta})$, the slopes vary across the fitted surface and, therefore, must be evaluated specifically at s_{k0}. Consider the matrix representation of equation (4.9) for the first element of $\mathbf{Y}$ and for the $n_p + 1$ sample points at time t_i:

$$\mathbf{Y}_{1i} = \mathbf{X}\mathbf{S}_{1i} \tag{4.11}$$

Note that y_k is the kth element of the output vector $\mathbf{Y}$ and y_k^j is the jth element of the surface-fitting vector, $\mathbf{Y}_k$. Matrix $\mathbf{X}$ contains $n_p + 1$ rows defining the $n_p + 1$ sample points. Expanding equation (4.11)

to show the vector and matrix elements gives

$$\begin{bmatrix} y_{1i}^0 \\ y_{1i}^1 \\ \vdots \\ y_{1i}^n \end{bmatrix} = \begin{bmatrix} 1 & \theta_1^0 & \theta_2^0 & \cdots & \theta_n^0 \\ 1 & \theta_1^1 & \theta_2^1 & \cdots & \theta_n^1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \theta_1^n & \theta_2^n & \cdots & \theta_n^n \end{bmatrix} \begin{bmatrix} s_{10} \\ s_{11} \\ \vdots \\ s_{1n} \end{bmatrix}_i \tag{4.12}$$

where $n = n_p$ for the computationally least demanding approach, or algebraic solution. In general, n may be greater than n_p producing a least squares solution. Since s_{10} is known, equation (4.12) can be simplified. The first line in equation (4.12) can be eliminated by subtracting it from the other n_p equations. Thus,

$$\Delta \mathbf{Y}_{1i} = \Delta \mathbf{X} \mathbf{S}_{1i} \tag{4.13}$$

or in matrix notation

$$\begin{bmatrix} y_{1i}^1 - y_{1i}^0 \\ y_{1i}^2 - y_{1i}^0 \\ \vdots \\ y_{1i}^n - y_{1i}^0 \end{bmatrix} = \begin{bmatrix} \Delta\theta_1^1 & \Delta\theta_2^1 & \cdots & \Delta\theta_n^1 \\ \Delta\theta_1^2 & \Delta\theta_2^2 & \cdots & \Delta\theta_n^1 \\ \vdots & \vdots & \ddots & \vdots \\ \Delta\theta_1^n & \Delta\theta_2^n & \cdots & \Delta\theta_n^n \end{bmatrix} \begin{bmatrix} s_{11} \\ s_{12} \\ \vdots \\ s_{1n} \end{bmatrix}_i \tag{4.14}$$

where

$$\Delta\theta_\ell^j = \theta_\ell^j - \theta_\ell^0$$

Thus, at time t_i, the sensitivities for the first element in $\mathbf{Y}$ are given by

$$\mathbf{S}_{1i} = (\Delta \mathbf{X})^{-1} \Delta \mathbf{Y}_{1i} \tag{4.15}$$

Because the $\Delta\mathbf{X}$ matrix is independent of time, the sensitivities can 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 $n_s n_p$ (number of states times number of parameters) sensitivity equations has been replaced by a set of n_o (number of outputs) matrix multiplications.

Figure 3 shows, geometrically, two iterations for the case where $\boldsymbol{\theta}$ is two-dimensional and a linear surface is used to fit a scalar y. The expansion at time t_i is

$$\begin{bmatrix} y_i^1 - y_i^0 \\ y_i^2 - y_i^0 \end{bmatrix} = \begin{bmatrix} \Delta\theta_1^1 & \Delta\theta_2^1 \\ \Delta\theta_1^2 & \Delta\theta_2^2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}_i \tag{4.16}$$

During the first iteration, this expansion requires that $y(\boldsymbol{\theta})$ be evaluated at $n_p + 1 = 3$ points: y^0, y^1, and y^2. Computationally, the first iteration is the most costly phase of the MNRES algorithm. Each

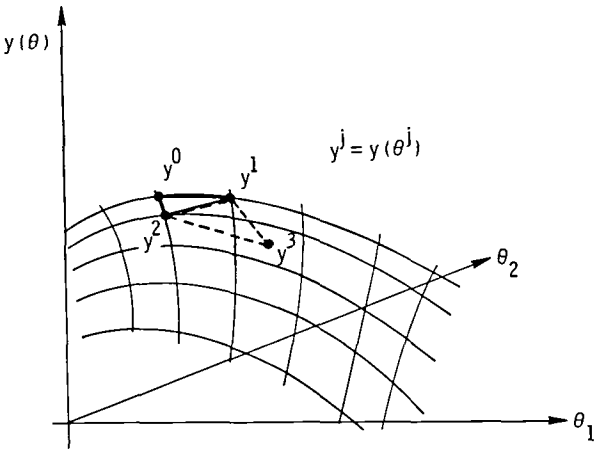


Figure 3. Linear-surface fit for two iterations of MNRES.

evaluation of $\mathbf{Y}$ requires that the equations of motion be integrated. The linear surface (indicated by the solid triangle in fig. 3) is fit and the slopes (sensitivities) are thereby determined. The algorithm proceeds, as in the ordinary MNR method, by using equation (4.8) to obtain

$$\hat{\boldsymbol{\theta}}_{r+1} = \hat{\boldsymbol{\theta}}_r + \Delta\hat{\boldsymbol{\theta}}_{r+1} \tag{4.17}$$

The estimated sensitivity values $s_{k\ell}$ are used to define the elements of matrix $\mathbf{G}_i$ in equation (4.7). The new $\mathbf{Y}$ is evaluated (by integration of equations of motion) at $\hat{\boldsymbol{\theta}}_{r+1}$ to get $y^3(\boldsymbol{\theta})$. At this point the MNRES algorithm has reduced the sensitivity problem to solving a set of simultaneous equations. This is done by eliminating the $\boldsymbol{\theta}^j$ in $\mathbf{X}$ which produced the greatest value of $J(\boldsymbol{\theta})$ and replacing that information with the newest estimate of $\boldsymbol{\theta}$. In the example in figure 3, y^0 was assumed to be the high cost point and so was eliminated to obtain the new fitted surface (indicated by dashed triangle). 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 that the new $\mathbf{Y}^j(\hat{\boldsymbol{\theta}}_{r+1})$ produces a smaller value of $J(\boldsymbol{\theta})$. Sometimes, 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 $\mathbf{Y}^j$. After this initializing pass, only one integration of the system equations is needed to evaluate the cost $J(\boldsymbol{\theta})$ and outputs $\mathbf{Y}(\boldsymbol{\theta})$ 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 $\boldsymbol{\theta}$ to obtain the corresponding perturbation in each element of $Y(\boldsymbol{\theta})$ is not optimum for

derivative calculations. Experience has shown that it is beneficial to use perturbation sizes that reflect the importance of the parameter to the model. By computing the sensitivities $\theta_\ell^2 M_{\ell\ell}$ 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 $\boldsymbol{\theta}$; however, because of the sometimes widely varying sensitivities of the parameters as well as the numerical precision limitations, it is beneficial to vary the perturbation size according to the aforementioned rule. The sensitivity defined as $\theta_\ell^2 M_{\ell\ell}$ was introduced in reference 36 and used again in reference 37 as a means of quantifying the significance of a parameter to the model.

2. Least Squares Solution

The least squares approach to fitting the surface $\mathbf{Y}(\boldsymbol{\theta})$ offers another advantage if a recursive least squares method is used. The recursive method reduces the storage requirements from $n_p + 1$ sets of output time histories to just two time histories: one corresponding to the new response predicted by the most recent estimate of $\boldsymbol{\theta}$ and the other corresponding to the outgoing $\boldsymbol{\theta}$ that produced the highest cost. The penalty for this advantage is the need to integrate equations of motion twice per iteration; still, the least squares solution 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 $\Delta\mathbf{X}$ matrix, and the second change is in the sensitivity calculation. The development of this formulation begins with equation (4.9) in condensed form (everything discussed up to this equation in the previous development applies here):

$$y_{ki}^j = \mathbf{X}^j \mathbf{S}_{ki} \tag{4.18}$$

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

$$\mathbf{Y}_k = \mathbf{X}\mathbf{S}_k \tag{4.19}$$

The least squares solution for the sensitivity vector is

$$\mathbf{S}_k = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}_k \tag{4.20}$$

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

$$\mathbf{P}_{r+1} = (\mathbf{X}_r^T\mathbf{X}_r)^{-1} \tag{4.21}$$

and the updating equation can be defined as

$$\mathbf{P}_{r+1} = (\mathbf{P}_r^{-1} + \mathbf{a}_r^T\mathbf{a}_r - \mathbf{a}_r^{oT}\mathbf{a}_r^o)^{-1} \tag{4.22}$$

where the row vector $\mathbf{a}_r$ is the new set of $\boldsymbol{\theta}$ to be included and $\mathbf{a}_r^o$ is the outgoing set of $\boldsymbol{\theta}$, which produced the high cost J. The recursive relation for $\mathbf{S}$, which has already been derived in section III.A, is now

$$\begin{aligned}\mathbf{S}_{r+1} = \mathbf{S}_r - \mathbf{P}_{r+1}[&\mathbf{a}_r^{oT} y_r^o - \mathbf{a}_r^T y_r \\ &+ (\mathbf{a}_r^T\mathbf{a}_r - \mathbf{a}_r^{oT}\mathbf{a}_r^o)\mathbf{S}_r]\end{aligned} \tag{4.23}$$

where y_r^o and y_r are the outgoing and incoming elements of $\mathbf{Y}_k$, respectively, at the rth iteration. With the new sensitivities determined, the algorithm proceeds as before.

3. Properties of MNRES

Properties of MNRES are discussed in comparison with the commonly used MNR algorithm. Thus, convergence characteristics and computational advantages and disadvantages of MNRES are compared with those of 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 $\Delta\boldsymbol{\theta}^j$ (the distance between points on the fitted surface for MNRES) becomes small enough, the sensitivities become identical to those given by a simple finite-difference method, regardless of the actual functional representation of $\mathbf{Y}(\boldsymbol{\theta})$. 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 those 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 n_p different $\boldsymbol{\theta}$ are chosen (a perturbation on each element of $\boldsymbol{\theta}$ is sufficient) to give n_p different response time histories $\mathbf{Y}(\boldsymbol{\theta})$. The surface given by $\mathbf{Y}(\boldsymbol{\theta})$ is fitted. The initial $\boldsymbol{\theta}^j$ can be chosen such that

$$|(\boldsymbol{\theta}^j - \boldsymbol{\theta}^0)/\boldsymbol{\theta}^0| \ll 1 \qquad (4.24)$$

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 $\Delta\boldsymbol{\theta}^j$ 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 $\boldsymbol{\theta}$ may be very different, as discussed later. Toward the end of optimization the MNRES scheme again becomes equivalent to a simple finite-difference scheme since the $\Delta\boldsymbol{\theta}$ become very small. This forces the surface that is fit to $\mathbf{Y}(\boldsymbol{\theta})$ 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, that is, after initialization and before convergence. During this part of the optimization, large $\Delta\boldsymbol{\theta}$ 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, so that the slopes or sensitivities no longer approximate the slope of the $\mathbf{Y}(\boldsymbol{\theta})$ surface at a "point," that is, 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 $\boldsymbol{\theta}$ which are far from $\boldsymbol{\theta}^*$, the optimal solution. This, in effect, contracts the expanding surface that is fitting $\mathbf{Y}(\boldsymbol{\theta})$ and balances the expansion process. As the updated estimates of $\boldsymbol{\theta}$ become close to $\boldsymbol{\theta}^*$, the contraction process dominates and slope (sensitivity) information approaches 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 generally improves the closer $\boldsymbol{\theta}$ gets to $\boldsymbol{\theta}^*$, and since initial estimates of $\boldsymbol{\theta}^*$ are often given by a least squares procedure or by a knowledgeable user, $\boldsymbol{\theta}_0$ tends to be "close" to $\boldsymbol{\theta}^*$. Thus, for aircraft estimation problems, MNRES generally begins in a region conducive to convergence. The third factor is that step-size control logic can always be incorporated. Carried to the extreme, MNRES could always be forced to produce the same derivative estimates as a finite-difference method. Of course, convergence would be very slow because of the very small steps. In practice, one lets the algorithm take steps determined by the NR logic (as done in this study); then if a convergence problem develops, step-size control can 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 during 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 (4.15), where $\Delta\mathbf{X}$ 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 during each pass is not equivalent between the two methods. MNRES performs less computation during each pass and, consequently, has less information with which to update the estimates. However, when sufficient passes have been performed to make the work done by MNRES equal to MNR, MNRES has already performed $n_p + 1$ parameter updates. This allows MNRES to step more quickly toward the final solution. MNRES achieves convergence faster than MNR as the cost function becomes more quadratic. 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 in 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 in which estimates of the unknowns are updated during each iteration. The updates are estimated by equation (4.8), 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 most 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 in terms of convergence, since there is no way to know for how many iterations the information matrix can be held 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 to be integrated during 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. The recursive least squares method discussed earlier reduces this storage requirement to two sets of time histories; however, the computational effort increases from n_s to $2n_s$ equations to be integrated during each pass. The user's computer system would dictate which approach is more appropriate.

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 that 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 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. 20). 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 modeling 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 Regions

Confidence regions are described by a surface in parameter space representing a certain confidence level. The surface is defined by a statistic reflecting 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 that 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 obviously varies according to the definition of the statistic. A useful statistic for composite hypothesis tests is created from the ratio of likelihood functions.

Let $Z_1, Z_2, \ldots, Z_N$ be N independent random variables with probability density functions

$$f(Z_i, \theta) \qquad (i = 1, 2, \ldots, N) \tag{5.1}$$

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

$$L(\Omega) = \prod_{i=1}^{N} f(Z_i, \theta) \quad (\theta \in \Omega) \tag{5.2}$$

$$L(\omega) = \prod_{i=1}^{N} f(Z_i, \theta) \quad (\theta \in \omega) \tag{5.3}$$

and denote the maxima of the likelihood functions as $L(\hat{\Omega})$ and $L(\hat{\omega})$; the likelihood ratio is defined as

$$\lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} \tag{5.4}$$

This ratio forms a statistic having 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, λ 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. 23) recommended that 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 on a correction to the linear problem, the following development begins with the linear case.

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 equations. The form of the linear estimation model (single output) is given by equation (3.5), repeated here with θ as the vector of unknown parameters (the number of measurements is taken to be N for this discussion):

$$\mathbf{Y} = \mathbf{X}\theta + \mathbf{e} \tag{5.5}$$

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

$$H_0\colon\ \theta_t = \theta \tag{5.6}$$

$$H_1: \ \theta_t \neq \theta \tag{5.7}$$

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

$$\lambda = \exp\left\{-\frac{1}{2\sigma^2}[J(\theta) - J(\hat{\theta})]\right\} \tag{5.8}$$

where

$$J(\theta) = (\mathbf{Y} - \mathbf{X}\theta)^T(\mathbf{Y} - \mathbf{X}\theta) \tag{5.9}$$

and

$$J(\hat{\theta}) = (\mathbf{Y} - \mathbf{X}\hat{\theta})^T(\mathbf{Y} - \mathbf{X}\hat{\theta}) \tag{5.10}$$

The statistic λ can be equivalently replaced by

$$\mu = J(\theta) - J(\hat{\theta}) \tag{5.11}$$

or in practice by the statistic

$$F_{\alpha_p} = \frac{N - n_p}{n_p}\frac{J(\theta) - J(\hat{\theta})}{J(\hat{\theta})} \tag{5.12}$$

where F_{α_p} 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. 28). In addition, for the linear estimation problem, it is known that (ref. 25):

1. $\hat{\theta}$ is an unbiased estimate of θ.
2. The Cramer-Rao bound is reached.

The confidence region R_c in parameter space can now be given as the set of θ for which

$$J(\theta) - J(\hat{\theta}) \leq n_p s^2 F_{\alpha_p}(n_p, N - n_p) \tag{5.13}$$

where $s^2 = J(\hat{\theta})/(N - n_p)$. Once the data are 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) in equation (5.5), the cost $J(\theta)$ can be expanded as

$$\begin{aligned} J(\theta) &= (\mathbf{Y} - \mathbf{X}\theta)^T(\mathbf{Y} - \mathbf{X}\theta) \\ &= \mathbf{Y}^T\mathbf{Y} - 2\theta^T\mathbf{X}^T\mathbf{Y} + \theta^T\mathbf{X}^T\mathbf{X}\theta \end{aligned} \tag{5.14}$$

Differentiating equation (5.14) with respect to θ and setting the derivative to zero, the normal equations are obtained

$$0 = -\mathbf{X}^T\mathbf{Y} + \mathbf{X}^T\mathbf{X}\theta \tag{5.15}$$

and the solution for θ is

$$\hat{\theta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \tag{5.16}$$

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

$$\begin{aligned} J(\theta) - J(\hat{\theta}) = \theta^T\mathbf{X}^T\mathbf{X}\theta - 2\theta^T\mathbf{X}^T\mathbf{Y} \\ + 2\hat{\theta}^T\mathbf{X}^T\mathbf{Y} - \hat{\theta}^T\mathbf{X}^T\mathbf{X}\hat{\theta} \end{aligned} \tag{5.17}$$

Substituting for $\mathbf{X}^T\mathbf{Y}$ from the normal equations gives

$$J(\theta) - J(\hat{\theta}) = (\theta - \hat{\theta})^T\mathbf{X}^T\mathbf{X}(\theta - \hat{\theta}) \tag{5.18}$$

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 depend 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 that 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 (5.13), 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 (5.18) into (5.13):

$$(\theta - \hat{\theta})^T\mathbf{X}^T\mathbf{X}(\theta - \hat{\theta}) \leq n_p s^2 F_{\alpha_p}(n_p, N - n_p) \tag{5.19}$$

The confidence limits are determined by realizing that 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 a nonzero vector $\mathbf{b}$ if and only if (see ref. 38)

$$|\mathbf{b}^T(\theta - \hat{\theta})| \leq (\mathbf{b}^T\mathbf{H}_\alpha^{-1}\mathbf{b})^{1/2} \tag{5.20}$$

where

$$\mathbf{H}_\alpha = [n_p s^2 F_{\alpha_p}(n_p, N - n_p)]^{-1}\mathbf{H} \tag{5.21}$$

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

$$|\mathbf{b}^T\theta - \mathbf{b}^T\hat{\theta}| \leq [n_p s^2 F_{\alpha_p}(n_p, N - n_p)(\mathbf{b}^T\mathbf{H}^{-1}\mathbf{b})]^{1/2} \tag{5.22}$$

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

$$|\theta_i - \hat{\theta}_i| \leq \sqrt{k} s \sqrt{d_{ii}} \tag{5.23}$$

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

$$\hat{\theta} - \sqrt{k}\sigma_{\theta} < \theta < \hat{\theta} + \sqrt{k}\sigma_{\theta} \tag{5.24}$$

where

$$k = n_p F_{\alpha_p}(n_p, N - n_p) \tag{5.25}$$

$$\sigma_{\theta_i} = s\sqrt{d_{ii}} \tag{5.26}$$

$$\mathbf{H}^{-1} = [d_{ij}] \tag{5.27}$$

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 change from those found in the linear case (ref. 28):

1. Assuming that $\mathbf{e}$ is normally distributed does not imply that $\boldsymbol{\theta}$ is normally distributed.
2. $s^2 = J(\hat{\boldsymbol{\theta}})/(N - n_p)$ 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.

Some results remain true, however:

1. The sum of squares, $J(\boldsymbol{\theta})$, 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 $J(\boldsymbol{\theta})$ still corresponds to finding a point in the estimation space closest to $(Z_1, Z_2, \ldots, Z_N)$.
3. Confidence regions can still be defined; however, the confidence level is an approximation.

In reference 23, 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 (5.13) becomes

$$J(\boldsymbol{\theta}) - J(\hat{\boldsymbol{\theta}}) = n_p s^2 F_{\alpha_p}(n_p, N - n_p) \times \left[1 + \frac{N(n_p + 2)}{(N - n_p)n_p} N_\phi\right] \tag{5.28}$$

and for the multi-output case,

$$J(\boldsymbol{\theta}) - J(\hat{\boldsymbol{\theta}}) = n_p s^2 F_{\alpha_p}(n_p, Nn_o - n_p) \times \left[1 + \frac{Nn_o(n_p + 2)}{(Nn_o - n_p)n_p} N_\phi\right] \tag{5.29}$$

where s^2 is now given as

$$s^2 = \frac{J(\hat{\boldsymbol{\theta}})}{Nn_o - n_p} \tag{5.30}$$

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

The confidence contours, defined by equation (5.29), cannot be determined analytically as done in the linear case since the contours are not necessarily ellipsoidal. The contours may be very irregular and may possibly have 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 $\boldsymbol{\theta}_{\min}$ and $\boldsymbol{\theta}_{\max}$; for the linear case, the confidence interval is given by the dashed vertical lines equidistant from $\hat{\boldsymbol{\theta}}$. The search algorithm used in this study for finding the contour boundaries was presented in reference 26. 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 search algorithm is computationally very demanding, even for problems with relatively few parameters.

B. Nonlinearity Measure for Aircraft Applications

The following is a generalization of Beale's development (ref. 23) of an intrinsic nonlinearity measure N_ϕ and its adaptation 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 $P(\boldsymbol{\theta})$ represents the estimation space (or solution locus) in sample space, then $P(\hat{\boldsymbol{\theta}})$ is the point on

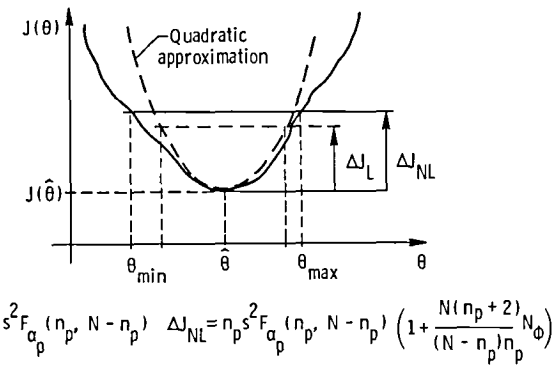


Figure 4. Construction of confidence intervals for one-dimensional problem.

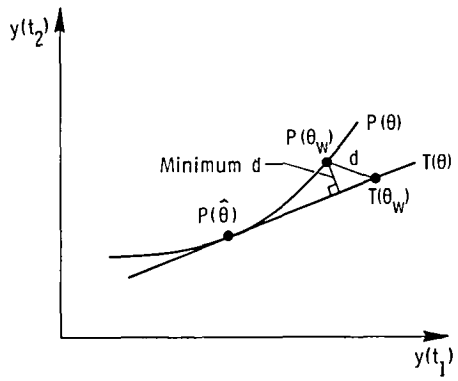


Figure 5. Two-dimensional sample space with solution locus $P(\boldsymbol{\theta})$ and tangent plane $T(\boldsymbol{\theta})$ at $P(\hat{\boldsymbol{\theta}})$.

the solution locus closest to the measurement Z and $\hat{\boldsymbol{\theta}}$ is the point in parameter space which minimizes the cost function. If $T(\boldsymbol{\theta})$ is defined as a point on a tangent plane at $P(\hat{\boldsymbol{\theta}})$ and W different values are chosen for the vector $\boldsymbol{\theta}$ near $\hat{\boldsymbol{\theta}}$ (i.e., $\boldsymbol{\theta}_w, w = 1, 2, \ldots, W$), then a crude measure of nonlinearity can be written as

$$Q_\psi = \sum_{w=1}^{W} |P(\boldsymbol{\theta}_w) - T(\boldsymbol{\theta}_w)|^2 \tag{5.31}$$

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

$$D = \sum_{w=1}^{W} |P(\boldsymbol{\theta}_w) - P(\hat{\boldsymbol{\theta}})|^4 \tag{5.32}$$

and, thus, Q_ψ is normalized as

$$N_\psi = \frac{n_p s^2 Q_\psi}{D} \tag{5.33}$$

The values of this measure may still depend on the configuration or orientation of the points $P(\boldsymbol{\theta}_w)$ relative to $P(\hat{\boldsymbol{\theta}})$, but it should not depend significantly on the number of points $P(\boldsymbol{\theta}_w)$ chosen or on their distances from $P(\hat{\boldsymbol{\theta}})$ (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 $\boldsymbol{\theta}$ are chosen such that $T(\boldsymbol{\theta})$ is always at the foot of a perpendicular from $P(\boldsymbol{\theta})$ onto a tangent plane at $P(\hat{\boldsymbol{\theta}})$. 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 as follows. 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

$$\hat{N}_\phi = \frac{n_p s^2 \hat{Q}_\phi}{D} \tag{5.34}$$

where n_p is the number of unknown parameters and s^2 is the sum of squares of residuals. For the multiple output case, s^2 is given by equation (5.30).

The denominator in equation (5.34), D, can be formulated as

$$D = \sum_{w=1}^{W} \left\{ \sum_{i=1}^{N} [\hat{\mathbf{Y}}_i(\boldsymbol{\theta}_w) - \hat{\mathbf{Y}}_i(\hat{\boldsymbol{\theta}})]^T \hat{\mathbf{R}}^{-1} \times [\hat{\mathbf{Y}}_i(\boldsymbol{\theta}_w) - \hat{\mathbf{Y}}_i(\hat{\boldsymbol{\theta}})] \right\}^2 \tag{5.35}$$

where i is the index of the N data points (time variable).

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

$$Q_\psi = \sum_{w=1}^{W} \sum_{i=1}^{N} [\hat{\mathbf{Y}}_i(\boldsymbol{\theta}_w) - \hat{\mathbf{Y}}_i(\hat{\boldsymbol{\theta}}) - \mathbf{G}_i \boldsymbol{\psi}_w]^T \times \hat{\mathbf{R}}^{-1} [\hat{\mathbf{Y}}_i(\boldsymbol{\theta}_w) - \hat{\mathbf{Y}}_i(\hat{\boldsymbol{\theta}}) - \mathbf{G}_i \boldsymbol{\psi}_w] \tag{5.36}$$

where

$$\mathbf{G}_i = \left[\frac{\partial y_k}{\partial \theta_\ell} \right]_i \tag{4.6}$$

$$\psi_w = \theta_w - \hat{\theta} \tag{5.37}$$

Let

$$\delta\hat{\mathbf{Y}}_i = \hat{\mathbf{Y}}_i(\theta_w) - \hat{\mathbf{Y}}_i(\hat{\theta}) \tag{5.38}$$

and rewrite equation (5.36) to obtain

$$Q_\psi \approx \sum_{w=1}^{W}\sum_{i=1}^{N}(\delta\hat{\mathbf{Y}}_i - \mathbf{G}_i\psi_w)^T\hat{\mathbf{R}}^{-1}(\delta\hat{\mathbf{Y}}_i - \mathbf{G}_i\psi_w) \tag{5.39}$$

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

$$d = |P(\theta_w) - T(\theta)|^2 \tag{5.40}$$

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

$$\Phi_w = \left(\sum_{i=1}^{N}\mathbf{G}_i^T\hat{\mathbf{R}}^{-1}\mathbf{G}_i\right)^{-1} \times \sum_{i=1}^{N}\mathbf{G}_i^T\hat{\mathbf{R}}^{-1}\delta\hat{\mathbf{Y}}_i \qquad (w = 1,2,\ldots,W) \tag{5.41}$$

Thus

$$\hat{Q}_\phi = \sum_{w=1}^{W}\sum_{i=1}^{N}(\delta\hat{\mathbf{Y}}_i - \mathbf{G}_i\Phi_w)^T\hat{\mathbf{R}}^{-1}(\delta\hat{\mathbf{Y}}_i - \mathbf{G}_i\Phi_w) \tag{5.42}$$

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

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

$$|\Delta\theta_i| = \hat{\sigma}_{\theta_i} \tag{5.43}$$

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 to computing N_ϕ. With 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. The error bounds determined by random search and N_ϕ can also indicate the effect of experimental design, especially input form and model error, on identifiability.

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 with commonly used techniques, which provide a benchmark for comparison. The commonly used techniques are the ML/MNR algorithm for parameter estimation and the Cramer-Rao (CR) bounds for confidence interval determination. ML/MNR is used with both analytically and numerically determined derivatives. The CR 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 is 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 1–3) and flight data (examples 4–6). Except for the first two examples, the parameters estimated are the nondimensional aircraft stability and control derivatives conforming to standard NASA notation. For examples 1 and 2, a simple linear system is integrated with an Euler integration method. Examples 3–6 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 (Control Data 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 flowchart for program MAX is given in figure 7 with table I defining the elements in figure 7.

TABLE I. PRIMARY SUBROUTINES FOR PROGRAM MAX AND DEFINITIONS OF FLOWCHART BLOCKS IN FIGURE 7

Name	Description
	Subroutines
AERO	Computes aerodynamic forces and moments with selected aerodynamic model
COST	Computes residuals, fit error, $\hat{\mathbf{R}}^{-1}$, and cost
DIFFEQ	Computes state derivatives from selected equations of motion
EST	Computes updated parameter estimates
HICOST	Determines whether new estimates reduce cost and updates storage of outgoing and incoming parameters and response time histories
INT	Main subroutine for management of model inputs and outputs; computes initial conditions and input arrays for RK4 and OUTPUT
MASTER	Primary subroutine represented by flowchart; handles initializations, input/output 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
FAIL	Test whether 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
$\hat{\mathbf{R}}$ UPDATES	Test for maximum number of weighting matrix updates

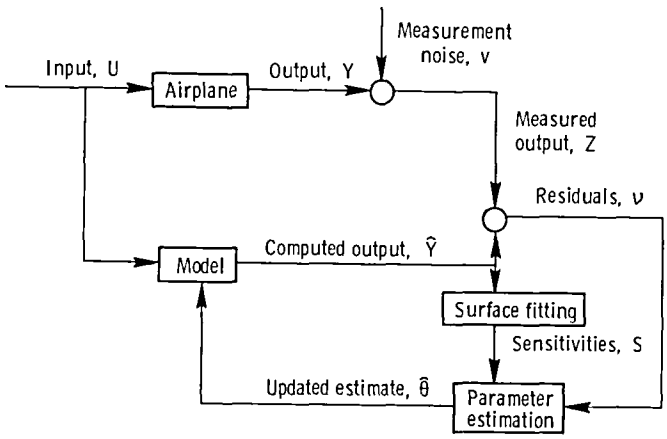


Figure 6. Computing scheme for ML/MNRES.

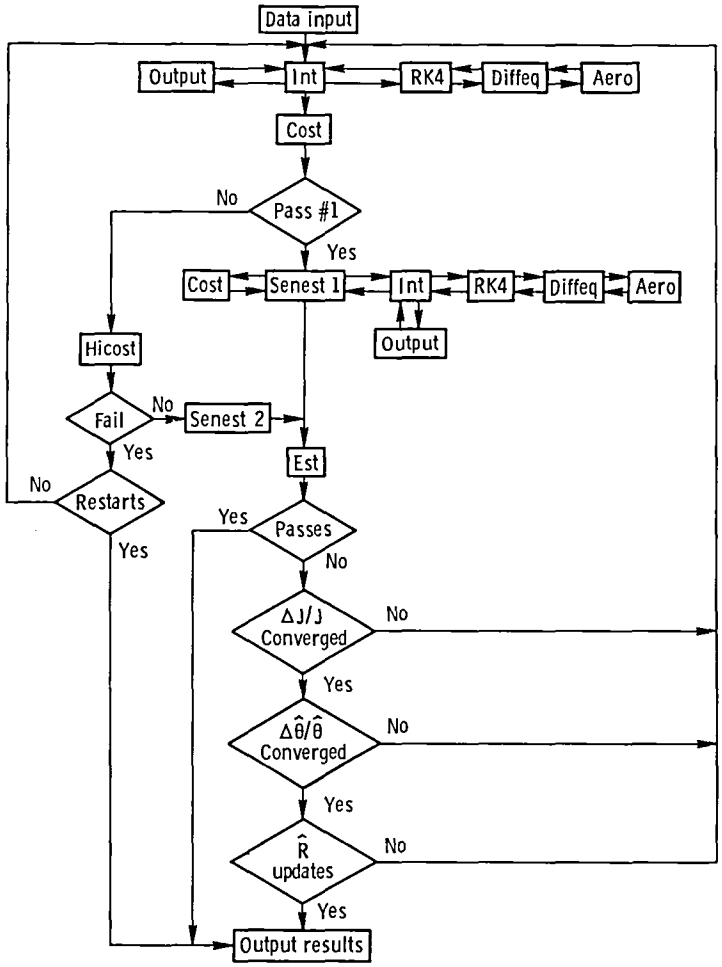


Figure 7. Flowchart for program MAX.

A. Simulated Data Studies

Simulated data offer 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 is that the problem can be well posed and defined without modeling error; and the third is that 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:

$$\dot{\mathbf{X}}_s = \mathbf{A}\mathbf{X}_s + \mathbf{B}\mathbf{U} \qquad \mathbf{X}_s(0) = 0 \qquad (6.1)$$

$$\mathbf{Y} = \mathbf{X}_s \qquad (6.2)$$

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.

1. Example 1

Example 1 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. 16). 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. The recursive form is normally used for conserving memory; however because of the small memory requirements to store time histories in this example, all time histories are stored so that only one integration per pass is needed. The purpose of this example is to compare the relative performance of each method on a problem involving a simple dynamic system.

Example 1 has six unknown parameters. The six unknown parameters in equation (6.1) are the four elements of the 2×2 system matrix $\mathbf{A}$ and two elements of the control input matrix $\mathbf{B}$. The input form was chosen as

$$U = \begin{cases} \sin t & (0 < t < 2\pi) \\ 0 & (t > 2\pi) \end{cases}$$

and data points were generated every 0.25 second. Process and measurement noise was 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, that is 715:28:12 for FPS:MNR:MNRES, respectively.

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 32. 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 1, further study concentrated only on the gradient methods.

2. Example 2

Example 2 is provided to compare the robustness of MNRES with that of 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 1 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 the information limited 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 that 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 to 130 percent. MNR had both the most accurate and the least accurate estimate. The importance (sensitivity) of a parameter to

the model significantly affects 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.

3. Example 3

In example 3, 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 (4.15) instead of the recursive least squares form with equation (4.23). 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 were created by a fourth-order Runge-Kutta integration with a step size of 0.05 second. 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 in figure 10 for the three cases. The control inputs were the same for all three cases and are shown in figure 11.

Program MAX was applied using two convergence criteria: $\Delta J/J \leq 10^{-3}$ and $\Delta\boldsymbol{\theta}/\boldsymbol{\theta} \leq 10^{-3}$. As expected, the estimates of the less easily identified nonlinear terms, such as $C_{n_{\alpha r}}$ and $C_{l_{\alpha p}}$, 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.

TABLE II. PARAMETER ESTIMATES AND COMPUTATION TIME FOR FPS, MNR, AND MNRES APPLIED TO A LINEAR SYSTEM WITHOUT MEASUREMENT NOISE (EXAMPLE 1)

Unknown parameters, $\boldsymbol{\theta}$	True value	Initial value	Final estimated values using method—		
			FPS	MNR	MNRES
θ_1	0	0.01	-0.12×10^{-3}	0.89×10^{-7}	0.73×10^{-6}
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			0.14×10^{-8}	0.61×10^{-10}	0.11×10^{-7}
Equivalent evaluation			715	28	12
CPU time,[a] sec			2948	106	47

[a]Central processing unit time on CYBER 175 computer.

TABLE III. PARAMETER ESTIMATES AND COMPUTATION TIME FOR MNR AND MNRES APPLIED TO A LINEAR SYSTEM WITH TWO NOISE LEVELS (EXAMPLE 2)

(a) Case 1

[Noise level with standard error of 0.0001]

Unknown parameters, θ	True value	Initial value	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			0.105×10^{-6}	0.105×10^{-6}
Equivalent evaluation			42	12
CPU time,[a] sec			77.54	24.08

(b) Case 2

[Noise level with standard error of 0.001]

Unknown parameters, θ	True value	Initial value	Final estimated values using method—	
			MNR	MNRES[b]
θ_1	0	0.01	−0.705	−0.410
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			0.104×10^{-4}	0.122×10^{-4}
Equivalent evaluation			70	27
CPU time,[a] sec			134.44	56.05

[a]Central processing unit time on CYBER 175 computer.
[b]MNRES did not converge.

TABLE IV. STANDARD ERRORS OF SIMULATED MEASUREMENT NOISE (EXAMPLE 3)

Output variable	Standard errors for—		
	Case 1	Case 2	Case 3
β, rad	0	0.010	0.02
p, rad/sec	0	.010	.02
r, rad/sec	0	.010	.02
ϕ, rad	0	.005	.01
a_y, g units	0	.005	.01

TABLE V. PARAMETER ESTIMATES FROM ML/MNRES APPLIED TO SIMULATED NONLINEAR AIRCRAFT SYSTEM (EXAMPLE 3)

Unknown parameters, θ	Simulation values	Parameter estimates for—		
		Case 1	Case 2	Case 3
$C_{Y,o}$	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
$C_{l,o}$	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
$C_{l_{\delta a}}$	−.182	−.1815	−.1778	−.1755
$C_{l_{\delta r}}$	.007	.0067	.0059	.00497
$C_{l_{\alpha p}}$	2.63	2.6254	2.519	2.4359
$C_{n,o}$	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

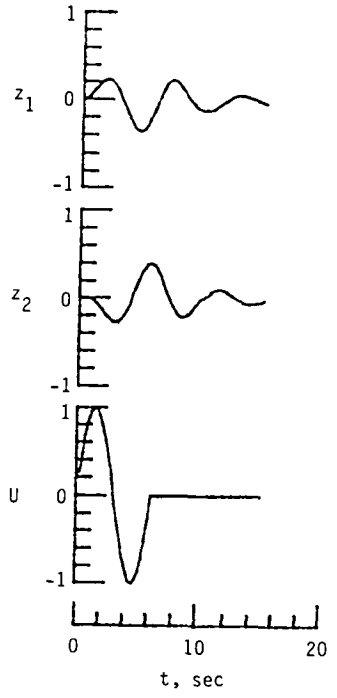
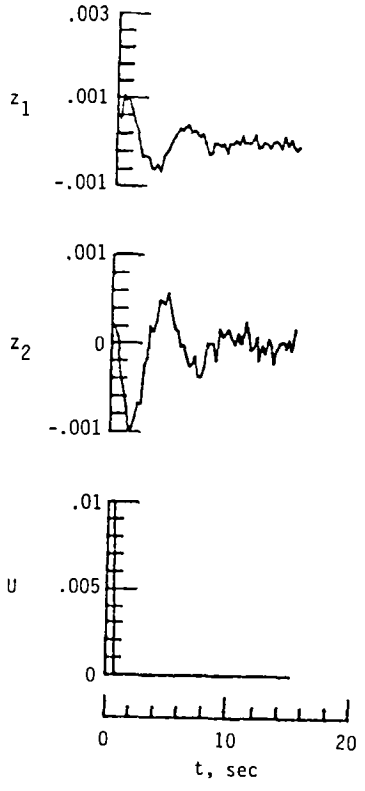
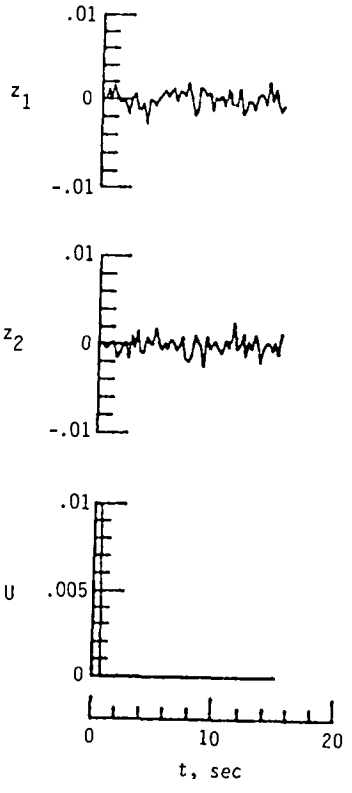


Figure 8. Time histories of input and response variables for example 1.

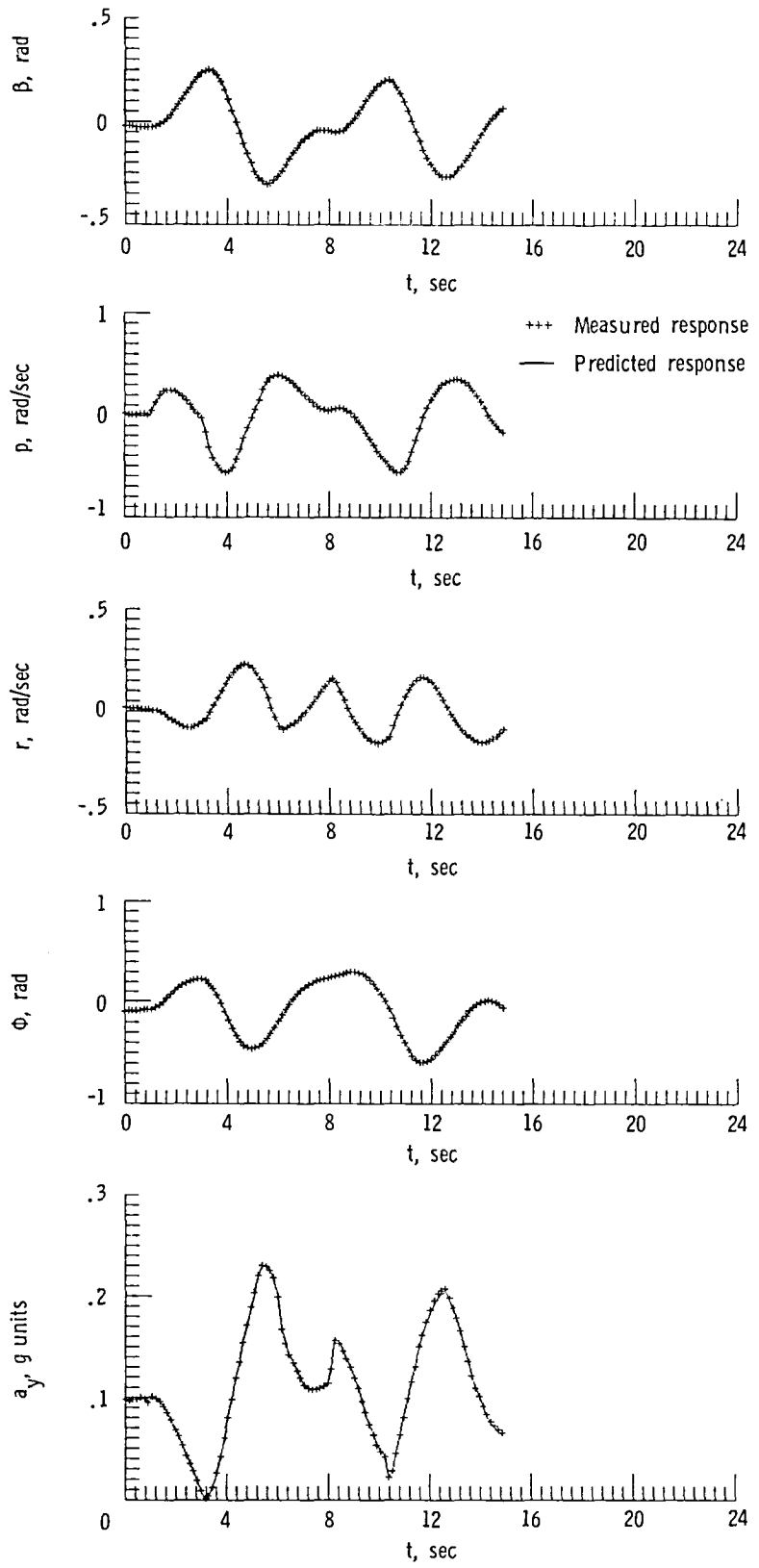


(a) Case 1.



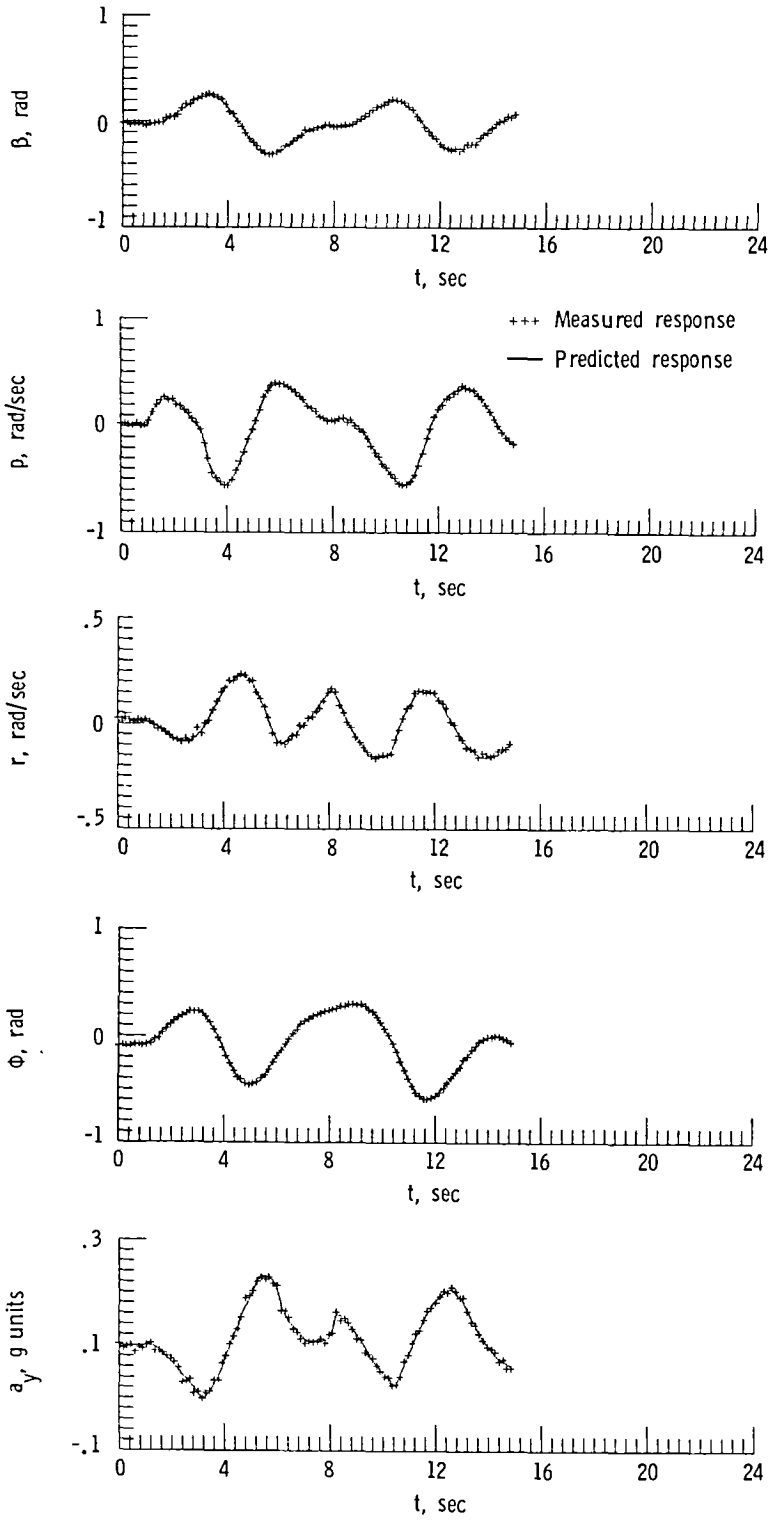
(b) Case 2.

Figure 9. Time histories of input and response variables for example 2.



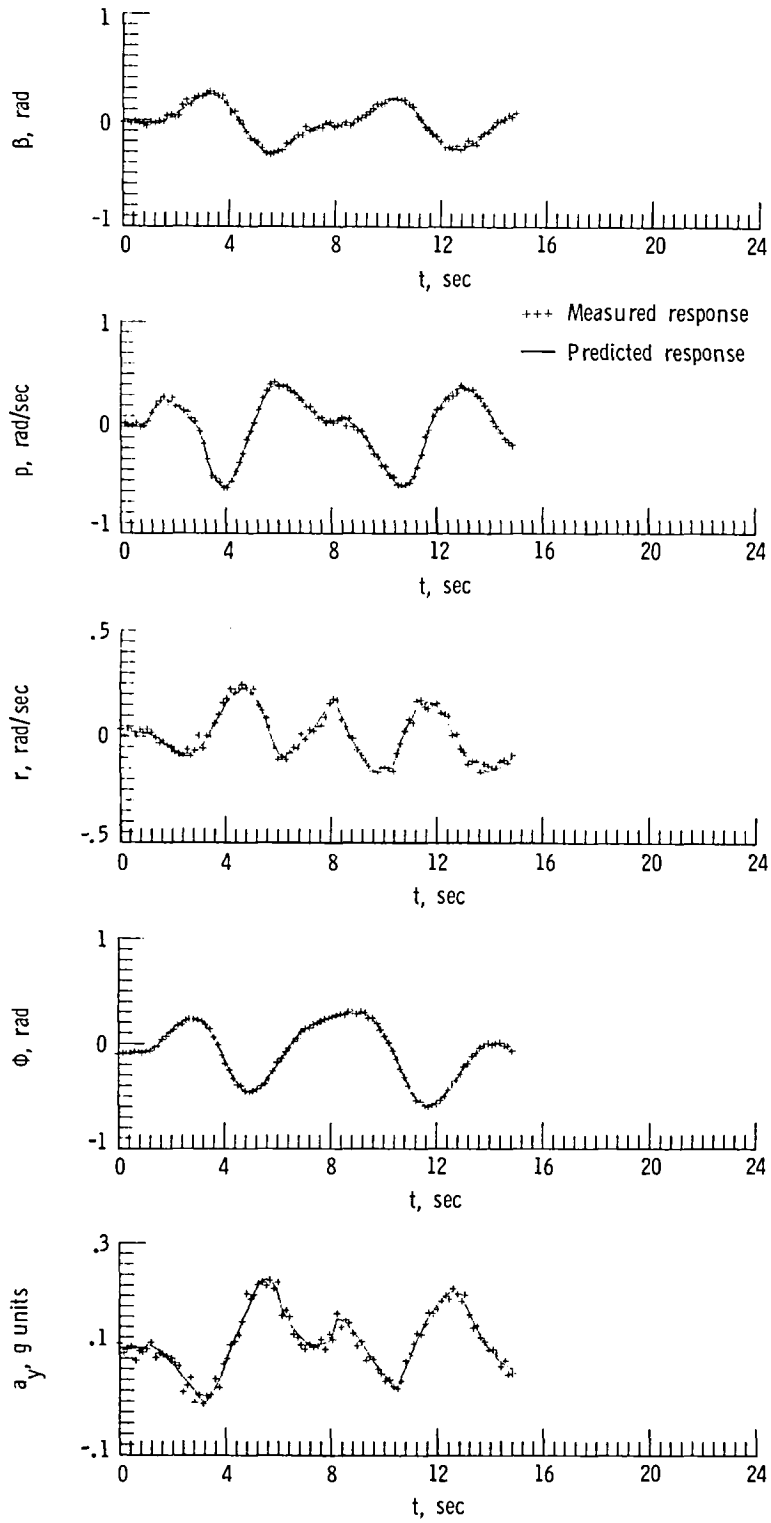
(a) Case 1.

Figure 10. Measured and predicted responses for lateral simulation (example 3).



(b) Case 2.

Figure 10. Continued.



(c) Case 3.

Figure 10. Concluded.

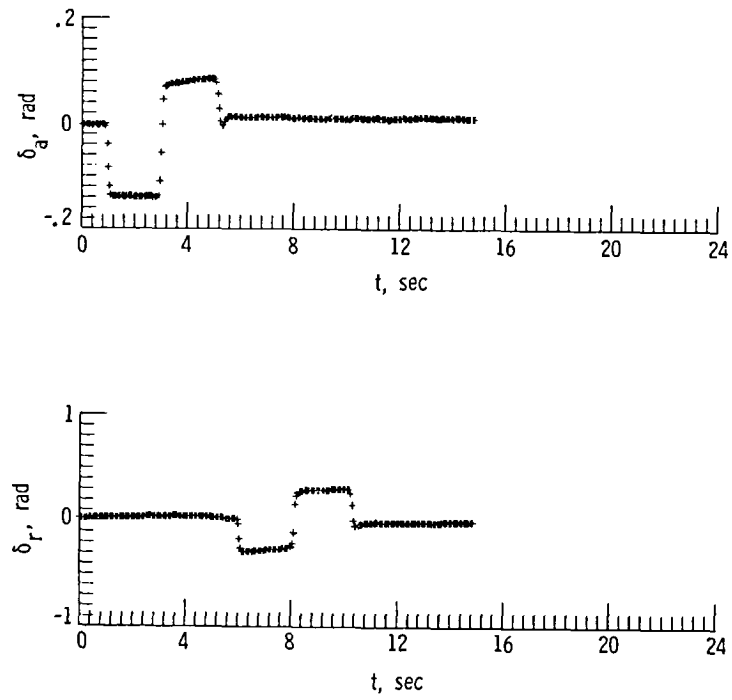


Figure 11. Control inputs for lateral simulation (example 3).

B. Real Flight Data Studies

In this section three examples are considered. In each example the model structure and initial parameter estimates were determined with the modified stepwise regression (MSR) program of reference 11. For examples 4 and 5, the parameter estimation problem is solved by using two ML programs. The first is program MAX, which uses the ML/MNRES algorithm as described in example 3. 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 34. In the last example, program MAX is used to compute parameter estimates and CR bounds. These bounds are adjusted to the 95-percent confidence level and compared with those obtained with the search method.

For comparison purposes, both program MAX and program MAXLIK use a fourth-order Runge-Kutta integration method with the same integration step size (0.05 sec in example 4 and 0.04 sec in examples 5 and 6). A convergence criterion is set at $\Delta J/J = 0.001$ for both codes. Program MAX normally uses an additional criterion restricting the parameter change $\Delta\boldsymbol{\theta}/\boldsymbol{\theta}$; 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 37. The same initial parameter values are used by both MAX and MAXLIK.

1. Example 4

Example 4 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 of parameter estimates and CR bounds from MAX and from MAXLIK. Initial values and sensitivities computed as $\theta_\ell^2 M_{\ell\ell}$ are also given. Again, there is reasonable agreement between the two approaches. CR 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, slightly improved 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 4 are 793 seconds for program MAX and 1036 seconds for program MAXLIK.

2. Example 5

Example 5 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 from MAXLIK and from 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 about one-third of the time. This example reflects the effectiveness of MNRES under fortunate conditions (that is, where the cost is well approximated by a quadratic function 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 0.003 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 those from MAXLIK.

Although the Fisher information matrix is updated during each iteration by both programs, program MAX delays updating the weighting matrix $\mathbf{R}^{-1}$ until convergence is achieved. The example is solved once more by program MAX and the weighting matrix is updated twice. These results are shown in table VII as case 2 for program MAX. 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 46 percent of the MAXLIK processing time.

3. Example 6

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 0.02; 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.

TABLE VI. PARAMETER ESTIMATES, CR BOUNDS, AND SENSITIVITIES FOR PROGRAMS MAX AND MAXLIK APPLIED TO REAL DATA FROM NONLINEAR AIRCRAFT SYSTEM (EXAMPLE 4)

Unknown parameters, θ	Initial value	Program MAX		Program MAXLIK		$\theta_\ell^2 M_{\ell\ell}$
		$\hat{\theta}$	$\hat{\sigma}$	$\hat{\theta}$	$\hat{\sigma}$	
$C_{Y,o}$	0.036	0.0061	0.0006	0.0213	0.0005	0.7533×10^3
C_{Y_β}	−.479	−.4603	.0075	−.4608	.0067	$.3756 \times 10^5$
C_{Y_p}	−.186	−.1378	.0485	−.0604	.0439	$.9373 \times 10^3$
C_{Y_r}	.522	.6677	.0289	.6209	.0256	$.1915 \times 10^4$
$C_{Y_{\delta a}}$	−.08	−.0504	.0166	−.0375	.0150	$.8129 \times 10^3$
$C_{Y_{\delta r}}$	.083	.0814	.0043	.0763	.0037	$.9922 \times 10^3$
$C_{Y_{\alpha\beta}}$	.45	.4300	.0592	.4512	.0504	$.9399 \times 10^3$
$C_{l,o}$	0	.0002	.00005	−.0001	.00005	$.9618 \times 10^3$
C_{l_β}	−.079	−.0872	.0015	−.08	.0013	$.1493 \times 10^6$
C_{l_p}	−.47	−.5320	.0102	−.4823	.0085	$.4529 \times 10^7$
C_{l_r}	.187	.1700	.0043	.1543	.0045	$.6838 \times 10^4$
$C_{l_{\delta a}}$	−.19	−.2035	.0036	−.1852	.0031	$.7969 \times 10^5$
$C_{l_{\delta r}}$	.01	.00055	.00024	−.0012	.00072	$.1340 \times 10^4$
$C_{l_{\alpha\beta}}$	−.26	−.2707	.0116	−.2105	.0091	$.6327 \times 10^4$
$C_{n,o}$	0	−.00063	.00003	−.002	.00002	$.1909 \times 10^4$
C_{n_β}	.04	.0323	.00045	.0329	.0004	$.1515 \times 10^6$
C_{n_p}	−.056	−.1043	.0026	−.0916	.0022	$.1400 \times 10^6$
C_{n_r}	−.15	−.1462	.002	−.1534	.0017	$.5039 \times 10^5$
$C_{n_{\delta a}}$	0	−.0044	.001	−.0037	.0009	$.1780 \times 10^4$
$C_{n_{\delta r}}$	−.053	−.0550	.0003	−.0532	.0003	$.9048 \times 10^7$
[a]$C_{Y_{\beta 3}}$	−.39	−.39		−.39		
[a]$C_{n_{\beta 3}}$	.08	.08		.08		

[a]Parameter held fixed.

TABLE VII. PARAMETER ESTIMATES, THEIR STANDARD ERRORS, AND TIME TO REACH CONVERGENCE FOR PROGRAMS MAX AND MAXLIK APPLIED TO REAL DATA FROM NONLINEAR AIRCRAFT SYSTEM (EXAMPLE 5)

Unknown parameters, θ	MAXLIK		MAX			
			Case 1[a]		Case 2[b]	
	$\hat{\theta}$	[c]$\hat{\sigma}$	$\hat{\theta}$	[c]$\hat{\sigma}$	$\hat{\theta}$	[c]$\hat{\sigma}$
$C_{X,o}$	0.17017	0.208×10^{-3}	0.17078	0.249×10^{-3}	0.17070	0.212×10^{-3}
C_{X_α}	.9464	$.616 \times 10^{-2}$	.9245	$.703 \times 10^{-2}$	.92511	$.555 \times 10^{-2}$
$C_{X_{\delta e}}$	.2789	$.404 \times 10^{-2}$	.2754	$.116 \times 10^{-1}$	.2755	$.876 \times 10^{-2}$
$C_{Z,o}$	−.83946	$.896 \times 10^{-3}$	−.84335	$.106 \times 10^{-2}$	−.84252	$.103 \times 10^{-2}$
C_{Z_α}	−5.311	$.230 \times 10^{-1}$	−5.197	$.223 \times 10^{-1}$	−5.194	$.215 \times 10^{-1}$
C_{Z_q}	−18.7	$.162 \times 10$	−20.3	$.189 \times 10$	−20.5	$.177 \times 10$
$C_{Z_{\delta e}}$	−.618	$.264 \times 10^{-1}$	−.566	$.350 \times 10^{-1}$	−.570	$.324 \times 10^{-1}$
$C_{m,o}$	−.001251	$.828 \times 10^{-4}$	−.001610	$.942 \times 10^{-4}$	−.001555	$.937 \times 10^{-4}$
C_{m_α}	−.5129	$.102 \times 10^{-2}$	−.5186	$.115 \times 10^{-2}$	−.5183	$.110 \times 10^{-2}$
C_{m_q}	−16.95	.157	−19.06	.144	−18.83	.144
$C_{m_{\delta e}}$	−1.3409	$.576 \times 10^{-2}$	−1.4150	$.447 \times 10^{-2}$	−1.4075	$.461 \times 10^{-2}$
CPU time,[d] sec . . .	342		105		157	

[a]Case 1 is for one update of weighting matrix $\mathbf{R}^{-1}$.
[b]Case 2 is for two updates of weighting matrix $\mathbf{R}^{-1}$.
[c]Cramer-Rao bound.
[d]Central processing unit time on CYBER 175 computer.

TABLE VIII. PARAMETER ESTIMATES AND CR BOUNDS FROM PROGRAM MAX AND CONFIDENCE LIMITS FROM A RANDOM SEARCH TECHNIQUE FOR REAL DATA FROM NONLINEAR AIRCRAFT SYSTEM (EXAMPLE 6)

Unknown parameters, θ	$\hat{\theta}$	95% confidence interval		
		Cramer-Rao bound	Random search	
			Upper bound	Lower bound
C_{Y_β}	−0.77	±0.27	0.92	−0.89
$C_{Y_{\delta r}}$	.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
$C_{l_{\delta a}}$	−.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

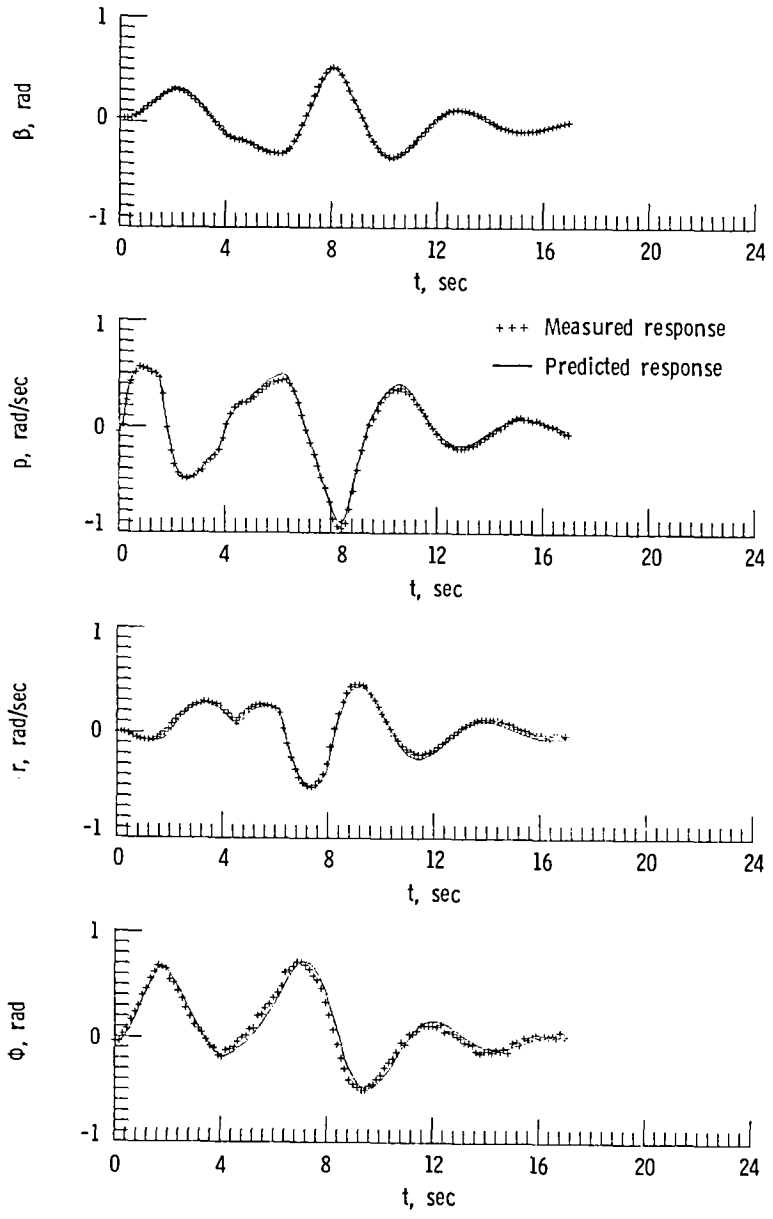


Figure 12. Measured and predicted responses for lateral maneuvering using real flight data (example 4).

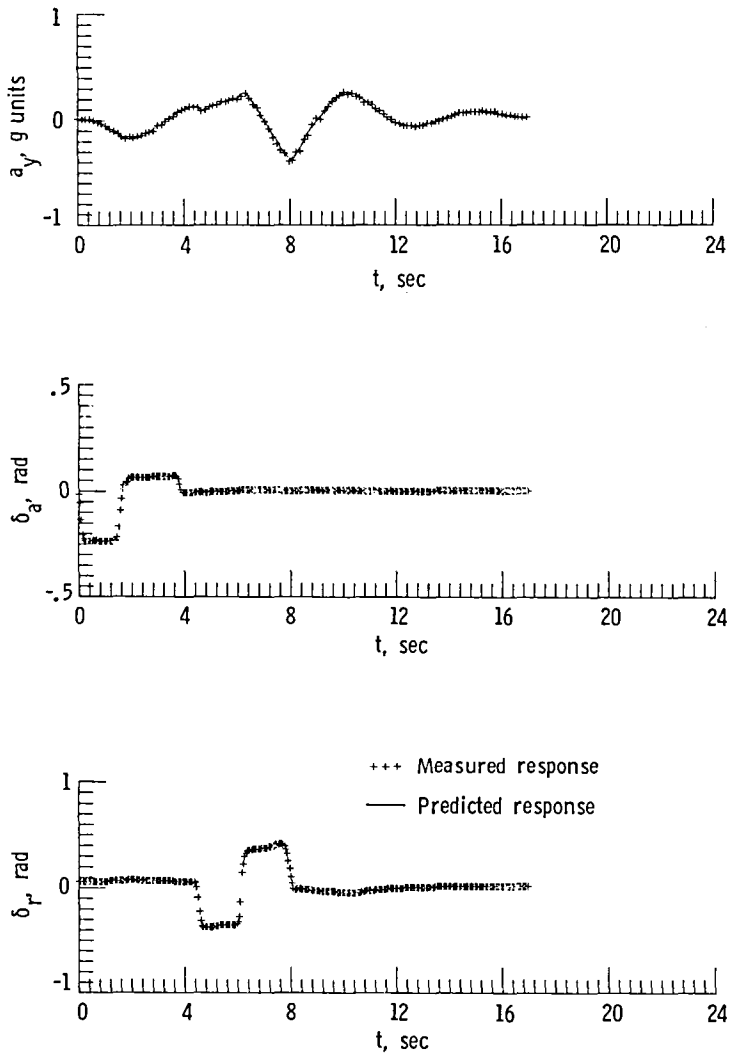


Figure 12. Concluded.

C. Discussion of Results

The general experience with ML/MNRES and the examples chosen for this study indicate that ML/MNRES performs 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 function). 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 CR bounds should agree with values determined by the search technique.

1. Parameter Estimation

Two 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 of dealing with 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 (2) the modular form of MAX allows easy application to any system. The second goal is achieved in three ways: (1) the algorithm provides the user many computationally efficient options for approximating the sensitivities (allowing for variation in accuracy and order of derivatives); (2) the algorithm allows for substantial reduction in memory requirements with only a small cost in additional computation; and (3) all the above options are readily incorporated because of the modular format of MAX. The second goal has been demonstrated in the examples and the following discussion clarifies 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. Because the initial values were relatively close to the final solution, a good quadratic approximation of the cost function was possible and thus provided 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 these equations provides information that 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; for example 1, the ratio of equivalent evaluations was 28:12. 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 conditons for estimation algorithms because problems such as modeling 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 used real flight data and were specially selected to reflect differing levels of difficulty for the estimation algorithms. Examples 4 and 6 relative to example 5 demonstrate a representative range in the degree of difficulty (measured by computational effort) for the algorithms, and it is no surprise that the degree of nonlinearity also varies widely (N_ϕ differed by an order of magnitude between examples 5 and 6, 6 being more nonlinear). ML/MNRES was again faster than the benchmark algorithm 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 46 percent of the benchmark time. These examples give some credence to the superiority of ML/MNRES.

However, these examples also indicate a large variability in the superiority of ML/MNRES. As the degree of nonlinearity increases, the two methods appear to 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 both example 5 and example 6 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 section 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

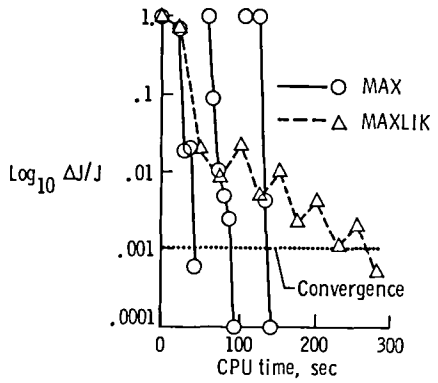


Figure 13. Convergence criterion versus CPU time for programs MAX and MAXLIK.

the nonlinearity increases, or the quadratic approximation becomes poorer, but the MNRES method depends much more on the quality of the quadratic approximation, since in effect it is an approximation of the MNR method. As the nonlinearity increases, MNRES loses its advantage over MNR.

Figure 13 offers a graphical view of the performance of MNR and MNRES in example 5, where the quadratic approximation is fairly good. The graph shows 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 $\mathbf{R}$ during each pass after the criterion falls below 0.01 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 during 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 $\mathbf{M}$ does not converge until the weighting matrix $\mathbf{R}^{-1}$ is updated sufficiently. This is understandable since the parameter estimates are asymptotically independent of the weighting matrix, whereas the CR bounds depend on the information matrix, which in turn depends on the value of $\mathbf{R}$. Therefore, two more cycles are made to convergence to ensure 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 the method of updating the information matrix. Although both ML/MNRES and ML/MNR update the information matrix during each pass, they do it quite differently. One well-known way to improve the speed of techniques such as ML/MNR involving the Hessian matrix (or approximations to it) is to hold the information matrix $\mathbf{M}$ constant for one or more iterations (ref. 14). This reduces the amount of integration required per pass to the same as required in MNRES because no sensitivity equations are integrated. The number of iterations that $\mathbf{M}$ 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 during each pass, requiring maximum computational effort but giving maximum accuracy to the optimization process. The other alternative is to integrate the sensitivity equations during 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, not updating the information matrix is dangerous. ML/MNRES offers an effective compromise between these two unsatisfactory alternatives.

The compromise is achieved by updating the information matrix during each pass, but incorporating only the information obtained from integrating the equations of motion once. Thus, 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 each pass in 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.

2. Confidence Interval Estimation

Confidence intervals obtained in example 5 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 5 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 example 6 were found to be 3 to 8 times the size of the CR bound adjusted to the 95-percent confidence level (table VIII). This is in agreement with references 20 and 25 on the values generally found in analyzing actual flight

data. Reference 25 used the search technique and reference 20 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 is not well approximated by a quadratic function at the error level considered. This is confirmed by the relative speed of convergence of MNRES and MNR. MNRES, with two updates of $\mathbf{R}$, 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, modeling error, 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 differ 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 whether the lengthy computations of the random search are needed. If very little nonlinearity exists, the user can be reasonably confident that the CR bounds provide accurate error bound information.

VII. CONCLUDING REMARKS

The primary contribution in this study is a 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 readily accommodates the widely varying model structure in nonlinear flight regimes. A maximum likelihood scheme with the optimization algorithm developed in this study (ML/MNRES) 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 nonlinear problems. 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 whether the extensive calculations of the random search are needed to estimate confidence limits.

The new optimization algorithm, MNRES, has three advantages over other comonly 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 and also provides much flexibility, allowing the model equations to be in any convenient format, 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 methods for fitting the output vector to a surface in parameter space (needed for sensitivity estimation), the user can choose a surface-fitting method best suited to the problem. Also storage requirements can be reduced with little additional computation. The third advantage of the algorithm is that it requires less computational effort than 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 accurate. 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 includes 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 problems 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 problem. 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 is closely related to the degree to which the Cramer-Rao bounds and the random search confidence limits agree. Therefore, it is recommended that this or some similar measure be used 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 how the confidence intervals and the nonlinearity of the cost function relate to other factors such as bias errors, modeling error, input form, and noise spectra. In addition, measures of nonlinearity and the best schemes for computing them need more investigation. 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 with gradient techniques rather than the computationally demanding search scheme used in this study.

NASA Langley Research Center
Hampton, VA 23665-5225
October 17, 1985

REFERENCES

1. Åström, K. J.; and Eykhoff, P.: System Identification—A Survey. *Automatica,* vol. 7, Mar. 1971, pp. 123–162.
2. Greenberg, Harry: *A Survey of Methods for Determining Stability Parameters of an Airplane From Dynamic Flight Measurements.* NACA TN 2340, 1951.
3. Shinbrot, Marvin: *On the Analysis of Linear and Nonlinear Dynamical Systems From Transient-Response Data.* NACA TN 3288, 1954.
4. Shinbrot, Marvin: *A Least Squares Curve Fitting Method With Applications to the Calculation of Stability Coefficients From Transient-Response Data.* NACA TN 2341, 1951.
5. Rediess, Herman A.: An Overview of Parameter Estimation Techniques and Applications in Aircraft Flight Testing. *Parameter Estimation Techniques and Applications in Aircraft Flight Testing,* NASA TN D-7647, 1974, pp. 1–18.
6. Larson, Duane B.: *Identification of Parameters by the Method of Quasilinearization.* CAL Rep. No. 164, Cornell Aeronaut. Lab., Inc., May 14, 1968.
7. Taylor, Lawrence W., Jr.; and Iliff, Kenneth W.: A Modified Newton-Raphson Method for Determining Stability Derivatives From Flight Data. *Computing Methods in Optimization Problems—2,* Lofti A. Zadeh, Lucien W. Neustadt, and A. V. Balakrishnan, eds., Academic Press, 1969, pp. 353–364.
8. Mehra, Raman K.: Maximum Likelihood Identification of Aircraft Parameters. *Eleventh Joint Automatic Control Conference of the American Automatic Control Council,* American Soc. Mech. Eng., c.1970, pp. 442–444.
9. Stepner, David E.; and Mehra, Raman K.: *Maximum Likelihood Identification and Optimal Input Design for Identifying Aircraft Stability and Control Derivatives.* NASA CR-2200, 1973.
10. Rault, André: Identification Applications to Aeronautics. *Identification and System Parameter Estimation, Part 1,* P. Eykhoff, ed., American Elsevier Pub. Co., Inc., 1973, pp. 49–65.
11. Klein, Vladislav; Batterson, James G.; and Murphy, Patrick C.: *Determination of Airplane Model Structure From Flight Data by Using Modified Stepwise Regression.* NASA TP-1916, 1981.
12. Klein, Vladislav: Identification Evaluation Methods. *Parameter Identification,* AGARD-LS-104, Nov. 1979, pp. 2-1–2-21.
13. Himmelblau, David M.: *Applied Nonlinear Programming.* McGraw-Hill, Inc., c.1972.
14. Gupta, Narendra K.; and Mehra, Raman K.: Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Calculations. *IEEE Trans. Autom. Control,* vol. AC-19, no. 6, Dec. 1974, pp. 774–783.
15. Bowles, Roland L.; and Straeter, Terry A.: System Identification Computational Considerations. *System Identification of Vibrating Structures—Mathematical Models From Test Data,* W. D. Pilkey and R. Cohen, eds., American Soc. Mech. Eng., 1972, pp. 23–43.
16. Trankle, T. L.; Vincent, J. H.; and Franklin, S. N.: Recent Methods for Nonlinear System Identification. AIAA-81-1865, Aug. 1981.
17. 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,* C. T. Leondes, ed., AGARD-AG-256, Mar. 1982, pp. 7-1–7-26.
18. Ralston, Mary L.; and Jennrich, Robert I.: Dud, A Derivative-Free Algorithm for Nonlinear Least Squares. *Technometrics,* vol. 20, no. 1, Feb. 1978, pp. 7–13.
19. Murphy, Patrick C.: *An Algorithm for Maximum Likelihood Estimation Using an Efficient Method for Approximating Sensitivities.* NASA TP-2311, 1984.
20. Klein, Vladislav: *Determination of Stability and Control Parameters of a Light Airplane From Flight Data Using Two Estimation Methods.* NASA TP-1306, 1979.
21. Balakrishnan, A. V.; and Maine, Richard E.: *Improvements in Aircraft Extraction Programs.* NASA CR-145090, [1976].
22. Maine, Richard E.; and Iliff, Kenneth W.: *The Theory and Practice of Estimating the Accuracy of Dynamic Flight-Determined Coefficients.* NASA RP-1077, 1981.
23. Beale, E. M. L.: Confidence Regions in Non-linear Estimation. *J. R. Stat. Soc.,* ser. B, vol. 22, no. 1, 1960, pp. 41–88.
24. Bates, Douglas M.; and Watts, Donald G.: Relative Curvature Measures of Nonlinearity. *J. R. Stat. Soc.,* ser. B, vol. 42, no. 1, 1980, pp. 1–25.
25. Mereau, P.; and Prevost, G.: Calculation of Uncertainty Intervals in Non Linear Estimation. *Identification and System Parameter Estimation, Volume 1,* R. Isermann, ed., Pergamon Press, 1979, pp. 365–372.
26. Mereau, P.; and Raymond, J.: *Computation of the Intervals of Uncertainties About the Parameters Found for Identification.* NASA TM-76978, 1982.
27. 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.
28. Draper, N. R.; and Smith, H.: *Applied Regression Analysis,* Second ed. John Wiley & Sons, Inc., c.1981.
29. Klein, Vladislav; and Batterson, James G.: *Determination of Airplane Model Structure From Flight Data Using Splines and Stepwise Regression.* NASA TP-2126, 1983.
30. Bard, Yonathan: *Nonlinear Parameter Estimation.* Academic Press, 1974.
31. Cramér, Harald: *Mathematical Methods of Statistics.* Princeton Univ. Press, 1946.
32. Jategaonkar, Ravindra; and Plaetschke, Ermin: *Non-Linear Parameter Estimation From Flight Test Data Using Minimum Search Methods.* DFVLR-FB 83-15, Apr. 1983.
33. Olsson, Donald M.; and Nelson, Lloyd S.: The Nelder-Mead Simplex Procedure for Function Minimization. *Technometrics,* vol. 17, no. 1, Feb. 1975, pp. 45–51.

34. Grove, Randall D.; Bowles, Roland L.; and Mayhew, Stanley 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.
35. Adams, William M., Jr.; and Tiffany, Sherwood H.: Design of a Candidate Flutter Suppression Control Law for DAST ARW-2. NASA paper presented at 1983 AIAA Conference on Guidance and Control (Gatlinburg, Tenn.), Aug. 15-17, 1983.
36. Klein, V.: *On the Adequate Model for Aircraft Parameter Estimation.* Rep. Aero. No. 28, Cranfield Inst. Technol., Mar. 1975.
37. Klein, Vladislav; and Schiess, James R.: *Compatibility Check of Measured Aircraft Responses Using Kinematic Equations and Extended Kalman Filter.* NASA TN D-8514, 1977.
38. Schweppe, Fred C.: *Uncertain Dynamic Systems.* Prentice-Hall, Inc., c.1973.

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA RP-1153		

4. Title and Subtitle	5. Report Date
A Methodology for Airplane Parameter Estimation and Confidence Interval Determination in Nonlinear Estimation Problems	April 1986
	6. Performing Organization Code: 505-34-03-02
7. Author(s): Patrick C. Murphy	8. Performing Organization Report No.: L-16009
9. Performing Organization Name and Address: NASA Langley Research Center, Hampton, VA 23665-5225	10. Work Unit No.
	11. Contract or Grant No.
12. Sponsoring Agency Name and Address: National Aeronautics and Space Administration, Washington, DC 20546-0001	13. Type of Report and Period Covered: Reference Publication
	14. Sponsoring Agency Code

15. Supplementary Notes

This report was originally a Ph.D. dissertation submitted to The George Washington University, April 1985. Vladislav Klein directed the dissertation. Editorial revisions have been made.

16. Abstract

An algorithm for maximum likelihood (ML) estimation is developed with an efficient method for approximating the sensitivities. 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. With the fitted surface, sensitivity information can be updated at each iteration with less computational effort than that required by either a finite-difference method or integration of the analytically determined sensitivity equations. MNRES eliminates the need to derive sensitivity equations for each new model, and thus provides flexibility to use model equations in any convenient format. 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. 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. Beale's measure of nonlinearity is developed in this study for airplane identification problems; it is used to empirically correct confidence levels and to predict the degree of agreement between CR bounds and search estimates.

17. Key Words (Suggested by Authors(s))	18. Distribution Statement
System identification Aircraft identification Maximum likelihood Sensitivities Optimization Parameter estimation Nonlinear estimation Confidence intervals Cramer-Rao bounds Random search	Unclassified—Unlimited Subject Category 08

19. Security Classif.(of this report)	20. Security Classif.(of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	57	A04

For sale by the National Technical Information Service, Springfield, Virginia 22161

NASA-Langley, 1986