NASA Technical Memorandum 85905

(NASA-TM-85905) MORE THAN YOU WANT TO KNOW ABOUT MAXIMUM LIKELIHCCD ESTIMATICN Final Report (NASA) 26 p HC A03/MF A01 CSCL 01C

N85-15752

Unclas
G3/08 13484

More Than You May Want To Know About Maximum Likelihood Estimation

Kenneth W. Iliff and Richard E. Maine

January 1985

NASA
National Aeronautics and
Space Administration

NASA Technical Memorandum 85905

More Than You May Want To Know About Maximum Likelihood Estimation

Kenneth W. Iliff and Richard E. Maine
NASA Ames Research Center, Dryden Flight Research Facility, Edwards, California 93523

1985

National Aeronautics and
Space Administration
Ames Research Center
Dryden Flight Research Facility
Edwards, California 93523

MORE THAN YOU MAY WANT TO KNOW ABOUT MAXIMUM LIKELIHOOD ESTIMATION

Kenneth W. Iliff* and Richard E. Maine†
NASA Ames Research Center
Dryden Flight Research Facility
Edwards, California

Abstract

The maximum likelihood estimator has been used to extract stability and control derivatives from flight data for many years. Most of the literature on aircraft estimation concentrates on new developments and applications, assuming familiarity with basic estimation concepts. This paper presents some of these basic concepts. The paper briefly discusses the maximum likelihood estimator and the aircraft equations of motion that the estimator uses. The basic concepts of minimization and estimation are examined for a simple computed aircraft example. The cost functions that are to be minimized during estimation are defined and discussed. Graphic representations of the cost functions are given to help illustrate the minimization process. Finally, the basic concepts are generalized, and estimation from flight data is discussed. Some of the major conclusions for the computed example are also developed for the analysis of flight data.

Introduction

The maximum likelihood estimator has been used to obtain stability and control estimates from flight data for nearly 20 years. The results of many applications have been reported worldwide. Reference 1 contains a representative list of some of these reports. Several good texts (including Refs. 2 and 3) contain thorough treatments of the theory of maximum likelihood estimation. Experience reports[1,4,5] pointing out practical considerations for applying the maximum likelihood estimator have also been published. Stability and control derivatives estimated from flight data are currently required for correlation studies with predictive techniques, handling qualities documentation, design compliance, aircraft simulator enhancement and refinement, and control system design. Correlation, simulation, and control system design applications are discussed in Ref. 6. Current studies have concentrated on estimation model structure determination,[7,8] equation error with state reconstruction,[9,10,11] and maximum likelihood estimation in the frequency domain.[12,13]

Most of the reports in the estimation area concentrate on new developments and applications, assuming familiarity with the basic concepts of maximum likelihood estimation. In this paper some of these basic concepts are reviewed, concentrating on simple, idealized models. These simple models provide insights applicable to a wide variety of real problems.

This paper presents some fundamentals of maximum likelihood estimation as applied to the aircraft problem. It briefly discusses the maximum likelihood estimator and the aircraft equations of motion that the estimator uses. The basic aspects of minimization and estimation are then examined in detail for a simple computed aircraft example. Finally, the discussion is expanded to the general aircraft estimation problem.

*Senior Staff Scientist. Associate Fellow AIAA.
†Aerospace Engineer. Member AIAA.

Symbols

A,B,C,D,F,G	system matrices
a_y	lateral acceleration, g
b	reference span, ft
C_ℓ	coefficient of rolling moment
C_n	coefficient of yawing moment
C_Y	coefficient of side-force
f(·), g(·)	general functions
GG*	measurement noise covariance matrix
g	acceleration due to gravity, ft/sec^2
H	approximation to the information matrix
I_x, I_y, I_z, I_{xz}	moment of inertia about subscripted axis, slug-ft^2
i	general index
J	cost function
K_β	sidewash factor
L	rolling moment divided by I_x, deg/sec^2, or number of iterations
L′	rolling moment, ft-lb
L_{YJ}	rolling moment due to yaw jet, ft-lb
m	mass, slug
N	number of time points or cases
n	state noise vector or number of unknowns
p	roll rate, deg/sec
q	pitch rate, deg/sec
$\bar{q}$	dynamic pressure, lb/ft^2
R	innovation covariance matrix
r	yaw rate, deg/sec

s	reference area, ft^2
t	time, sec
u	control input vector
V	forward velocity, ft/sec
x	state vector
$x_{a_y}, y_{a_y}, z_{a_y}$	distance between lateral accelerometer and the center of gravity along the appropriate axis, ft
z	observation vector
$\tilde{z}$	predicted Kalman-filtered estimate
α	angle of attack, deg
β	angle of sideslip, deg
Δ	time sample interval, sec
δ	control deflection, deg
δ_a	aileron deflection, deg
δ_e	elevon deflection, deg
δ_r	rudder deflection, deg
η	measurement noise vector
θ	pitch angle, deg
μ	mean
ξ	vector of unknowns
σ	standard deviation
τ	time, sec
φ	transition matrix or bank angle, deg
ψ	integral of transition matrix

Subscripts:

$p, q, r, \alpha, \dot{\alpha}, \beta, \dot{\beta}, \delta, \delta_a, \delta_r, \delta_e$	partial derivative with respect to subscripted quantity
0	bias or at time zero
m	measured quantity

Other nomenclature:

~	predicted estimate
^	estimate
*	transpose

Maximum Likelihood Estimation

The concept of maximum likelihood is discussed in this section. First the general heuristic problem is discussed, and then the specific equations for obtaining maximum likelihood estimates for the aircraft problem are given. In the following sections, both the concepts and the computations involved in a simple but realistic example are discussed in detail.

The aircraft parameter estimation problem can be defined quite simply in general terms. The system investigated is assumed to be modeled by a set of dynamic equations containing unknown parameters. To determine the values of the unknown parameters, the system is excited by a suitable input, and the input and actual system response are measured. The values of the unknown parameters are then inferred based on the requirement that the model response to the given input match the actual system response. When formulated in this manner, the problem of identifying the unknown parameters can be easily solved by many methods; however, complicating factors arise when application to a real system is considered.

The first complication results from the impossibility of obtaining perfect measurements of the response of any real system. The inevitable sensor errors are usually included as additive measurement noise in the dynamic model. Once this noise is introduced, the theoretical nature of the problem changes drastically. It is no longer possible to exactly identify the values of the unknown parameters; instead, the values must be estimated by some statistical criterion. The theory of estimation in the presence of measurement noise is relatively straightforward for a system with discrete time observations, requiring only basic probability.

The second complication of real systems is the presence of state noise. State noise is random excitation of the system from unmeasured sources, the standard example for the aircraft stability and control problem being atmospheric turbulence. If state noise is present and measurement noise is neglected, the analysis results in the regression algorithm.

When both state and measurement noise are considered, the problem is more complex than in the cases that have only state noise or only measurement noise. Reference 14 develops the Maine-Iliff formulation of the maximum likelihood estimator in continuous/discrete time, which accounts for both state and measurement noise. This formulation has a continuous system model with discrete sampled observations.

The final problem for real systems is modeling. It has been assumed throughout the above discussion that for some value (called the "correct" value) of the unknown parameter vector, the system is correctly described by the dynamic model. Physical systems are seldom described exactly by simple dynamic models, so the question of modeling error arises. No comprehensive theory of modeling error is available. The most common approach is to ignore it: Any modeling error is simply treated as state noise or measurement noise, or both, in spite of the fact that the modeling error may be deterministic rather than random. The assumed noise statistics can then be adjusted to include the contribution of the modeling error. This procedure is not rigorously justifiable, but combined with a carefully chosen model, it is probably the best approach available.

With the above discussion in mind, it is possible to make a more precise, mathematically probabilistic statement of the parameter estimation problem. The first step is to define the general system model (aircraft equations of motion). This model can be written in the continuous/discrete form as

$$x(t_0) = x_0 \tag{1}$$

$$\dot{x}(t) = f[x(t),u(t),\xi] + F(\xi)n(t) \tag{2}$$

$$z(t_i) = g[x(t_i),u(t_i),\xi] + G(\xi)\eta_i \tag{3}$$

where x is the state vector, z is the observation vector, f and g are system state and observation functions, u is the known control input vector, ξ is the unknown parameter vector, n is the state noise vector, and η is the measurement noise vector. The state noise vector is assumed to be zero-mean white Gaussian and stationary, and the measurement noise vector is assumed to be a sequence of independent Gaussian random variables with zero mean and identity covariance. For each possible estimate of the unknown parameters, a probability that the aircraft response time histories attain values near the observed values can then be defined. The maximum likelihood estimates are defined as those that maximize this probability. Maximum likelihood estimation has many desirable statistical characteristics; for example, it yields asymptotically unbiased, consistent, and efficient estimates.[15]

If there is no state noise and the matrix G is known, then the maximum likelihood estimator minimizes the cost function

$$J(\xi) = \frac{1}{2}\sum_{i=1}^{N}[z(t_i) - \tilde{z}_\xi(t_i)]^*(GG^*)^{-1}[z(t_i) - \tilde{z}_\xi(t_i)] + \frac{1}{2}N \ln |(GG^*)| \tag{4}$$

where GG* is the measurement noise covariance matrix, and $\tilde{z}_\xi(t_i)$ is the computed response estimate of z at t_i for a given value of the unknown parameter vector ξ. The cost function is a function of the difference between the measured and computed time histories.

If Eqs. (2) and (3) are linearized (as is the case for the stability and control derivatives in the aircraft problem),

$$x(t_0) = x_0 \tag{5}$$

$$\dot{x}(t) = Ax(t) + Bu(t) + Fn(t) \tag{6}$$

$$z(t_i) = Cx(t_i) + Du(t_i) + G\eta_i \tag{7}$$

For the no-state-noise case, the $\tilde{z}_\xi(t_i)$ term of Eq. (4) can be approximated by

$$\tilde{x}_\xi(t_0) = x_0(\xi) \tag{8}$$

$$\tilde{x}_\xi(t_{i+1}) = \phi\tilde{x}_\xi(t_i) + \psi[u(t_i) + u(t_{i+1})]/2 \tag{9}$$

$$\tilde{z}_\xi(t_i) = C\tilde{x}_\xi(t_i) + Du(t_i) \tag{10}$$

where

$$\phi = \exp\,[A(t_{i+1} - t_i)]$$

$$\psi = \int_{t_i}^{t_{i+1}} \exp\,(A\tau)\,d\tau\,B$$

When state noise is important, the nonlinear form of Eqs. (1) to (3) is intractable. For the linear model defined by Eqs. (5) to (7), the cost function that accounts for state noise is

$$J(\xi) = \frac{1}{2}\sum_{i=1}^{N}[z(t_i) - \tilde{z}_\xi(t_i)]^*R^{-1}[z(t_i) - \tilde{z}_\xi(t_i)] + \frac{1}{2}N \ln |R| \tag{11}$$

where R is the innovation covariance matrix. The $\tilde{z}_\xi(t_i)$ term in Eq. (11) is the Kalman-filtered estimate of z, which, if the state noise covariance is zero, reduces to the form of Eq. (4). If there is no state noise, the second term of Eq. (11) is of no consequence, (unless one wishes to include elements of the G matrix) and R can be replaced by GG* which makes Eq. (11) the same as Eq. (4).

To minimize the cost function $J(\xi)$, we can apply the Newton-Raphson algorithm which chooses successive estimates of the vector of unknown coefficients, $\hat{\xi}$. Let L be the iteration number. The L + 1 estimate of $\hat{\xi}$ is then obtained from the L estimate as follows:

$$\hat{\xi}_{L+1} = \hat{\xi}_L - [\nabla_\xi^2 J(\hat{\xi}_L)]^{-1}[\nabla_\xi^* J(\hat{\xi}_L)] \tag{12}$$

If R is assumed fixed the first and second gradients are defined as

$$\nabla_\xi J(\xi) = -\sum_{i=1}^{N}[z(t_i) - \tilde{z}_\xi(t_i)]^*(GG^*)^{-1}[\nabla_\xi \tilde{z}_\xi(t_i)] \tag{13}$$

$$\nabla_\xi^2 J(\xi) = \sum_{i=1}^{N}[\nabla_\xi \tilde{z}_\xi(t_i)]^*(GG^*)^{-1}[\nabla_\xi \tilde{z}_\xi(t_i)] - \sum_{i=1}^{N}[z(t_i) - \tilde{z}_\xi(t_i)]^*(GG^*)^{-1}[\nabla_\xi^2 \tilde{z}_\xi(t_i)] \tag{14a}$$

The Gauss-Newton approximation to the second gradient is

$$\nabla_\xi^2 J(\xi) \cong \sum_{i=1}^{N}[\nabla_\xi \tilde{z}_\xi(t_i)]^*(GG^*)^{-1}[\nabla_\xi \tilde{z}_\xi(t_i)] \tag{14b}$$

The Gauss-Newton approximation, which is sometimes referred to as modified Newton-Raphson, is computationally much easier than the Newton-Raphson approximation because the second gradient of the innovation never needs to be calculated. In addition, it can have the advantage of speeding the convergence of the algorithm, as is discussed in the Simple Aircraft Example section.

Figure 1 illustrates the maximum likelihood estimation concept. The measured response of the aircraft is compared with the estimated response, and the difference between these responses is called the response error. The cost functions of Eqs. (4) and (11) include this response error. The Gauss-Newton computational algorithm is used to find the coefficient values that maximize the cost function. Each iteration of this algorithm provides a new estimate of the unknown coefficients on the basis of the response error. These new estimates of the coefficients are then used to update the mathematical model of the aircraft, providing a new estimated response and, therefore, a new response error. The updating of the mathematical model continues iteratively until a convergence criterion is satisfied. The estimates resulting from this procedure are the maximum likelihood estimates.

The maximum likelihood estimator also provides a measure of the reliability of each estimate based on the information obtained from each dynamic maneuver. This measure of the reliability, analogous to the standard deviation, is called the Cramèr-Rao bound[16] or the uncertainty level. The Cramèr-Rao bound as computed by current programs should generally be used as a measure of relative accuracy rather than absolute accuracy. The bound is obtained from the approximation of the information matrix, H. This matrix equals the approximation to the second gradient given by Eq. (14b). The bound for each unknown is proportional to the the square root of the corresponding diagonal element of H. That is for the ith unknown, the Cramèr-Rao bound is $\sqrt{H(i,i)}$.

The formulation and minimization algorithm discussed above is implemented with the Iliff-Maine code (MMLE3 maximum likelihood estimation program). The program and computational algorithms are described fully in Ref. 17. All the computations shown and described in the remainder of this paper use the algorithms exactly as described in Ref. 17.

Aircraft Equations of Motion

For the discussion that follows in later sections of this paper, some knowledge of the aircraft equations of motion is assumed. To clarify some of that discussion, the aircraft equations are discussed briefly in this section.

Generalized nonlinear equations of motion are given in detail in Ref. 17, which fully describes the Iliff-Maine code (MMLE3 program). All computations and aircraft examples in this paper use the linearized form for the lateral-directional equations. These equations are given below and referred to in the remainder of the paper.

$$\dot{\beta} = \frac{\bar{q}s}{mV}(C_Y + \dot{\beta}_0) + \frac{g}{V}\cos\theta\sin\phi + p\sin\alpha - r\cos\alpha \qquad (15)$$

$$\dot{p}I_x - \dot{r}I_{xz} = \bar{q}sbC_\ell + qr(I_y - I_z) + pqI_{xz} \qquad (16)$$

$$\dot{r}I_z - \dot{p}I_{xz} = \bar{q}sbC_n + pq(I_x - I_y) - qrI_{xz} \qquad (17)$$

$$\dot{\phi} = p + r\cos\phi\tan\theta + q\sin\phi\tan\theta + \dot{\phi}_0 \qquad (18)$$

where

$$C_Y = C_{Y_\beta}\beta + C_{Y_p}\frac{pb}{2V} + C_{Y_r}\frac{rb}{2V} + C_{Y_\delta}\delta + C_{Y_0} \qquad (19)$$

$$C_\ell = C_{\ell_\beta}\beta + C_{\ell_p}\frac{pb}{2V} + C_{\ell_r}\frac{rb}{2V} + C_{\ell_\delta}\delta + C_{\ell_0} + C_{\ell_{\dot{\beta}}}\frac{\dot{\beta}b}{2V} \qquad (20)$$

$$C_n = C_{n_\beta}\beta + C_{n_p}\frac{pb}{2V} + C_{n_r}\frac{rb}{2V} + C_{n_\delta}\delta + C_{n_0} + C_{n_{\dot{\beta}}}\frac{\dot{\beta}b}{2V} \qquad (21)$$

where the δ term is summed over all controls.

The observation equations are

$$\beta_m = K_\beta\left(\beta - \frac{z_\beta}{V}p + \frac{x_\beta}{V}r\right) \qquad (22)$$

$$p_m = p \qquad (23)$$

$$r_m = r \qquad (24)$$

$$\phi_m = \phi \qquad (25)$$

$$a_{y_m} = \frac{\bar{q}s}{mg}C_Y - \frac{z_{a_Y}}{g}\dot{p} + \frac{x_{a_Y}}{g}\dot{r} - \frac{y_{a_Y}}{g}(p^2 + r^2) \qquad (26)$$

$$\dot{p}_m = \dot{p} + \dot{p}_0 \qquad (27)$$

$$\dot{r}_m = \dot{r} + \dot{r}_0 \qquad (28)$$

The state, control, and observation vectors for the lateral-directional mode can then be defined as

$$x = (\beta\ p\ r\ \phi)^* \qquad (29)$$

$$u = (\delta_a\ \delta_r)^* \qquad (30)$$

$$z = (\beta_m\ p_m\ r_m\ \phi_m\ a_{y_m}\ \dot{p}_m\ \dot{r}_m)^* \qquad (31)$$

Simple Aircraft Example

The basic concepts involved in a parameter estimation problem can be illustrated by using a simple example representative of a realistic aircraft problem. The example chosen here is representative of an aircraft that exhibits pure rolling motion from an aileron input. This example, although simplified, typifies the motion exhibited by many aircraft in particular flight regimes, such as the F-14 aircraft flying at high dynamic pressure, the F-111 aircraft at moderate speeds with the wing in the forward position, and the T-37 aircraft at low speed.

Derivation of an equation describing this motion is straightforward. Figure 2 shows a sketch of an aircraft with the x-axis perpendicular to the plane of the figure (positive forward on the aircraft). The rolling moment (L'), roll rate (p), and aileron deflection (δ_a) are positive as shown.

For this example, the only state is p and the only control is δ_a. The result of summing moments is

$$I_x\dot{p} = L'(p,\delta_a) \qquad (32)$$

The first-order Taylor expansion then becomes

$$\dot{p} = L_p p + L_{\delta_a}\delta_a \qquad (33)$$

where

$$L' = I_x L$$

Since the aileron is the only control, it is notationally simpler to use δ instead of δ_a for the discussion of this example. Equation (33) can then be written as

$$\dot{p} = L_p p + L_\delta \delta \qquad (34)$$

An alternate approach that results in the same equation is to combine Eq. (16) with Eq. (20), substituting for C_ℓ, and then eliminate the terms that are zero for our example. This yields

$$\dot{p}I_x = \bar{q}sb\left(C_{\ell_p}\frac{pb}{2V} + C_{\ell_\delta}\delta\right) \qquad (35)$$

where p is the roll rate and δ is the aileron deflection. Rearranging terms, the equation can be put into the dimensional derivative form of Eq. (34).

Equation (34) is a simple aircraft equation where the forcing function is provided by the aileron and the damping by the damping-in-roll term, L_p. In subsequent sections we examine in detail the parameter estimation problem where Eq. (34) describes the system. For this single-degree-of-freedom problem, the maximum likelihood estimator is used to estimate either L_p or L_δ or both for a given computed time history.

We will assume that the system has measurement noise, but no state noise as in Eqs. (1), (2), and (3). Equation (4) then gives the cost function for maximum likelihood estimation. The weighting GG* is unimportant for this problem, so let it equal 1. For our example, Eqs. (2) and (3) become $x_i = p_i$ and $z_i = x_i$. Therefore, Eq. (4) becomes

$$J(L_p,L_\delta) = \frac{1}{2}\sum_{i=1}^{n}[p_i - \tilde{p}_i(L_p, L_\delta)]^2 \qquad (36)$$

where p_i is the value of the measured response p at time t_i and $\tilde{p}_i(L_p,L_\delta)$ is the computed time history of $\tilde{p}$ at time t_i for $L_p = \hat{L}_p$ and $L_\delta = \hat{L}_\delta$. Throughout the rest of the paper, where computed data (not flight) are used, the measured time history refers to p_i, and the computed time history refers to $\tilde{p}_i(L_p,L_\delta)$. The computed time history is a function of the current estimates of L_p and L_δ, but the measured time history is not.

The most straightforward method of obtaining $\tilde{p}_i$ is with Eqs. (8) and (9). In terms of the notation stated above,

$$\tilde{p}_{i+1} = \phi\tilde{p}_i + \psi(\delta_i + \delta_{i+1})/2 \qquad (37)$$

where

$$\phi = \exp(L_p\Delta)$$

$$\psi = \int_0^\Delta \exp(L_p\tau)\,d\tau\, L_\delta = \frac{L_\delta[1 - \exp(L_p\Delta)]}{L_p}$$

and Δ is the length of the sample interval $(t_{i+1} - t_i)$. Simplifying the notation

$$\delta_{i+1/2} = (\delta_i + \delta_{i+1})/2 \qquad (38)$$

then

$$\tilde{p}_{i+1} = \phi\tilde{p}_i + \psi\delta_{i+1/2} \qquad (39)$$

The maximum likelihood estimate is obtained by minimizing Eq. (36). The Gauss-Newton method described earlier is used for this minimization. Equation (12) is used to determine successive values of the estimates of the unknowns during the minimization.

For this simple problem, $\hat{\xi} = [\hat{L}_p\ \hat{L}_\delta]^*$ and successive estimates of $\hat{L}_p$ and $\hat{L}_\delta$ are determined by updating Eq. (12). The first and second gradients of Eq. (12) are defined by Eqs. (13) and (14b). The complete set of equations are given in Ref. 17.

The entire procedure can now be written for obtaining the maximum likelihood estimates for this simple example. To start the algorithm, an initial estimate of L_p and L_δ is needed. This is the value of $\hat{\xi}_0$. With Eq. (12), $\hat{\xi}_1$ and subsequently $\hat{\xi}_L$ are defined by using the first and second gradients of $J(L_p,L_\delta)$ from Eq. (36). The gradients for this particular example from Eqs. (13) and (14b) are

$$\nabla_\xi J(\hat{\xi}_L) = -\sum_{i=1}^{N}(p_i - \tilde{p}_i)\nabla_\xi \tilde{p}_i \qquad (40)$$

$$\nabla_\xi^2 J(\hat{\xi}_L) \cong \sum_{i=1}^{N}(\nabla_\xi\tilde{p}_i)^*(\nabla_\xi\tilde{p}_i) \qquad (41)$$

With the specific equations defined in this section for this simple example, we can now proceed in the next section to the computational details of a specific example.

Computational Details of Minimization

In the previous section we specified the equations for a simple example and described the procedure for obtaining estimates of the unknowns from a dynamic maneuver. In this section we give the computational details for obtaining the estimates. Some of the basic concepts of parameter estimation are best shown with computed data where the correct answers are known. Therefore, in this section we study two examples involving computed time histories. The first example is based on data that have no measurement noise, which results in estimates that are the same as the correct value. The second example contains significant measurement noise; consequently, the estimates are not the same as the correct values. Throughout the rest of the paper, where computed data is used, the term "no-noise case" is used for the case with no noise added and "noisy case" for the case where noise has been added.

Since we are studying a simple computed example, it is desirable to keep it simple enough to complete some or all of the calculations on a home computer or, with some labor, on a calculator. With this in mind, the number of data points needs to be kept small. For this computed example, 10 points (time samples) are used. The simulated data, which we refer to as the measured data, are based on Eq. (34). We use the same correct values of L_p and L_δ (-0.2500 and 10.0, respectively) for both examples. In addition, the same input (δ) is used for both examples, the sample interval (Δ) is 0.2 sec, and the initial conditions are zero. Tables of all the significant intermediate values are given for each example. These values are given to four significant digits, although to obtain exactly the same values with a computer or calculator requires the use of 13 significant digits, as in the computation of these tables. If the four-digit numbers are used in the computation, the answers will be a few tenths of a percent off, but will still serve to illustrate the minimization accuracy. In both examples, the initial values of L_p and L_δ (or $\hat{\xi}_0$) are -0.5 and 15.0, respectively.

Example With No Measurement Noise

The measurement time history for no measurement noise (no-noise case) is shown in Fig. 3. The aileron input starts at zero, goes to a fixed value, and then returns to zero. The resulting roll-rate time history is also shown. The values of the measured roll rate to 13 significant digits are given in Table 1 along with the aileron input.

Table 2 shows the values for $\hat{L}_p$, $\hat{L}_\delta$, and J for each iteration, along with the values of ϕ and ψ needed for calculations of $\tilde{p}_i$. In three iterations the algorithm converges to the correct values to four significant digits for both L_p and L_δ. $\hat{L}_\delta$ overshoots slightly on the first iteration and then comes quickly to the correct answer. $\hat{L}_p$ overshoots slightly on the second iteration.

Figure 4 shows the match between the measured data and the computed data for each of the first three iterations. The match is very good after two iterations. The match is nearly exact after three iterations.

Although the algorithm has converged to four-digit accuracy in L_p and L_δ, the value of the cost function, J, continues to decrease rapidly between iterations 3 and 4. This is a consequence of using the maximum likelihood estimator on data with no measurement noise. Theoretically, using infinite accuracy the value of J at the minimum should be zero. However, with finite accuracy the value of J becomes small but never quite zero. This value is a function of the number of significant digits that are being used. For the 13-digit accuracy used here, the cost eventually decreases to approximately 0.3×10^{-28}.

Example With Measurement Noise

The data used in this example (noisy case) are the same as those used in the previous section, except that pseudo-Gaussian noise has been added to the roll rate. The time history is shown in Fig. 5. The signal-to-noise ratio is quite low in this example, as is readily apparent by comparing Figs. 3 and 5. The exact values of the time history to 13-digit accuracy are shown in Table 3. The values of $\hat{L}_p$, $\hat{L}_\delta$, ϕ, ψ, and J are shown for each iteration in Table 4. The algorithm converges in four iterations. The behavior of the coefficients as they approach convergence is much like the no-noise case. The most notable result of this case is the converged values of $\hat{L}_p$ and $\hat{L}_\delta$, which are somewhat different from the correct values. The match between the measured and computed time history is shown in Fig. 6 for each iteration. No change in the match is apparent for the last two iterations. The match is very good considering the amount of measurement noise.

In Fig. 7, the computed time history for the no noise estimates of L_p and L_δ is compared to that for the noisy-case estimates of L_p and L_δ. Because the algorithm converged to values somewhat different than the correct values, the two computed time histories are similar but not identical.

The accuracy of the converged elements can be assessed by looking at the Cramèr-Rao inequality[16,17] discussed earlier. The Cramèr-Rao bound can be obtained from the information matrix corrected for observed noise amplitude as follows.

$$H = \frac{2}{\xi}(J_{minimum})(\nabla^2 J)^{-1}/(N-1)$$

The Cramèr-Rao bounds for L_p and L_δ are the square roots of the diagonal elements of the H matrix, or $\sqrt{H(1,1)}$ and $\sqrt{H(2,2)}$, respectively. The Cramèr-Rao bounds are 0.1593 and 1.116 for L_p and L_δ, respectively. The errors in $\hat{L}_p$ and $\hat{L}_\delta$ are less than the bounds.

Cost Functions

In the previous section we obtained the maximum likelihood estimates for computed time histories by minimizing the values of the cost function. To fully understand what occurs in this minimization, we must study in more detail the form of the cost functions and some of their more important characteristics. In this section, the cost function for the no-noise case is discussed briefly. The cost function of the noisy case is then discussed in more detail. The same two time histories studied in the previous section are examined here. The noisy case is more interesting because it has a meaningful Cramèr-Rao bound and is more representative of aircraft flight data.

First we will look at the one-dimensional case where L_δ is fixed at the correct value, because it is easier to grasp some of the characteristics of the cost function in one dimension. Then we will look at the two-dimensional case, where both L_p and L_δ are varying. It is important to remember that everything shown in this paper on cost functions is based on computed time histories that are defined by Eq. (36). For every time history we might choose (computed or flight data), a complete cost function is defined. For the ca·, of n variables, the cost function defines a hypersurface of n + 1 dimensions. It might occur to us that we could just construct this surface and look for the minimum, avoiding the need to bother with the minimization algorithm. This is not a reasonable approach because, in general, the number of variables is greater than two. Therefore, the cost function can be described mathematically but not pictured graphically.

One-Dimensional Case

To illustrate the many interesting aspects of cost functions, it is easiest to first look at cost functions having one variable. In an earlier section, the cost function of L_p and L_δ was minimized. That cost function is most interesting in the L_p direction. Therefore, the one-variable cost function studied here is $J(L_p)$. All discussions in this section are for $J(L_p)$ with L_δ equal to the correct value of 10. Figure 8 shows the cost function plotted as a function of L_p for the case where there is no measurement noise (no-noise case). As expected for this case, the minimum cost is zero and occurs at the correct value of $L_p = -0.2500$. It is apparent that the cost increases much more slowly for a more negative L_p than for a positive L_p. In fact, the slope of the curve tends to become less negative where L_p is more negative than -1.0. Physically this makes sense since the more negative values of L_p represent cases of high damping, and the positive L_p represents an unstable system. Therefore, the p_i for positive L_p becomes increasingly different from the measured time history for small positive increments in L_p. For very large damping (very negative L_p), the system would show essentially no response. Therefore, large increases in damping result in relatively small changes in the value of $J(L_p)$.

In Fig. 9, the cost function based on the time history with measurement noise (noisy case) is plotted as a function of L_p. The correct value of L_p (-0.2500) and the value of L_p (-0.3218) at the minimum of the cost (3.335) are both indicated on the figure. The general shape of the cost function in Fig. 9 is similar to that shown in Fig. 8. Figure 10 shows the comparison between the cost functions based on the noisy and no-noise cases. The comments relating to the cost function of the no-noise case also apply to the cost function based on the noisy case. Figure 10 shows clearly that the two cost functions are shifted by the difference in the value of L_p at the minimum and increased by the difference in the minimum cost. One would expect only a small difference in the value of the cost when far from the minimum. This is because the "estimated" time history is so far from the measured time history that it becomes irrelevent as to whether the measured time history has noise added. Therefore, for large values of cost, the difference in the two cost functions should be small in comparison to the total cost.

Figure 11 shows the gradient of $J(L_p)$ plotted as a function of L_p for the noisy case. This is the function for which we were trying to find the zero (or equivalently, the minimum of the cost function) using the Gauss-Newton method that is discussed in a previous section. The gradient is zero at $L_p = -0.3218$, which corresponds to the value of the minimum of $J(L_p)$.

The difference between the Newton-Raphson method (Eq. (14a)) and the Gauss-Newton method (Eq. 14b)) of minimization has been mentioned previously. For this simple one-dimensional case, we can easily compute the second gradient both with the second term of Eq. (14a) (Newton-Raphson), and without the second term (Gauss-Newton, Eq. (14b,). Figure 12 shows a comparison between the Newton-Raphson and the Gauss-Newton approximation second gradients. The Gauss-Newton second gradient (dashed line) always remains positive because it is the sum of quadratic terms (squared for the one-dimensional example). The Newton-Raphson second gradient can be positive or negative, depending upon the value of the second partial derivative with respect to L_p. Other than the difference in sign for the more negative L_p, the two curves have similar shapes.

As stated earlier, the Gauss-Newton method can be shown to be superior to Newton-Raphson in certain cases. We can demonstrate obvious cases of this with our example. An easy way to select a spot where problems with the Newton-Raphson method will occur is to look for places where the second gradient (slope of the gradient) is near zero or negative. Figure 11 has such a region near $L_p = -1.0$. If we choose a point where the gradient slope is exactly zero, we are forced to divide by zero in Eq. (12) with the Newton-Raphson method. This point is at $L_p = -1.13$ in Fig. 12. If the value of the slope of the gradient is negative, then the Newton-Raphson method will go to very negative values of L_p. For very negative values of L_p, the cost becomes asymptotically constant and the gradient becomes nearly zero. In that region, the Newton-Raphson algorithm would diverge to negative infinity. If the slope of the gradient is positive but small, we still have a problem with the Newton-Raphson method. Figure 13 shows the first iteration starting from $L_p = -0.95$ for both Gauss-Newton and Newton-Raphson. The Newton-Raphson method selects a point where the tangent of the gradient at $L_p = -0.95$ intersects the zero line. This results in the selection of an L_p of approximately 2.6 in the first iteration. From that value it requires many iterations to return to the actual minimum. On the other hand, the Gauss-Newton method selects a value for L_p of approximately -0.09 and converges to the minimum to four-digit accuracy in two more iterations. With more complex examples a comparison of the convergence properties of the two algorithms becomes more difficult to visualize, but the problems are generalizations of the situation we have observed with the one-dimensional example.

The usefulness of the Cramèr-Rao bound was discussed in the Examples With Measurement Noise section. At this point it is useful to digress briefly to discuss some of the ramifications of the Cramèr-Rao bound for the one-dimensional case. The Cramèr-Rao bound only has meaning for the noisy case. In the noisy example, the estimate of L_p is -0.3218 and the Cramèr-Rao bound is 0.0579. The calculation of the Cramèr-Rao bound was defined in the previous section for both the one-dimensional and two-dimensional examples. The Cramèr-Rao bound is an estimate of the standard deviation of the estimate. One would expect the scatter in the estimates of L_p to be of about the same magnitude as the estimate of the standard deviation. For the one-dimensional case discussed here, the range (L_p (-0.3218) plus or minus the Cramèr-Rao bound (0.0579)) nearly includes the correct value of L_p (-0.2500). If noisy cases are generated for many time histories (adding different measurement noise to each time history), then the sample mean and sample standard deviation of the estimates for these cases can be calculated. Table 5 gives the sample mean, sample standard deviation, and the standard deviation of the sample mean (standard deviation divided by the square root of the number of cases) for 5, 10, and 20 cases. The sample mean, as expected, gets closer to the correct value of -0.2500 as the number of cases increases. This is also reflected by the decreasing values in column 3 of Table 5, which are estimates of the error in the sample mean. Column 2 of Table 5 shows the sample standard deviations, which indicate the approximate accuracy of the individual estimates. This standard deviation, which stays more or less constant, is approximately equal to the Cramèr-Rao bound for the noisy case being studied here. In fact, the Cramèr-Rao bounds for each of the 20 noisy cases used here (not shown in the table) do not change much from the values found for the noisy case being studied. Both of these results are in good agreement with the theoretical characteristics[16] of the Cramèr-Rao bounds and maximum likelihood estimators in general.

The examples shown here indicate the value of obtaining more sample time histories (maneuvers). More samples improve confidence in the estimate of the unknowns. The same result holds true in analyzing actual flight time histories (maneuvers); thus it is always advisable to obtain several maneuvers at a given flight condition to improve the best estimate of each derivative.

The size of the Cramèr-Rao bounds and of the error between the correct value and the estimated value of L_p is determined to a large extent by the length of the time history and the amount of noise added to the correct time history. For the example being studied here, it is apparent from Fig. 5 that the amount of noise being added to the time history is large. The effect of the power of the measurement noise (GG^*, Eqs. (3) and (4)) on the estimate of L_p for the time history is given in Table 6. The estimate of L_p is much improved by decreasing the measurement noise power. A reduction in the value of G to one-tenth of the value in the noisy example being studied yields an acceptable estimate of L_p. For flight data, the measurement noise is reduced by improving the accuracy of the output of the measurement sensors.

Two-Dimensional Case

In this section the cost function (dependent on both L_p and L_δ) is studied. The no-noise case is examined first, followed by the noisy case.

No-Noise Case. Even though the cost function is a function of only two unknowns, it becomes much more difficult to visualize than the one-unknown case. The cost function over a reasonable range of L_p and L_δ is shown in Fig. 14. The cost increases very rapidly in the region of positive L_p and large values of L_δ. The reason is just an extension of the argument for positive L_p given in the previous section. The shape of the surface can be depicted in greater detail if we examine only the values of the cost function less than 200 for L_p less than 1.0. Figure 15 shows a view of this restricted surface from the upper end of the surface. The minimum must lie in the curving valley that gets broader as we go to the far side of the surface. Now that we have a picture of the surface, we can look at the isoclines of constant cost on the L_p-versus-L_δ plane. These isoclines are shown in Fig. 16. The minimum of the cost function is inside the closed isocline. The steepness of the cost function in the positive-L_p direction is once again apparent. Inside the closed isocline the shape is more nearly elliptical, indicating that the cost is nearly quadratic here, so fairly rapid convergence in this region would be expected. The L_p axis becomes an asymptote in cost as L_δ approaches zero. The cost is constant for $L_\delta = 0$ because no response would result from any aileron input. The estimated response is zero for all values of L_p, resulting in constant cost.

Figure 16 shows the minimum value of the cost function, which, as seen in the earlier example (Table 2), occurs at the correct values for L_p and L_δ of -0.2500 and 10, respectively. This is also evident by looking at the cost function surface shown in Fig. 17. The surface has its minimum at the correct value. As expected, the value of the cost function at the minimum is zero.

Noisy Case. As shown before in the one-dimensional case, the primary difference between the cost functions for the no-noise and noisy cases was a shift in the cost function. In that instance, the noisy case was shifted so that the minimum was at a higher cost and a more negative value of L_p. In the two-dimensional case, the no-noise and noisy cost functions exhibit a similar shift. For two dimensions the shift is in both the L_p and L_δ directions. The shift is small enough that the difference between the two cost functions is not visible at the scale shown in Fig. 14 or from the perspective of Fig. 15. Figure 18 shows the isoclines of constant cost for the noisy case. The figure looks much like the isoclines for the no-noise case shown in Fig. 16. The difference between Figs. 16 and 18 is a shift in L_p of about 0.1. This is the difference in the value of L_p at the minimum for the no-noise and noisy cases. Heuristically, one can see that the same would be true for cases with more than two unknowns. The primary difference between the two cost functions is near the minimum.

The next logical part of the cost function to examine is near the minimum. Figure 19 shows the same view of the cost function for the noisy case as was shown in Fig. 17 for the no-noise case. The shape is roughly the same as that shown in Fig. 17, but the surface is shifted such that its minimum lies over $L_p = -0.3540$ and $L_\delta = 10.24$, and is shifted upward to a cost function value of approximately 3.3.

To get a more precise idea of the cost of the noisy case near the minimum, we once again need to examine the iscolines. The isoclines (Fig. 20) in this region are much more like ellipses than they are in Figs. 16 and 18. We can follow the path of the minimization example used before by including the results from Table 4 on Fig. 20. The first iteration ($L = 1$) brought the values of L_p and L_δ very close to the values at the minimum. The next iteration essentially selected the values at the minimum when viewed at this scale. One of the reasons the convergence is so rapid in this region is that the isoclines are nearly elliptical, demonstrating that the cost is very nearly quadratic in this region. If we had started the Gauss-Newton algorithm at a point where the isoclines are much less elliptical (as in some of the border regions in Fig. 18), the convergence would have been much slower initially, but much the same as it entered the nearly quadratic region of the cost function.

Before concluding our examination of the two-dimensional case, we need to examine the Cramèr-Rao bound. Figure 21 shows the uncertainty ellipsoid, which is based on the Cramèr-Rao bounds defined in an earlier section. The relationships between the Cramèr-Rao bound and the uncertainty ellipsoid are discussed in Ref. 16. The uncertainty ellipsoid almost includes the correct value of L_p and L_δ. The Cramèr-Rao bound for L_p and L_δ can be determined from the projection of the uncertainty ellipsoid onto the L_p and L_δ axes, and compared with the values given earlier, which were 0.1593 and 1.116 for L_p and L_δ, respectively.

Estimation Using Flight Data

In the previous several sections we examined the basic mechanics of obtaining maximum likelihood estimates from computed examples with one or two unknown parameters. Now that we have a grasp of these basics, we can explore the estimation of stability and control derivatives from actual flight data. For the computationally much more difficult situation usually encountered using actual flight data, we will obtain the maximum likelihood estimates with the Illiff-Maine code (MMLE3 program) described in Ref. 17. The equations of motion that are of interest are given in the Aircraft Equations of Motion section of this paper; the remainder of the equations are given in Ref. 17.

In general, flight data estimation is fairly complex, and programs such as the Iliff-Maine code must usually be used to assist in the analysis. However, one must still be cautious about accepting the results; that is, the estimates must fit the phenomenology, and the match between the measured and computed time histories must be acceptable. This is true in all flight regimes, but one must be particularly careful in potential problem situations such as (1) in separated flow at high Mach numbers or high angle of attack, (2) with unusual aircraft configurations such as the oblique wing,[18] or (3) with modern high-performance aircraft with high-gain feedback loops. In any of the above cases, one should be particularly careful where there are even small anomalies in the match. These anomalies may indicate ignored terms in the equations of motion, separated flow, nonlinearities, sensor problems, or any of a long list of other problems.

The following brief examples are intended to show how the above caveats and the computed examples of previous sections can be used to assist in the analysis. In the computed example, the desirability of low-noise sensors, an adequate model, and several maneuvers at a given flight condition is shown.

Hand Calculation Example

Sometimes evaluation of a fairly complex flight maneuver can be augmented with a simple hand calculation. One example of this can be found for the space shuttle. The space shuttle is a large double-delta-winged vehicle designed to enter the atmosphere from space and land horizontally. The entry control system consists of 12 vertical reaction-control-system (RCS) jets (six up-firing and six down-firing), 8 horizontal RCS jets (four left-firing and four right-firing), 4 elevon surfaces, a body flap, and a split rudder surface. The locations of these devices are shown in Fig. 22. The vertical jets and the elevons are used for both pitch and roll control. The jets and elevons are used symmetrically for pitch control and asymmetrically for roll control. The space shuttle control system is described briefly in Ref. 6.

The shuttle example used here is from a maneuver obtained at a Mach number of approximately 21 and an angle of attack of approximately 40°. The controls being used for this lateral-directional maneuver are the differential elevons and the side-firing jets (yaw jets). The maneuver is shown in Fig. 23. Equations (15) to (31) describe the equations of motion. A simplified approach can be used to determine some of the derivatives by hand. The approach is one that has been used since the beginning of dynamic analysis of flight maneuvers. In particular, for this maneuver the slope of the rates can be used to determine the yaw jet control derivatives. This is possible for this example, even with a high-gain feedback system, because the yaw jets are essentially step functions and the slope of rates p and r can be determined before the vehicle and the differential elevon (aileron) responses become significant. The rolling moment due to yaw jet (L_{YJ}) is particularly important for the shuttle and is, in general, more difficult to obtain than the more dominant yawing moment due to yaw jet. Therefore, as an illustrative example, L_{YJ} is determined by hand. Figure 24 shows yaw jet and smoothed roll rate plotted at expanded scales. The equation for L_{YJ} is given by

$$L_{YJ} = \dot{p} I_x / (\text{Number of yaw jets}) \qquad (42)$$

$$\dot{p} \cong \Delta p/\Delta t = \left(\frac{-.07}{57.3}\right) \div (0.1) \qquad (43)$$

Therefore, given that $I_x \cong 900{,}000$ slug-ft^2 and the number of yaw jets is 4, $L_{YJ} \cong -2750$ ft-lb.

The same maneuver was analyzed with MMLE3, and the resulting match is shown in Fig. 25. The match is very good except for a small mismatch in p at about 6 sec. This small mismatch was studied separately with MMLE3 and found to be caused by a nonlinearity in the aileron derivative. The value from MMLE3 for L_{YJ} is -2690 ft-lb, which for the accuracy used here is essentially the same value as obtained by the simplified method. The aileron derivatives would be difficult to determine as accurately as the yaw jet derivatives. Although good estimates can seldom be obtained with the slope method discussed here, rough estimates can usually be obtained to gain some insight into values obtained with MMLE3 (or any other maximum likelihood program). These rough estimates can then be used to help explain unexpected values of estimates from an estimation program.

Sometimes a flight example becomes too complex to get anything other than qualitative estimates by hand. The determination of the rudder derivative for the F-8 aircraft with the yaw augmentation system on demonstrates this difficulty. Figure 26 shows an example of this difficulty for the F-8 data. This example, taken from Ref. 19, includes an aileron pulse and a rudder pulse. Although an independent pilot rudder pulse is input during the maneuver, the rudder is largely responding to the lateral acceleration feedback. When the rudder is moving, several other variables are also moving, thus making it difficult to use the simplified approach just discussed. However, $C_{n_{\delta_r}}$ can be roughly determined when the rudder moves, approximately 1.7 sec from the start of the maneuver. Most of the slope of yaw rate is caused by the rudder, but a poor estimate would be obtained using the hand calculation.

Cost Function for Full Aircraft Problem

The analysis of a lateral-directional maneuver obtained in flight typically has from 15 to 25 unknown parameters (as shown in Eqs. (15) and (31)), in contrast to the one or two in the simple aircraft example. This makes detailed examples unwieldy and any graphic presentation of the cost function impossible. Therefore, in this section we are primarily examining the estimation procedure and the process of the minimization.

For our flight example, we have chosen a lateral-directional maneuver, with both aileron and rudder inputs, that has 17 unknown parameters. The data are from the oblique wing aircraft[18] with the wing unskewed during the maneuver. This example was chosen because it is a typical maneuver. The time history of the data and the subsequent output of MMLE3 have been published in Ref. 20. The tabular results of the analysis are shown in Table 7. The match between the measured time history (solid lines) and the estimated (calculated) time history (dashed lines) is shown as a function of iteration in Fig. 27. Figures 27(a) to (e) are for iterations 0 to 4, respectively. Table 7 shows that the cost remains unchanged after four iterations. A similar result was obtained for the two-dimensional simple aircraft example in Fig. 6 and Table 4.

Of the many things the analyst must consider in obtaining estimates, the two most important ones are how good is the match and how good is the convergence. A satisfactory match and monotonic convergence are desirable, but not sufficient, conditions for a successful analysis. Figure 27(e), although not perfect, is a very good match. The convergence can best be evaluated by looking at the normalized cost in the last row of Table 7. The cost has rapidly and monotonically converged in four iterations, and it remains at the converged cost. These factors are convincing evidence that the convergence is complete. Therefore, the criteria of match and convergence are satisfied in our example. In some cases we might encounter cost that does not converge rapidly (in four to six iterations) or monotonically, or stay "exactly" at the minimum value. These situations usually indicate at least a small problem in the analysis. These problems, if found, are usually traced to a data problem, an inadequate mathematical model, or a maneuver that contains a marginal amount of information.

Table 7 also shows that the startup values of all the coefficients are zero for the control and bias variables. Wind tunnel estimates could have been used for starting values, but the convergence of the algorithm is not very dependent on the startup values. As part of the startup algorithm, the MMLE3 program normally holds the derivatives of the state variables constant until after the first iteration, as is evident in Table 7.

Figure 27(a) shows the match between the measured and computed data for the startup values. The match is very poor because the startup values for the control derivatives are all zero, so the only motion is in response to the initial conditions. The control derivatives and biases are determined on the first iteration, resulting in the much improved match shown in Fig. 27(b). The match after two iterations, shown in Fig. 27(c), is improved as the program further modifies the control derivatives and, for the first time, adjusts the derivatives affecting the natural frequency (C_{n_β} and C_{ℓ_β}). By the third iteration (Fig. 27(d)), the improvement in the match is almost complete, because minor adjustments to the frequency are made and the damping derivatives are changed. Fig. 27(e) shows the match when all but the most minor derivatives have ceased to change.

Several general observations can be made based on this well behaved example. The strong or most important coefficients have essentially converged in three iterations. The same effect was seen in the simple example — that is, L_δ converged faster than L_p (Table 4). Some of the less important or second-order coefficients have only converged to two places after three iterations and are still changing by one digit in the fourth place at the end of six iterations. Another observation is that for some coefficients (C_{ℓ_r}, $C_{n_{\delta_a}}$, and $C_{\ell_{\delta_r}}$) even though the sign is wrong after the first iteration, the algorithm quickly selects their correct values once the important derivatives have stabilized.

In general, if the analysis of a maneuver has gone well, we do not need to spend much time inspecting a table analogous to Table 7. However, if there have been problems in convergence or in the quality of the fit, a detailed inspection of such a table may be necessary. The data may show an important coefficient going unstable at an early iteration, which could cause problems later. If the starting values are grossly in error, the algorithm is driven a long way from reasonable values and then for many reasons does not behave well. Occasionally the algorithm alternately selects from two diverse sets of values of two or more coefficients on successive iterations, behaving as if the shape of the cost function were a narrow multidimensional valley analogous to but more extreme than the two-dimensional valley shown in Figs. 18 and 20.

Cramèr-Rao Bounds

The earlier sections regarding the computed example have shown that the Cramèr-Rao bound is a good indicator of the accuracy of an estimated parameter. The Cramèr-Rao bounds can be used in a similar, but somewhat more qualitative, fashion on flight data. The Cramèr-Rao bounds that are included in MMLE3 (as well as many other maximum likelihood estimation programs) have been useful in determining whether estimates are good or bad. The aircraft example discussed here has been reported previously (for example, in Refs. 1 and 16). However, this example of the use of the Cramèr-Rao bound in the assessment of flight-derived estimates is pertinent to the thrust of this paper. Figure 28 shows estimates of C_{n_p} as a function of angle of attack for the PA-30 twin-engine general aviation aircraft[21] at three flap settings. There is a significant amount of scatter, which makes the reliability of the information on C_{n_p} questionable. The data shown are the estimates from the MMLE3 program, which also provides the Cramèr-Rao bounds for each estimate. Past experience[1] has shown that if the Cramèr-Rao bound is multiplied by a scale factor (the result sometimes called the uncertainty level[1,16]) and plotted as a vertical bar with the associated estimate, it helps in the interpretation of flight-determined results. Figure 29 shows the same data as Fig. 28, with the uncertainty levels now included as vertical bars. The estimates with small uncertainty levels (Cramèr-Rao bounds) are the best estimates, as was discussed earlier in the section on Cramèr-Rao bounds for the one-dimensional case. The fairing shown in Fig. 29 goes through the estimates with small Cramèr-Rao bounds and ignores the estimates with large bounds. One can have great confidence in the fairing of the estimates, because the fairing is well defined and consistent when the Cramèr-Rao bound information is included. In this particular instance, the estimates with small bounds were from maneuvers where the aileron forced the motion, and the large bounds were from maneuvers where the rudder forced the motion. Therefore, in addition to aiding in the fairing of the estimates, the Cramèr-Rao bounds help show that the aileron-forced maneuvers are superior for estimating C_{n_p} for the PA-30 aircraft.

This example illustrates that the Cramèr-Rao bounds are a useful tool in assessing flight-determined estimates, just as they were found useful for the simple aircraft example with computed data.

Concluding Remarks

The computed simple aircraft example showed the basics of minimization and the general concepts of cost functions themselves. In addition, the example showed the advantage of low measurement noise, multiple estimates at a given condition, and the Cramèr-Rao bounds, and the quality of the match between the measured and computed data. The flight data showed that many of these concepts still hold true even though the dimensionality of the cost function makes it impossible to plot or visualize.

References

[1]Iliff, K. W., "Aircraft Identification Experience," AGARD-LS-104, 1979, pp. 6-1 to 6-35.

[2]Goodwin, Graham C., and Payne, Robert L., Dynamic System Identification: Experiment Design and Data Analysis, Academic Press, New York, 1977.

[3]Sorenson, Harold W., Parameter Estimation; P· nciples and Problems, Marcel Dekker, Inc., I. York, 1980.

[4]Klein, V., "On the Adequate Model for Aircraft Parameter Estimation," CIT, Cranfield Report Aero No. 28, Mar. 1975.

[5]Hamel, P., "Determination of Stability and Control Parameters From Flight Testing," AGARD-LS-114, 1981, pp. 10-1 to 10-42.

[6]Iliff, Kenneth W., and Maine, ichard E., "NASA Dryden's Experience in Parameter Estimation and Its Use in Flight Test," AIAA Paper 82-1373, Atmospheric Flight Mechanics Conference, San Diego, Calif., Aug. 1982.

[7]Gupta, N. K., Hall, Earl W., and Trankle, T. L., "Advanced Methods of Model Structure Determination from Test Data," AIAA Paper 77-1170, 1977.

[8]Trankle, T. L., Vincent, J. H., and Franklin, S. N., "System Identification of Nonlinear Aerodynamic Models," The Techniques and Technology of Nonlinear Filtering and Kalman Filtering, AGARD-AG-256, 1982.

[9]Klein, V., and Schiess, J. R., "Compatibility Check of Measured Aircraft Responses Using Kinematic Equations and Extended Kalman Filter," NASA TN D-8514, 1977.

[10]Bach, R. E., and Wingrove, R. C., "Applications of State Estimation in Aircraft Flight Data Analysis," AIAA Paper 83-2087, 1983.

[11]Mulder, J. A., Jonkers, H. L., Horsten, J. J., Breeman, J. H., and Simons, J. H., "Analysis of Aircraft Performance, Stability and Control Measurements," AGARD-LS-104, 1979, pp. 5-1 to 5-87.

[12]Klein, Vladislav, "Maximum Likelihood Method for Estimating Airplane Stability and Control Parameters From Flight Data in Frequency Domain," NASA TP-1637, 1980.

[13]Fu, K. H., and Marchand, M., "Helicopter System Identification in the Frequency Domain," Ninth European Rotorcraft Forum, Paper 96, Stresa, Italy, Sept. 13-15, 1983.

[14]Maine, Richard E., and Iliff, Kenneth W., "Formulation and Implementation of a Practical

Algorithm for Parameter Estimation with Process and Measurement Noise," SIAM J. Appl. Math., vol. 41, 1981, pp. 558-579.

[15]Balakrishnan, A. V., Communication Theory, McGraw-Hill Book Co., c.1968.

[16]Maine, Richard E., and Iliff, Kenneth W., "The Theory and Practice of Estimating the Accuracy of Dynamic Flight-Determined Coefficients," NASA RP-1077, 1981.

[17]Maine, Richard E., and Iliff, Kenneth W., "User's Manual for MMLE3, A General FORTRAN Program for Maximum Likelihood Parameter Estimation," NASA TP-1563, 1980.

[18]Maine, Richard E., "Aerodynamic Derivatives for an Oblique Wing Aircraft Estimated From Flight Data by Using a Maximum Likelihood Technique," NASA TP-1336, 1978.

[19]Shafer, M. F., "Flight-Determined Correction Terms for Angle of Attack and Sideslip," AIAA Paper 82-1374, 1982.

[20]Maine, Richard E., "Programmer's Manual for MMLE3, A General FORTRAN Program for Maximum Likelihood Parameter Estimation," NASA TP-1690, 1981.

[21]Fink, Marvin P., and Freeman, Delma C., Jr., "Full-Scale Wind-Tunnel Investigation of Static Longitudinal and Lateral Characteristics of a Light Twin-Engine Airplane," NASA TN D-4983, 1969.

Table 1 Values of computed time history with no measurement noise

i	δ, deg	p, deg/sec
1	0	0
2	1	0.9754115099857
3	1	2.878663149266
4	1	4.689092110779
5	1	6.411225409939
6	1	8.049369277012
7	1	9.607619924937
8	0	10.11446228200
9	0	9.621174135646
10	0	9.151943936071

Table 2 Pertinent values as a function of iteration

L	$\hat{L}_p(L)$	$\hat{L}_\delta(L)$	$\phi(L)$	$\psi(L)$	J_L
0	-0.5000	15.00	0.9048	2.855	21.21
1	-0.3005	9.888	0.9417	1.919	0.5191
2	-0.2475	9.996	0.9517	1.951	5.083×10^{-4}
3	-0.2500	10.00	0.9512	1.951	1.540×10^{-9}
4	-0.2500	10.00	0.9512	1.951	1.060×10^{-14}

Table 3 Values of computed time history with added measurement noise

i	δ, deg	p, deg/sec
1	0	0
2	1	0.4875521781881
3	1	3.238763570696
4	1	3.429117357944
5	1	6.286297353361
6	1	6.953798550097
7	1	10.80572930119
8	0	9.739367269447
9	0	9.788844525490
10	0	7.382568353168

Table 4 Pertinent values as a function of iteration

L	$\hat{L}_p(L)$	$\hat{L}_\delta(L)$	$\phi(L)$	$\psi(L)$	J_L
0	-0.5000	15.00	0.9048	2.855	30.22
1	-0.3842	10.16	0.9260	1.956	3.497
2	-0.3518	10.23	0.9321	1.976	3.316
3	-0.3543	10.25	0.9316	1.978	3.316
4	-0.3542	10.24	0.9316	1.978	3.316
5	-0.3542	10.24	0.9316	1.978	3.316

Table 5 Mean and standard deviations for estimates of L_p

Number of cases, N	Sample mean, $\mu(\hat{L}_p)$	Sample standard deviation, $\sigma(\hat{L}_p)$	Sample standard deviation of the mean, $\sigma(\hat{L}p)/\sqrt{N}$
5	-0.2668	0.0739	0.0336
10	-0.2517	0.0620	0.0196
20	-0.2452	0.0578	0.0129

Table 6 Estimate of L_p and Cramèr-Rao bound as a function of the square root of noise power

Square root of noise power, G	Estimate of $\hat{L}_p$	Cramèr-Rao bound
0.0	-0.2500	-------
0.01	-0.2507	0.00054
0.05	-0.2535	0.00271
0.10	-0.2570	0.00543
0.2	-0.2641	0.0109
0.4	-0.2783	0.0220
0.8	-0.3071	0.0457
1.0	-0.3218	0.0579
2.0	-0.3975	0.1248
5.0	-0.6519	0.3980
10.0	-1.195	1.279

Table 7 Stability and control derivatives as a function of iteration for flight maneuver

Iteration	Iteration						
	L	1	2	3	4	5	6
C_{Y_β}	-0.008500	-0.008500	-0.007959	-0.008347	-0.008375	-0.008364	-0.008364
C_{ℓ_β}	-0.0002500	-0.0002500	-0.0003141	-0.0003580	-0.0003572	-0.0003571	-0.0003571
C_{n_β}	0.001000	0.001000	0.001159	0.001243	0.001230	0.001230	0.001230
C_{ℓ_p}	-0.2500	-0.2500	-0.3393	-0.3584	-0.3581	-0.3581	-0.3581
C_{n_p}	-0.02500	-0.02500	-0.04356	-0.04537	-0.04512	-0.04599	-0.04600
C_{ℓ_r}	-0.05000	-0.05000	0.06790	0.07044	0.06972	0.06973	0.06974
C_{n_r}	-0.00800	-0.08000	-0.1327	-0.1033	-0.1065	-0.1062	-0.1062
$C_{\ell_{\delta_a}}$	0	0.0008009	0.001000	0.001067	0.001069	0.001069	0.001069
$C_{n_{\delta_a}}$	0	-0.00004604	6.786×10^{-7}	0.00001129	0.00001096	0.00001068	0.00001069
$C_{Y_{\delta_r}}$	0	0.005935	0.002064	0.001456	0.001546	0.001548	0.001548
$C_{\ell_{\delta_r}}$	0	-0.00005068	0.00005764	0.0001043	0.0001059	0.0001055	0.0001055
$C_{n_{\delta_r}}$	0	-0.0007329	-0.0009333	-0.0008875	-0.0008972	-0.0008961	-0.0008961
C_{Y_0}	0	-0.05109	-0.02691	-0.02362	-0.02420	-0.02419	-0.02420
$C_{Y_0} + \dot{\beta}_0$	0	-0.03370	-0.01370	-0.01117	-0.0115	-0.01156	-0.01156
C_{ℓ_0}	0	-0.0007096	-0.001629	-0.002021	-0.002031	-0.002028	-0.002028
C_{n_0}	0	0.005864	0.007300	0.007140	0.007175	0.007169	0.007169
$\dot{\phi}_0$	0	0.2121	0.1626	0.1482	0.1506	0.1506	0.1506
J/(N - 1)	731.5	65.00	11.23	4.8265	4.701	4.701	4.701

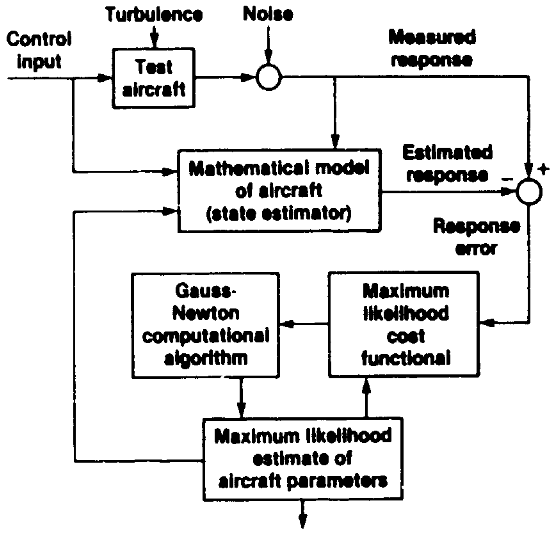


Fig. 1 Maximum likelihood estimation concept.

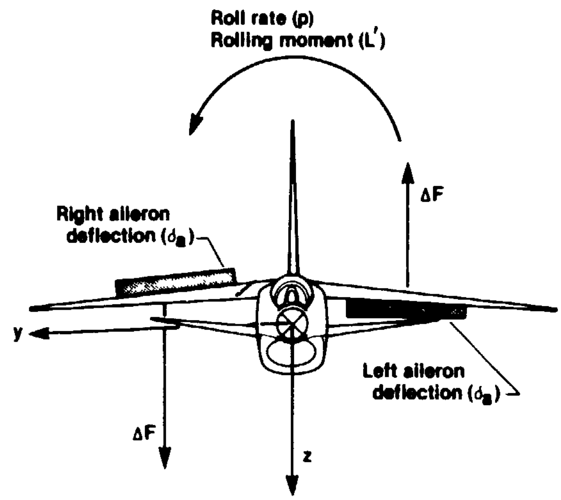


Fig. 2 Simplified aircraft nomenclature.

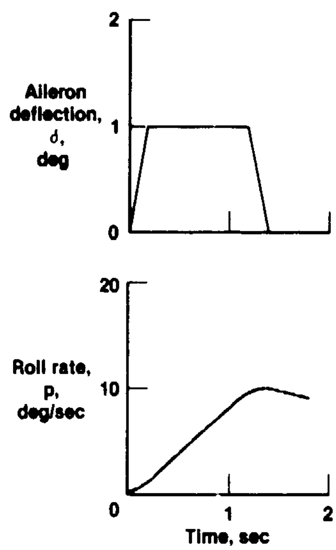


Fig. 3 Time history with no measurement noise.

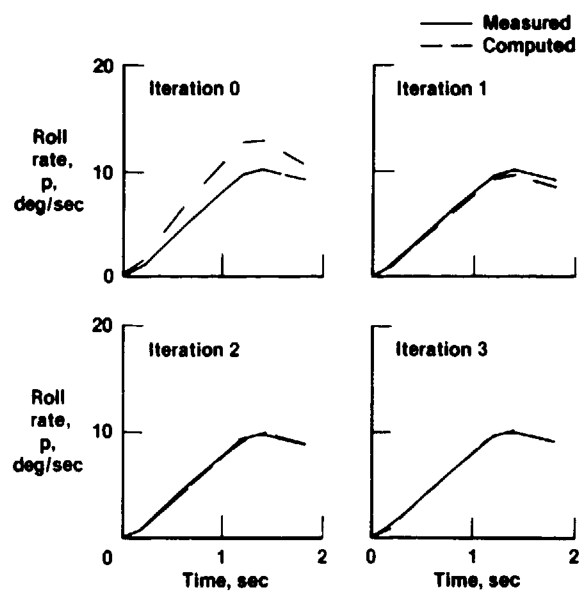


Fig. 4 Comparison of measured and computed data for each of the first three iterations.

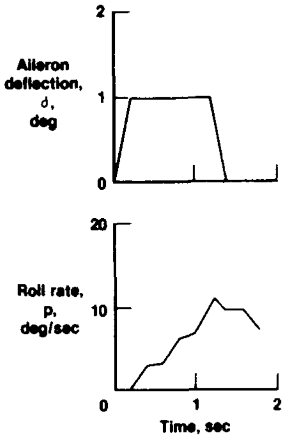


Fig. 5 Time history with measurement noise.

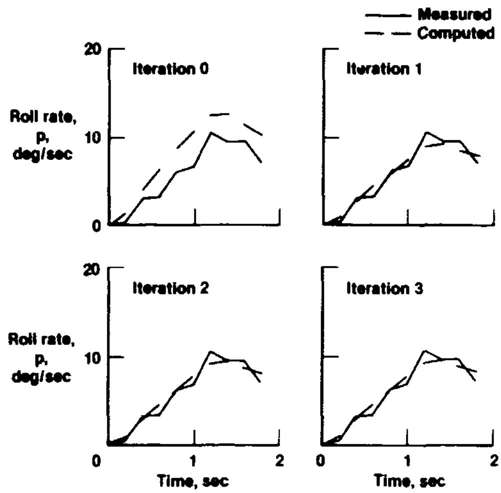


Fig. 6 Comparison of measured and computed data for each iteration.

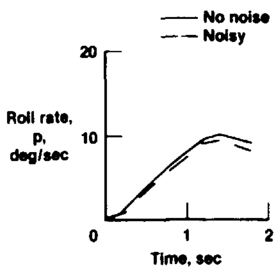


Fig. 7 Comparison of estimated roll rate from no-noise and noisy cases.

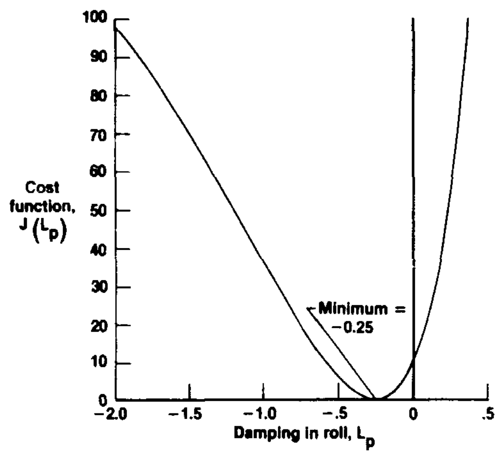


Fig. 8 Cost function ($J(L_p)$) as a function of L_p for no-noise case.

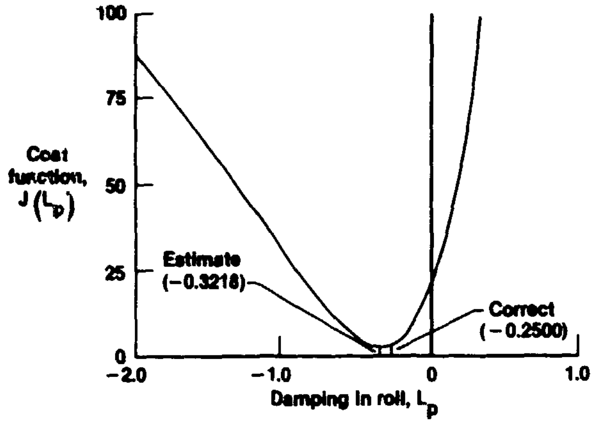


Fig. 9 Cost function as a function of L_p for noisy case.

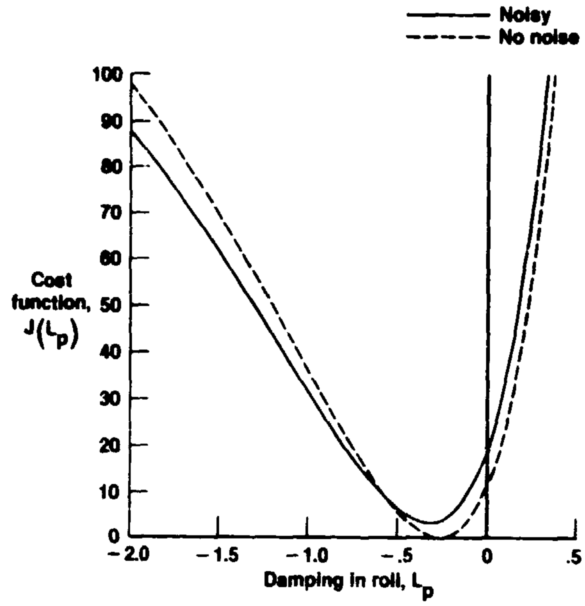


Fig. 10 Comparison of the cost functions for the no-noise and noisy cases.

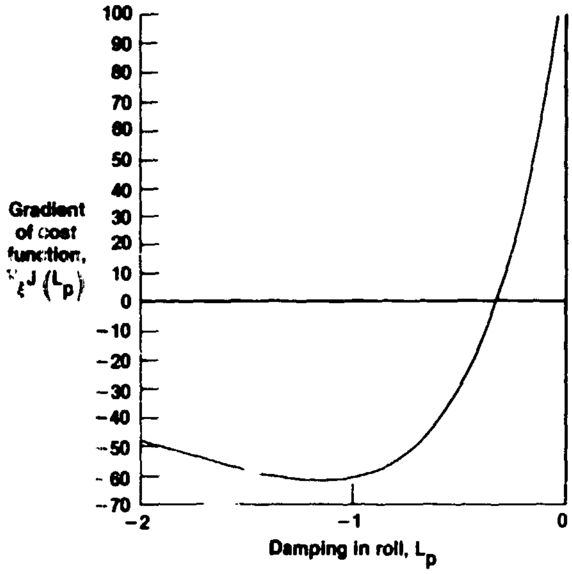


Fig. 11 Gradient of $J(L_p)$ as a function of L_p for noisy case.

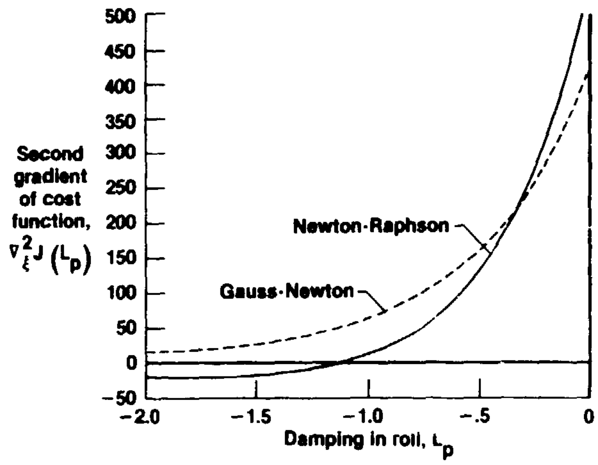


Fig. 12 Comparison of Newton-Raphson and Gauss-Newton values of the second gradient for the noisy case.

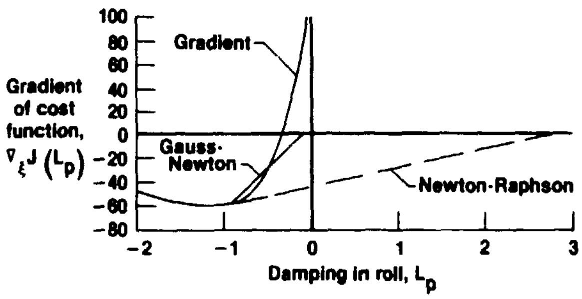


Fig. 13 Comparison of first iteration step size for the Newton-Raphson and Gauss-Newton algorithms for the noisy case.

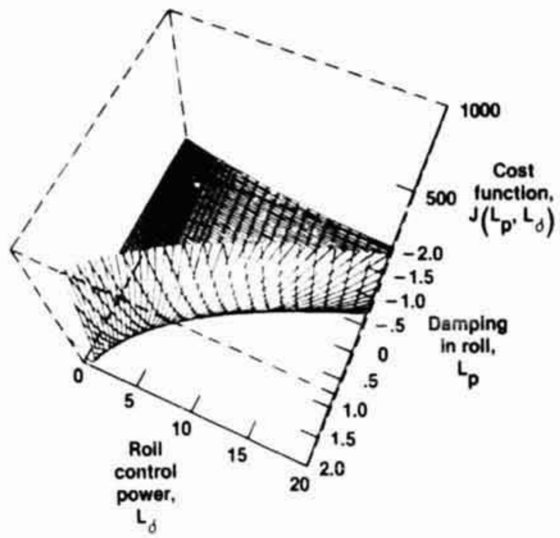


Fig. 14 Large-scale view of cost function surface.

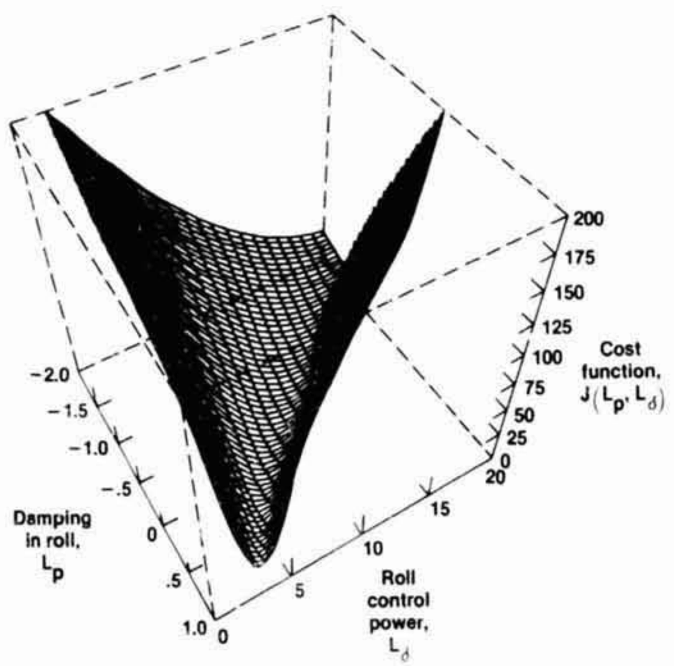


Fig. 15 Restricted view of cost function surface.

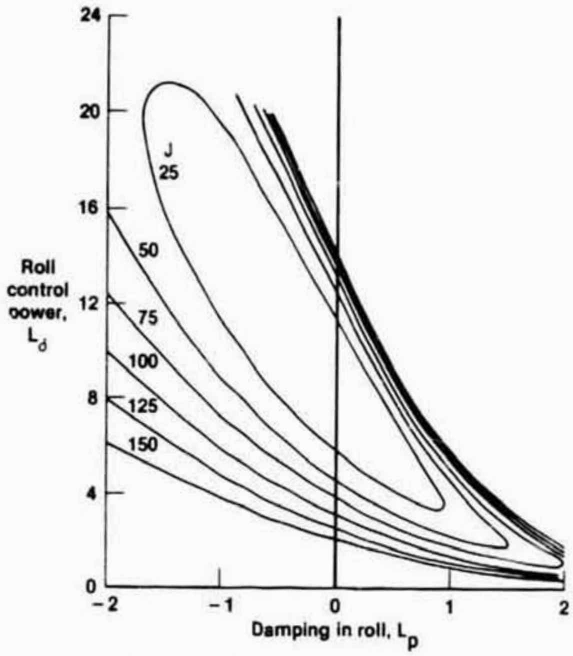


Fig. 16 Isoclines of constant cost of L_p and L_δ for the no-noise case.

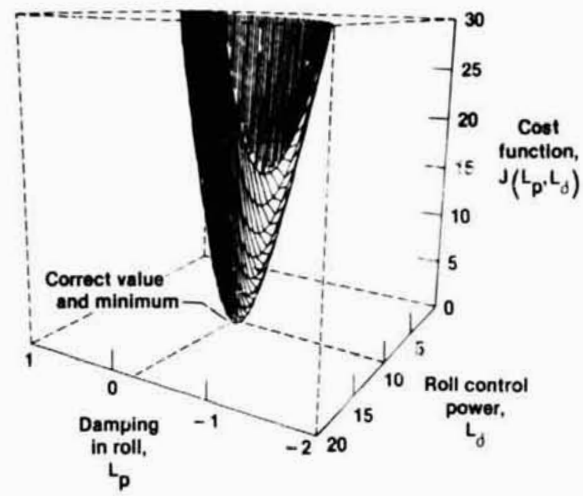


Fig. 17 Detailed view of cost function surface for no-noise case.

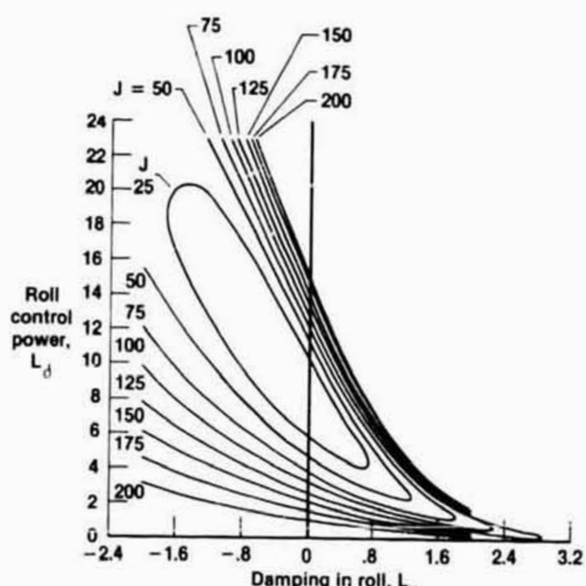


Fig. 18 Isoclines of constant cost in L_p and L_δ for the noisy case.

ORIGINAL PAGE IS
OF POOR QUALITY

ORIGINAL PAGE IS OF POOR QUALITY

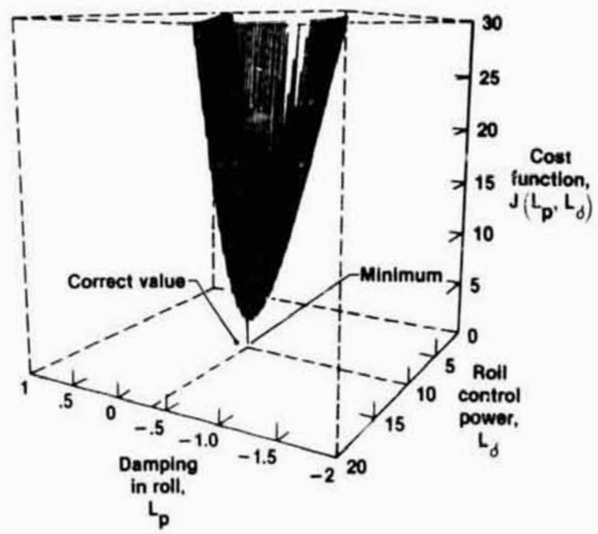


Fig. 19 Detailed view of cost function surface for noisy case.

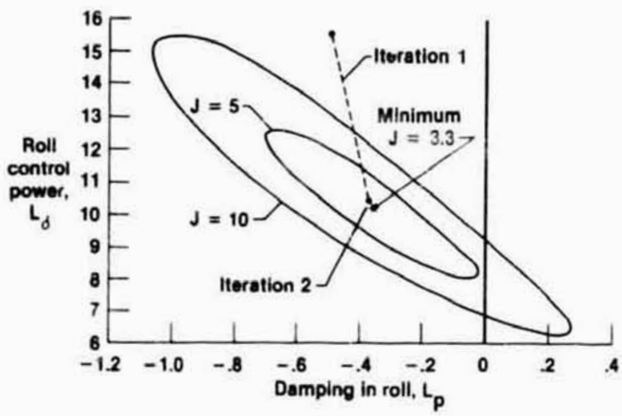


Fig. 20 Isoclines of constant cost for region near minimum for noisy case.

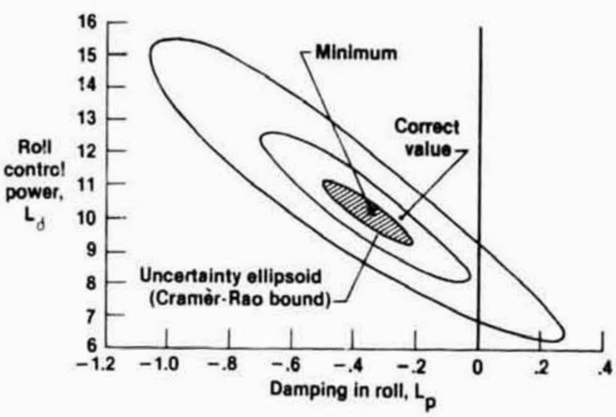


Fig. 21 Isoclines and uncertainty ellipsoid of the cost function for the noisy case.

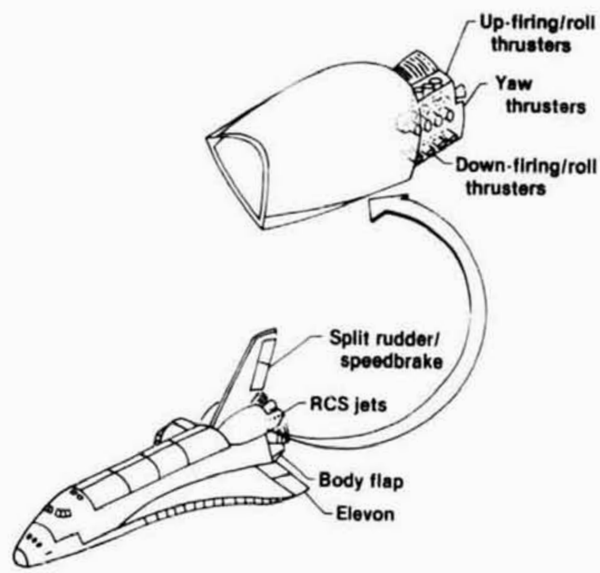


Fig. 22 Space shuttle configuration.

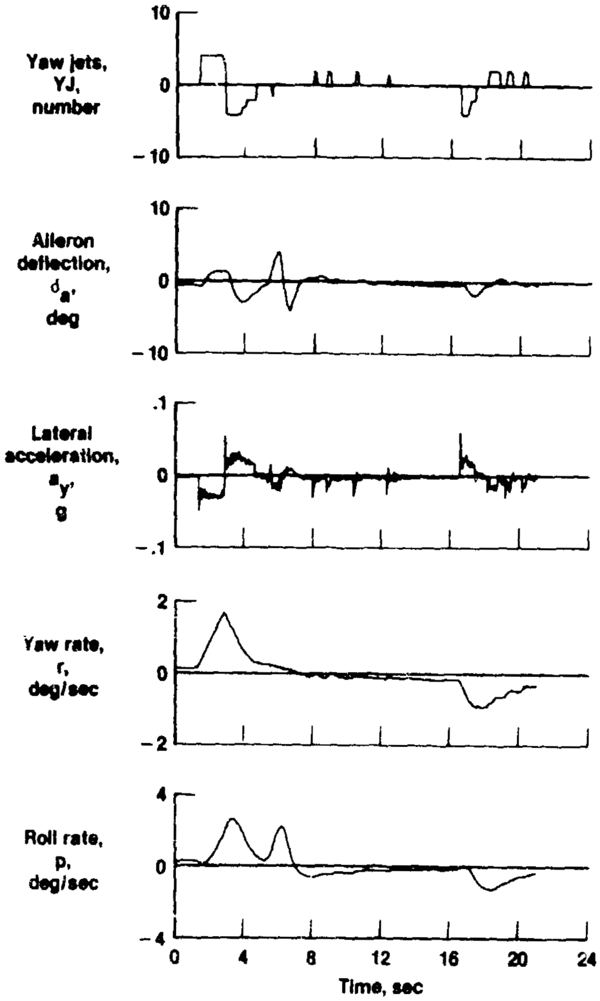


Fig. 23 Lateral-directional space shuttle maneuver at a Mach number of 21.

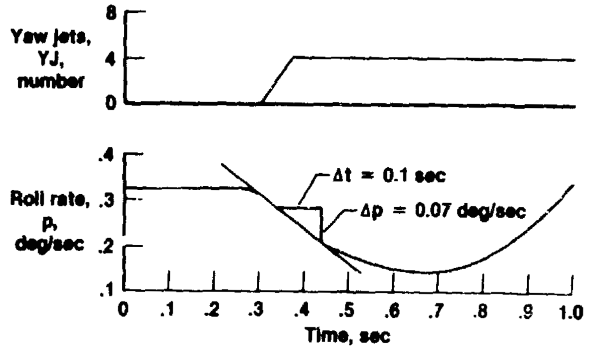


Fig. 24 Examples of obtaining L_{YJ} by simple calculations for the shuttle data from Fig. 23.

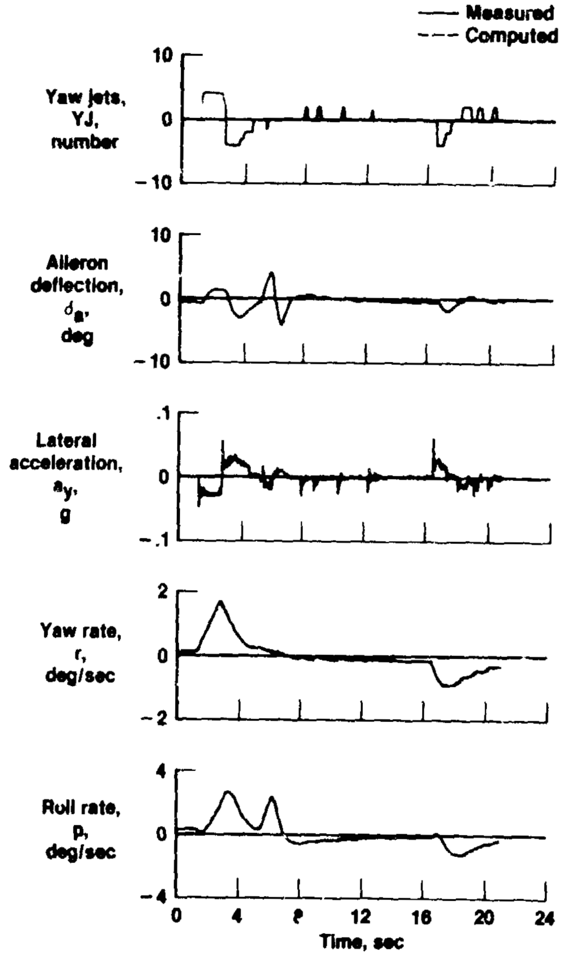


Fig. 25 MMLE3 match of maneuver shown in Fig. 23.

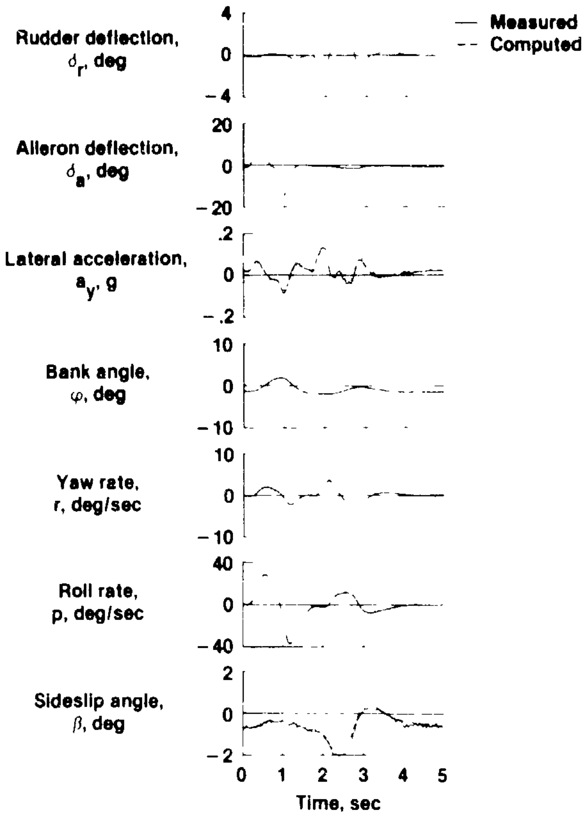
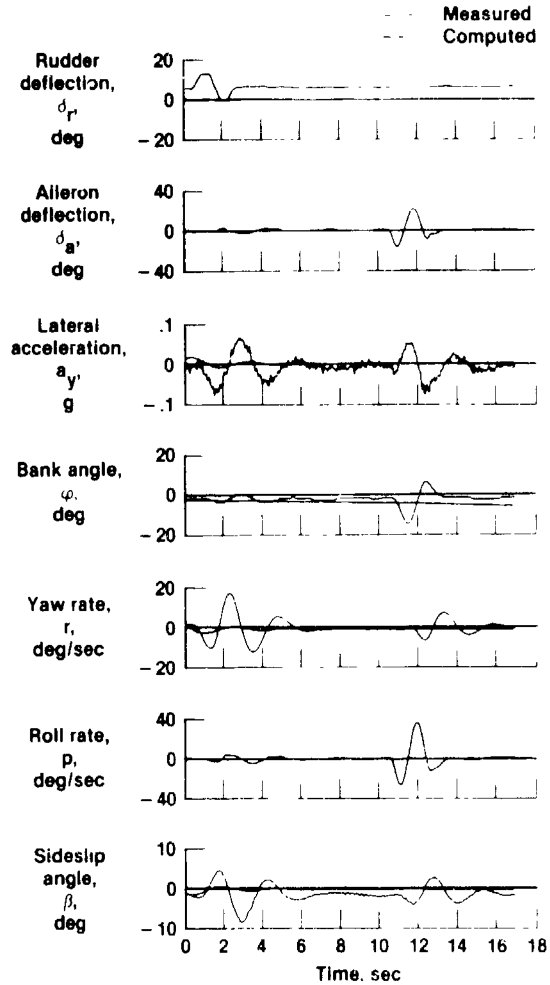
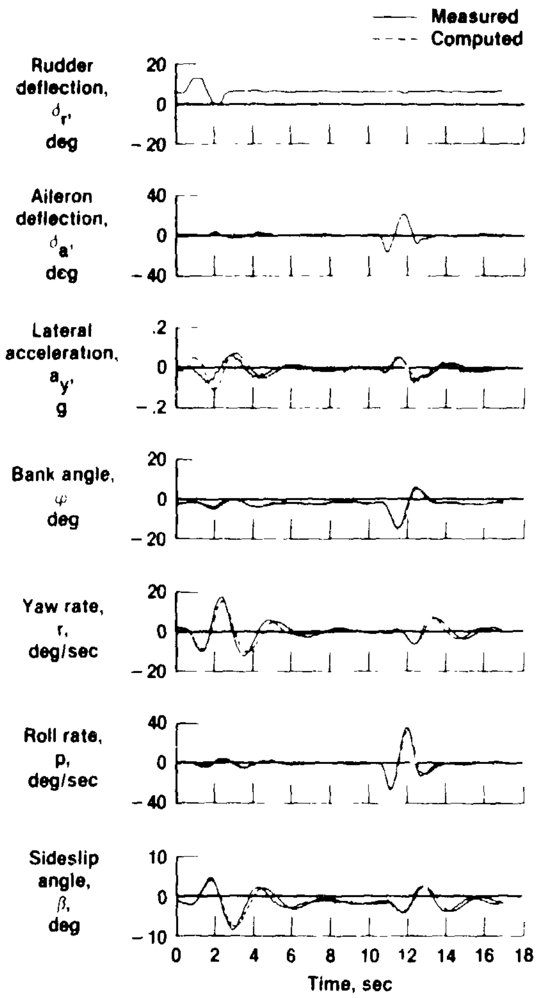


Fig. 26 Lateral-directional maneuver from F-8 aircraft with augmentation on.

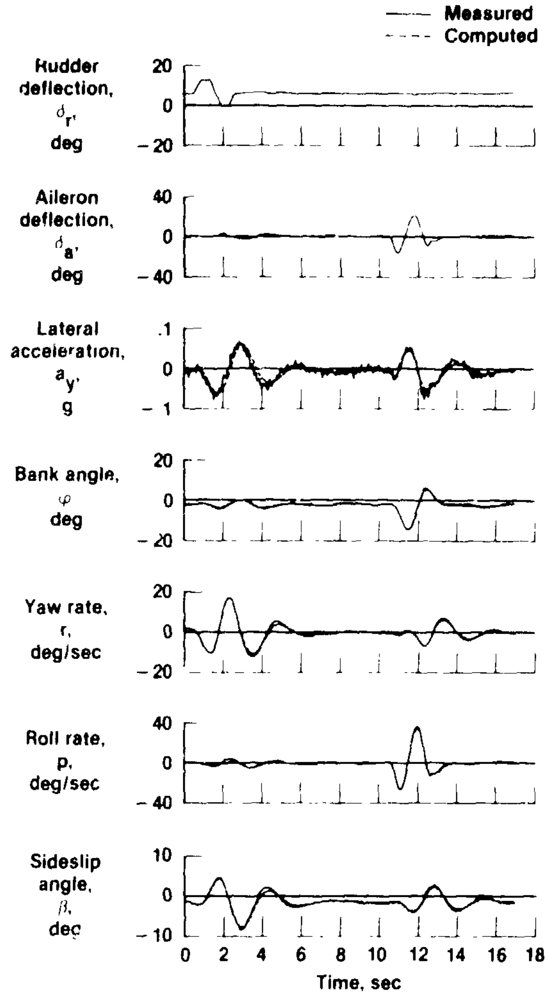


(a) Zero iteration.

Fig. 27 Match between measured and computed time histories as a function of iteration.

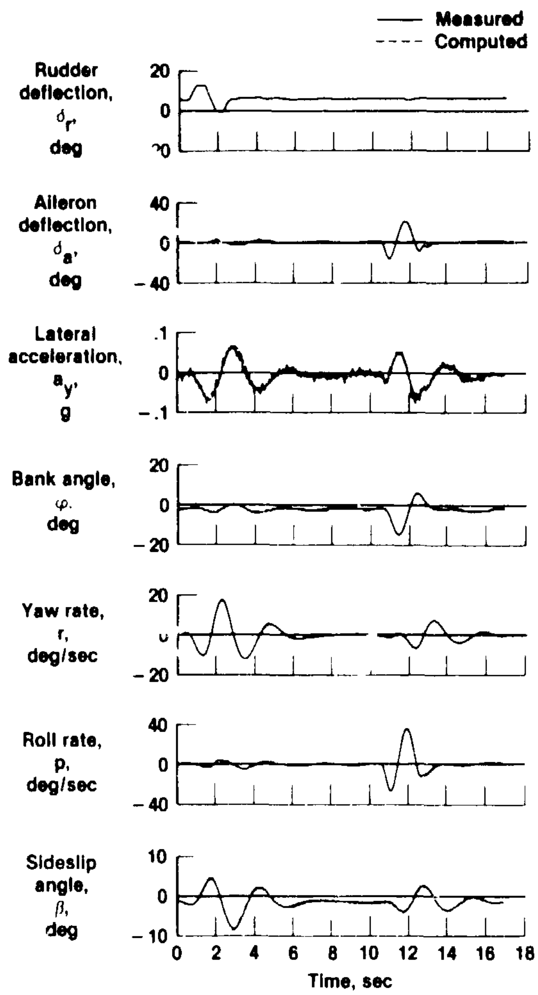


(b) One iteration.

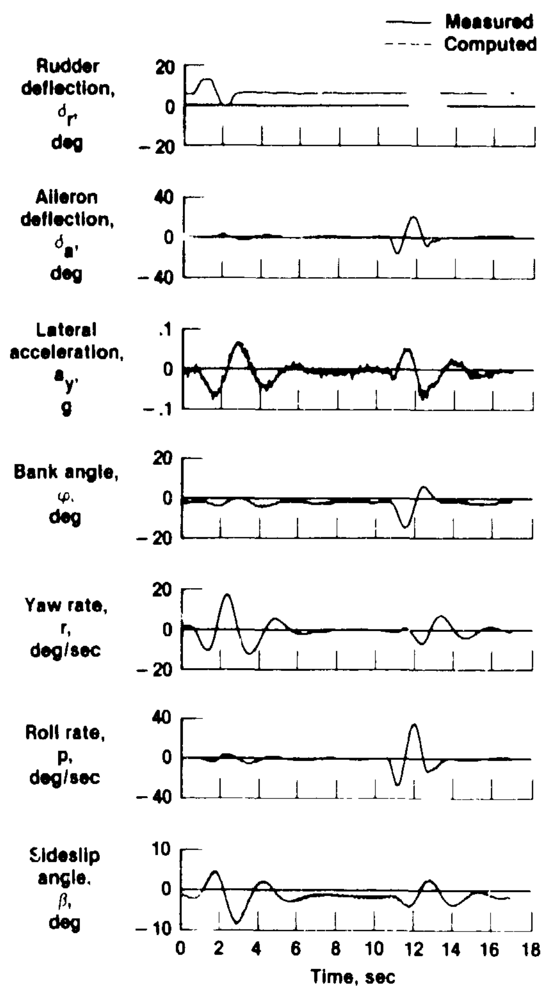


(c) Two iterations.

Fig. 27 Continued.



(d) Three iterations.



(e) Four iterations.

Fig. 27 Concluded.