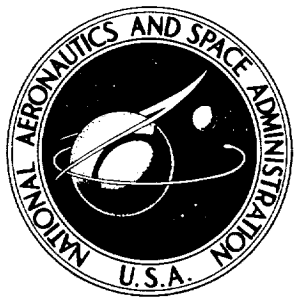
NASA CONTRACTOR REPORT



NASA CR-2200

NASA CR-2200

MAXIMUM LIKELIHOOD IDENTIFICATION AND OPTIMAL INPUT DESIGN FOR IDENTIFYING AIRCRAFT STABILITY AND CONTROL DERIVATIVES

by David E. Stepner and Raman K. Mehra

Prepared by
SYSTEMS CONTROL, INC
Palo Alto, Calif. 94306
for Langley Research Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • MARCH 1973

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CR-2200		

4. Title and Subtitle	5. Report Date
MAXIMUM LIKELIHOOD IDENTIFICATION AND OPTIMAL INPUT DESIGN FOR IDENTIFYING AIRCRAFT STABILITY AND CONTROL DERIVATIVES	March 1973
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
David E. Stepner and Raman K. Mehra	
9. Performing Organization Name and Address	10. Work Unit No.
Systems Control, Inc. 260 Sheridan Avenue Palo Alto, California 94306	11. Contract or Grant No. NAS1-10700
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
National Aeronautics and Space Administration Washington, D.C. 20546	Contractor Report
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

A new method of extracting aircraft stability and control derivatives from flight test data is developed based on the maximum likelihood criterion. It is shown that this new method is capable of processing data from both linear and nonlinear models, both with and without process noise and includes output error and equation error methods as special cases. The first application of this method to flight test data is reported for lateral maneuvers of the HL-10 and M2/F3 lifting bodies, including the extraction of stability and control derivatives in the presence of wind gusts. All the problems encountered in this identification study are discussed. Several different methods (including a priori weighting, parameter fixing and constrained parameter values) for dealing with identifiability and uniqueness problems are introduced and the results given.

The method for the design of optimal inputs for identifying the parameters of linear dynamic systems is also given. The criterion used for the optimization is the sensitivity of the system output to the unknown parameters. Several simple examples are first given and then the results of an extensive stability and control derivative identification simulation for a C-8 aircraft are detailed.

17. Key Words (Suggested by Author(s))	18. Distribution Statement
Parameter identification Maximum likelihood Aerodynamic derivatives Aircraft modeling Control design for parameter extraction	Unclassified - Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	205	$3.00

*For sale by the National Technical Information Service, Springfield, Virginia 22151

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

FIGURES

FIGURES

FIGURES

TABLES

MAXIMUM LIKELIHOOD IDENTIFICATION AND OPTIMAL INPUT DESIGN FOR IDENTIFYING AIRCRAFT STABILITY AND CONTROL DERIVATIVES

By David E. Stepner and Raman K. Mehra

Systems Control, Inc.

I

INTRODUCTION

Aircraft parameter identification is the process of extracting numerical values for the aerodynamic stability and control derivatives, and other subsidiary parameters (wind gusts, sensors errors, etc.), from a set of flight test data (a time history of the flight control inputs and the resulting aircraft response variables). The field of identification is one that has been pursued by diverse interests for many years. The practical application of this work to aircraft flight testing has existed for over 25 years. In spite of the wealth of experience which has been accumulated in this span of time, important requirements still exist for improving the techniques for extracting stability and control derivatives.

First, there exists today a greater need for stability and control derivatives. There are currently two principal requirements for the mathematical models that these coefficients provide. These are (1) to provide inputs to simulators*, and (2) to provide a basis for the design of flight control systems. A third potential may also exist. Because the stability and control derivatives define a given aircraft more uniquely than the response mode criteria such as those in the

---

*This may include digital computer simulations, fixed and moving base ground simulators, and in-flight simulators such as variable stability aircraft.

Flying Qualities Military Specification MIL-F-8785 there is reason to believe that these parameters will ultimately play more of a major role in the design, testing, and certification of aircraft.

Second, with the continuing advances in aircraft design and performance capabilities, the ability to extrapolate wind tunnel test results is diminishing and the importance of flight testing is growing. This is aided by the new Department of Defense policy of building prototype aircraft and thoroughly flight testing them before a production commitment is made.

The principal elements of the aircraft identification process (see Fig. 1.1) are: (1) the identification algorithm, (2) the flight control input and (3) the instrumentation. The ultimate success of the identification process is totally dependent on all three of these elements, not just one of them alone. This study was concerned with the first two points, namely, the development of a general advanced digital identification technique based on the maximum likelihood criterion and the design of control inputs which will enhance the ability to identify specific aircraft stability and control derivatives. Digital parameter identification techniques have already reached a stage where they are being used increasingly over analog matching techniques for extracting stability and control derivatives from flight test data. Systems Control, Inc., (SCI) under this present contract developed the maximum likelihood identification technique, which was used successfully to reduce data from flight tests where gusts were present. In such cases both the measurement noise and process noise statistics were identified.

The importance of the control input in the identifiability of stability and control derivatives from response data has been apparent for a long time. Under this contract, SCI has developed and applied an efficient computational technique to design the optimal inputs for identifying specific stability and control derivatives.

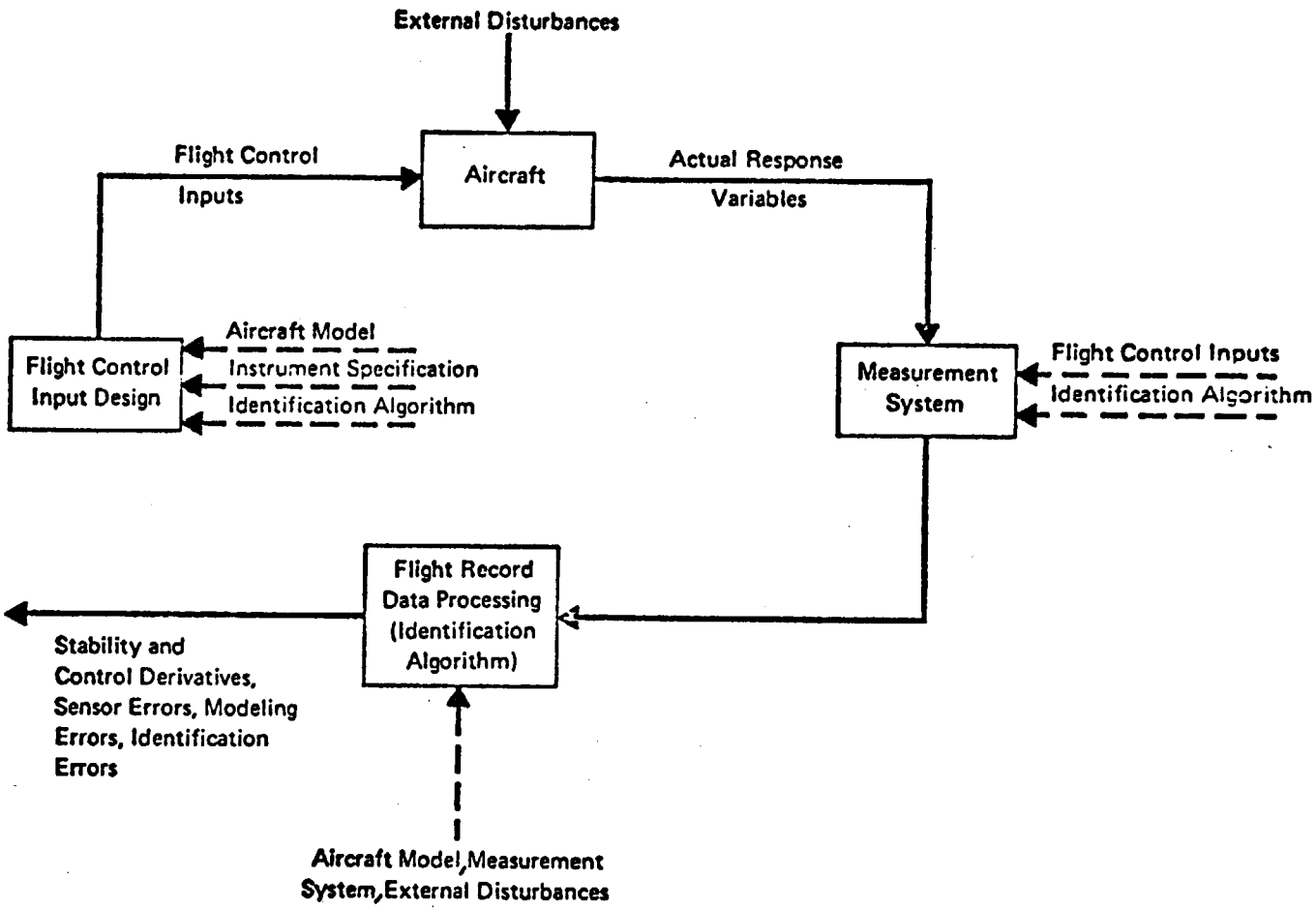


FIGURE 1.1 THE INTEGRATED AIRCRAFT IDENTIFICATION PROCESS

This report is organized as follows:

- Section II includes the specific task objectives of this contract and a summary of the principal results.
- Section III discusses the background material for the identification of aircraft stability and control derivatives.
- Section IV describes, in detail, the SCI Maximum Likelihood Identification Method. The derivation is carried out for both linear and non-linear models with and without process noise. The relationship of the technique to the output error and equation error methods is described and the related identifiability problems are discussed. Included also is a detailed description of the SCI Maximum Likelihood Identification Program.
- Section V presents the results on the identification of the stability and control derivatives for several different aircraft and under a variety of noise conditions. This includes simulated data for a nonlinear model of an X-22 VTOL, actual flight data from an HL-10 lifting body (linear model) and flight data containing gust effects for an M2/F3 lifting body (linear model).
- Section VI covers the requirements and the background material relating to the design of inputs for aircraft parameter identification.
- Section VII describes the details of the theoretical development and the computational technique of computing optimal inputs for identifying aircraft stability and control derivatives. Several examples for which analytical solutions for the optimal input exist are presented to illustrate the form which the optimal inputs take. Included also is a detailed flow diagram of the SCI Optimal Input Design Program.

- Section VIII presents numerical results showing the characteristics of optimal inputs, and comparing the performance of the optimal input with a doublet input of equal energy and duration. The results of a Monte Carlo simulation of the identification process for the short period longitudinal derivatives of a C-8 aircraft are presented indicating a substantial advantage in using the optimal input.

- Section IX states the conclusions based on the results of this study.

- Section X contains recommendations for further work.

II
OBJECTIVES AND SUMMARY OF RESULTS

This section presents (1) a statement of the study objectives, (2) an outline of the SCI Maximum Likelihood Identification Method and the SCI technique for designing optimal inputs for aircraft stability and control derivative identification and (3) a summary of the principal results of this study. The section is self-contained and is intended to provide the reader with an overview of the report.

2.1 Study Objectives

The objectives of this study were two-fold. First, it was desired to further develop the Maximum Likelihood Identification technique, originated by SCI in 1970, to the extent that it would successfully process actual flight test data containing random flight disturbances. This processing was to include identifying the correlation function of the random disturbance and determining a model for its representation. The second objective was to develop the theoretical foundation and construct a computer program for designing flight test inputs which would enhance the ability to identify specified stability and control derivatives.

To achieve the above objectives the following tasks were defined and completed:

Maximum Likelihood Identification:

(1) Investigate the effects of different sensitivity terms in the identification of stability and control derivatives of X-22 VTOL aircraft, from simulated data, for a nonlinear aerodynamical model

(2) Check out completed ML identification algorithm on HL-10 (Lifting Body) flight data, for a linear aerodynamical model, for which previous results had been obtained

(3) Identify stability and control derivatives of the M2/F3 (Lifting Body) from flight data containing random disturbances, for a linear model for which previous identification attempts with output error method have not succeeded.

(4) Investigate symptoms, causes and remedies of parameter identifiability and uniqueness problems.

Optimal Input Design:

(1) Construct operational computer program, based on results of theoretical study, for designing optimal inputs

(2) Perform a Monte Carlo simulation of the identification process, comparing the optimal input with a suboptimal input, for a model of the C-8 linear longitudinal equations of motion.

2.2 Maximum Likelihood Identification Technique

For the last 20 years various techniques such as fourier analysis, analog matching, and the time-vector method have been used to extract numerical values for the aerodynamic stability and control derivatives from records of flight test data. It has only been in recent years that modern digital computer techniques have been proposed for this problem. One of the most successful of these computer techniques is the Maximum Likelihood method developed by Systems Control, Inc. This technique holds great promise for future identification problems involving new aircraft configurations (VTOL,STOL), high angle of attack transonic flight regime, flight test data containing gusts and for aircraft with stability augmentation systems.

In the most general case, the maximum likelihood identification technique is a combination of three steps: (1) Kalman filtering to estimate the states and generate a residual or "innovation" sequence, (2) a modified Newton-Raphson algorithm for the parameter estimates and (3) an algorithm to estimate the noise statistics (mean and variances of the measurement and

process noise). In addition, the maximum likelihood technique provides a lower bound on the variances of the parameter estimate, and models for the measurement and process noise disturbances.

Under this contract the maximum likelihood identification technique has been applied to a variety of flight test data both simulated and real. The objective has been to exercise the technique as much as possible and to investigate the problems that arose. As each problem was solved, the specialized algorithm needed for its solution (if any) was added to the complete maximum likelihood identification program. The goal was to develop a set of general computer algorithms capable of dealing with problems that arise in the identification of aircraft stability and control derivatives.

2.2.1 X-22 VTOL Simulated Data

The first phase of the identification study was the processing of data from a simulation of the X-22 VTOL Aircraft. The longitudinal aerodynamic equations of motion were nonlinear and the data contained both measurement noise and process noise. Each of the stability and control derivatives was expressed as first or second order polynomial expansions in terms of the longitudinal velocity. The objective was to identify 23 of these expansion coefficients and the quantitative effect of increased noise power on the quality of the parameter estimates.

The problems encountered were almost all associated with either the aircraft model or the control input sequence. It was discovered quite soon in the investigation that a simple step input did not sufficiently excite the aircraft modes to allow for the accurate identification of many of the derivatives. The use of a multistep input improved these parameters a great deal , because of the model structure the input and output noise sequences were correlated. Accounting for this correlation improved the parameter estimates by increasing the estimate error covariances, thereby bringing many estimates to within one standard deviation

of the actual values. During earlier investigations, some of the sensitivity terms were not included in the identification algorithm. When these terms were added, however, the quality of the parameter estimates changed very little, although the computer time, per iteration, more than doubled.

A compilation of the results indicated that for "low" measurement and process noise, the maximum likelihood identification technique was able to identify all of the expansion coefficients, except those for Z_δ, accurately. When the expansions were recombined to form the time-varying stability and control derivatives, the fit to the derivatives M_o, M_q, M_w, M_δ, X_o, X_w, X_δ, Z_o was very good, the fit to Z_w was acceptable and only the fit to Z_δ would be considered unsatisfactory. When "moderate" process noise was used all the estimates of the expansion coefficients degraded.

2.2.2 HL-10 Flight Data

The second phase of the identification study involved using the maximum likelihood technique in the output error mode to identify the linearized lateral stability and control derivatives from flight test data for an HL-10 lifting body. Although the technique had no difficulty in accurately fitting the observed data (p, r, ϕ, β and n_{a_y}), several of the derivative estimates had opposite signs from the wind tunnel/theoretical derived values which were used as initial estimates. These incorrect signs could be attributed to any one of the following factors: (1) insufficient excitation of particular aircraft modes due to inadequate input or action of the SAS system, (2) the linearized dynamics not sufficiently accurate for the flight conditions of the data, or (3) correlated measurement noise due to instrumentation system dynamics.

In an effort to correct these signs, a modification to the maximum likelihood technique was attemped. This modification was to add to the

likelihood criterion a quadratic term putting a weighted cost on the difference between the a priori parameter estimates and the latest estimates. Using the weights supplied by NASA Edwards FRC, this "a priori weighting" method resulted in the correct signs and only a slightly (10%) degraded fit to the observed data, as long as the measurement biases were identified, in addition to the other parameters.

2.2.3 M2/F3 Flight Data

This third phase of the study involved extracting the linearized, lateral stability and control derivatives of an M2/F3 lifting body from flight test data containing gusts. Unlike the HL-10 data which had been successfully processed earlier by the use of the Output Error technique neither a satisfactory set of stability and control derivative estimates nor a satisfacory fit to the observed data had not been obtained from the M2/F3 data. Using an approximation that the gust noise in the sideslip angle measurement was much greater than the measurement noise and the maximum likelihood method with a Kalman filter model to account for the process noise, an accurate fit to the observed data was obtained. However, as in the HL-10 case, some of the estimated derivatives had signs opposite to those of the a priori estimates.

The a priori weighting method, which was used successfully on the HL-10 data, proved to be not useful on the M2/F3 data. Two other techniques, both dealing with identifiability problems, were investigated. The first technique involved fixing at their a priori values one or more of a set of unknown parameters whose effect on the observed data was very similar (e.g. parameters that appear as a sum) or any parameter which has neglible effect on the data. The best fits to the M2/F3 data was obtained with the derivatives L_p, L_r, L_β, N_p, N_r and all the δ_r derivatives fixed. The quality of the fit, however, was below that obtained without any parameters fixed.

The other technique involved eliminating from the set of allowable

values for the parameter estimates those eigenvector directions about which very little information could be gained from the data. These singular directions are associated with the smaller eigenvalues of the information matrix. When applied to the M2/F3 data, three singular directtions were determined. Unlike the other two techniques of fixing parameters or a priori weighting, the fit to the observed data remained very good and most of the sign problems were corrected. It is felt that this method offers great promise in future applications.

2.3 Optimal Input Design

As was shown with the X-22 simulated data, the use of the proper control input sequence can greatly improve the quality of the parameter estimates. This is done by maximizing the sensitivity of the system response to the unknown parameters to be identified. The optimal input to be used in system identification would therefore be one which optimizes some criterion based on the output sensitivity with respect to all the parameters to be identified.

During the second part of this contract, a computer program was developed which determines, for an arbitrary linear system and an arbitrary selection of parameters to be identified, the optimal input for parameter identification. The two criteria for optimality used in this program are (1) maximum sum of the (squares of the) the output sensitivities and (2) maximum product of the squares of output sensitivities. The first criterion is related to the trace of the Fisher Information Matrix and the second criterion is related to the determinant of the same matrix. The Information Matrix itself is the inverse of the Cramer-Rao lower bound for the covariances of the parameter estimates.

The only constraint put on the input is one of total energy. State and input amplitude constraints can be imposed indirectly by changing the input energy content. In addition algorithms have been added which will

specify the optimal input for a specified data length as well as investigating the frequency content of the optimal input.

The major emphasis in the optimal design part of the contract was in two areas. The first was to investigate the properties of the optimal input with respect to frequency content, comparison with a suboptimal input and the effect of parameter uncertainties. The second involved a Monte Carlo simulation of the identification process involving comparisons of an optimal input and a suboptimal doublet input for identifying the short period dynamics of a C-8 aircraft.

2.3.1 Optimal Input for C-8 Aircraft Identification

The optimal input for identifying the five stability and control derivatives associated with the short period longitudinal dynamics of a C-8 aircraft (assuming a priori wind-tunnel parameter values) was derived using the trace of the Information matrix criterion. When compared to the suboptimal doublet input of equal energy and duration the optimization criterion was almost 20 time as large. Frequency domain analysis inferred that the input was made up of a DC component to identify the gain parameters (control derivatives) and a sinusoidal component at the system natural frequency. This was to maximize output signal power and optimize the identification of the stability parameters. It was further found that if the optimal input was determined based on estimates of the stability derivatives which were 10% in error, the qualitative character of the optimal input did not change, and the accuracy with which the parameters could be identified remained approximately the same.

The last exercise of this first part was to determine the optimal input based on the second performance criterion viz. maximizing the product of the diagonal elements of the information matrix. Based on the value of the expected standard deviations for the parameter estimates, the optimal input determined from this second criterion was much improved over that

determined from the first.

2.3.2 Monte Carlo Simulation

The more realistic test for the optimal input is to use it under actual identification conditions, to determine if the statistics of the parameter estimates and computed information matrices match those predicted from a priori analysis. A four state linear model of the full C-8 longitudinal dynamics was used in generating the simulated flight data, with the control input designed to identify only the five short period dynamics. In addition, the control input was designed with each of the short period stability and control derivative changed by 50% from the values used in the data generation. This was to model the situation where the control sequence for aircraft identification is detemined from a priori wind tunnel or theoretical derivative values. 50 complete identification runs were made both with the optimal input and with a suboptimal doublet input of equal energy and duration.

The parameter estimates from each run were compiled and total results evaluated based on 50 runs. With all measures of comparison, the optimal input greatly out performed the suboptimal input. Histograms of the parameter estimates were also compiled and compared. The results after 50 runs closely matched the results predicted by the Cramer-Rao lower bound. Experiments were also run by modifying the optimal and the doublet inputs through the servo transfer functions. The use of optimal inputs in flight testing for the determination of aircraft stability and control derivatives appears, therefore, to be a very useful and powerful tool.

III

BACKGROUND FOR AIRCRAFT PARAMETER IDENTIFICATION

Although extensive time and effort, over a period of the last 20 years, has been expended in the development of more exact aircraft stability and control derivative identification techniques, up until recently, the extraction of these derivatives from flight data remained a difficult and time-consuming problem. An emphasis on working directly with flight data, in addition to dealing with wind tunnel tests or theoretical calculations, has evolved as a result of what is often gross disagreement between wind tunnel and flight test derivatives, as well as the known difficulties of obtaining dynamic derivatives and extrapolating them to full scale derivatives from the wind tunnel values.

There have been many methods proposed and tried for extracting stability and control derivatives from flight data. Most of these have proved to be successful only under idealized conditions such as no wind gusts or modeling errors and known instrumentation accuracies. Very often a good deal of the data collected during a flight test program has to be discarded for lack of a technique which is general enough to process it under less than ideal conditions.

The emergence of the digital techniques during the past few years, resulting in the development of the Maximum Likelihood Identification Techniques, has given rise to the realization that much of the previously discarded data can be successfully processed. As the limitations of the instrumentation system, flight control input and inadequate aerodynamic model are recognized and compensated for, and the presence of wind gusts is included in the model structure and accounted for in the identification algorithm, the best set of identified values for the stability and control derivatives can be obtained.

This section will discuss several of the previous identification techniques which have been used to process flight data. This discussion will bring out the similarities that exist among these methods, and mention the aircraft flight data to which the methods have been applied. The next section, then, will provide a detailed explanation of the Maximum Likelihood Technique to be applied later on to flight test data containing gusts.

3.1 Previous Identification Methods

Although a large number of identification methods have been used in the past, only some of the more common methods which are currently in use will be described.

3.1.1 Time Vector Method

The time vector methods for derivative identification is derived from the time-invariance of the amplitude and phase relations between the state variables (degrees of freedom) of an exponentially damped second order system and the derivatives and integrals of the state variables. This invariance is used to determine the values of the amplitude-phase relations, thereby determining the aircraft stability and control derivatives (Ref. 1).

When more than one state variable (degree of freedom) is involved in the transient response, and there is a common natural frequency, the instantaneous value of any one state may be readily determined if the characteristics of any one of the motions are known, along with the amplitude ratio and phase angle relative to the characteristic motion. The time invariance of the amplitude ratios and their phase angles permits the representation of any one of the linearized equations of motion as vectors. The properties of these vectors, plus the requirement that the vector sum of the quantities in any one equation equal zero, makes possible the determination of two unknown derivatives in any one equation.

As Ref. 2 points out, the time vector method has the principal disadvantage that it can only be applied to control-fixed, transient-oscillation aircraft responses with damping ratios less than $\tilde{}$.3. Furthermore, the successful application of the time-vector method is highly dependent on the operators' skill.

3.1.2 Analog-Matching Methods

The analog matching technique is actually an output error method since it strives to iteratively minimize the errors of the various responses through operator manipulation of the values of the stability and control derivatives. It is often used as a backup method for validating the more modern digital techniques. However, there are several disadvantages to the analog matching technique. First, the method works most successfully only when a single control surface is moved at a time and then only when the maneuvers are simple.(Ref. 1). Second, when the maneuvers are made with a stability augmentation system or other form of dependent control input, the data is difficult to analyze. Finally, this method is extremely time consuming, even in face of the fact that recent procedures, through the use of hybrid computers, has reduced the time considerably. For example, the time involved in analyzing a lateral-directional flight maneuver, from receipt of flight data to final results, is approximately four hours (Ref. 2); the analog matching technique is also extremely susceptible to uniqueness problems since the success of a data analysis is very dependent upon the type of control maneuvers used. In such a case the skill and knowledge of the operator would play an important part in determining the ultimate success of the analysis.

The analog matching technique has been used by the Air Force Flight Test Center (Ref. 3), the Naval Air Test Center (Ref. 4), and the NASA Flight Test Center (Ref. 5) for the F-104, X-15, B-70, HL-10, M2/F3, X-24 and PA-30 aircraft.

Most of the remaining identification techniques, almost all of which require the use of a digital computer, can be classified as either

1. Equation error methods,
2. Output error methods, or
3. Advanced methods.

These methods differ by (1) the performance criterion that they are developed from, (2) the kinds of estimates they produce, and (3) the problems to which they can be applied.

3.1.3 Equation Error Methods

Equation error methods (Ref. 6) assume a performance criterion that minimizes the square of the equation error (process noise). All of these methods are basically least squares techniques and, in general, it is necessary to measure all the response variables and their derivatives. The procedure is to express the stability and control derivatives as functions of the measured responses using the equations of motion. This results in n or more linear equations in n unknowns. For those cases where the time derivatives are not measured, various "method functions" are used to operate on these equations (take time derivatives, Laplace or Fourier transforms, etc.) to obtain equations that are linear in the unknown stability and control derivatives. Since these methods do not allow for measurement errors (instrumentation errors), they result in biased estimates when this type of error does exist. The principal use of these methods are as start-up techniques for the output error and advanced methods.

The equation error methods have been used or are being used by Cornell Aeronautical Laboratory (Ref. 6), Air Force Flight Test Center (Ref. 3), and Delft University of Technology (Ref. 7).

3.1.4 Output Error Methods

Output Error Methods (Refs. 8 through 17) minimize the square of the error between the actual system output and the output of a model used to represent the actual system. This method assumes measurement noise but no process noise. Typical output error methods include Newton-Raphson, Gradient methods, the Kalman Filter (without process noise), and modified Newton-Raphson, differential correction, and quasilinearization (all three of which are the same method).

The modified Newton-Raphson method has been used extensively in flight test applications for the past several years. It is the one method that has been used on an operational basis and for which the most experience exists. This method has been or is being used by (among others): (1) the NASA Flight Test Center (Ref. 5) on the X-24, X-14, XB-70, 990, HL-10, M2/F3, Jet Star and PA-30 aircraft; (2) the NASA Ames Research Center (Reference 18) on the Learjet, XV-5, 990 and the C-8 aircraft; and (3) the NASA Langley Research Center (Refs. 19,20) on the XC-142, Navion and F-4 aircraft. (NASA Langley program has automatic update of the weighting matrix based on the maximum likelihood criterion.)

The principal disadvantages of the output error methods is that, because they do not include process noise in their performance criterion, the results degrade when process noise (gusts, modelling errors) exists. This may result in the computer program not converging or in estimates that have large variances or poor estimates (Ref. 21). However, as long as these methods are applied to linear flight regions, or where the form of the equations is known, or where gusts do not exist, they work very well.

3.1.5 Advanced Methods

The most general identification problem is one of extracting stability and control derivatives, for non-linear aircraft models, from flight data

containing both measurement and process noise. The one advanced technique that has demonstrated the capability of extracting stability and control derivatives from flight data under these circumstances is an implementation of the maximum likelihood criterion. This numerical algorithm, developed by SCI, is a combination of three steps: (1) Kalman filtering to estimate the states and generate a residual sequence, (2) a modified Newton-Raphson algorithm for the parameter estimates, and (3) an algorithm to estimate the noise statistics (means and variances of the measurement and process noise). The details of the numerical method are outlined in the next section.

The success of the SCI maximum likelihood technique can be attributed to several important attributes of this method:

1. It does not require a priori knowledge of the process noise covariance, measurement noise covariance or the initial parameter estimate covariance. These covariances are determined as part of the identification procedure.
2. When process noise does not exist, this method simplifies to the modified Newton-Raphson output error method (although with a specific weighting matrix).
3. When no measurement noise exists (an unlikely event) this method simplifies to the least squares equation error method.
4. The Cramer-Rao lower bound on the covariance of the error in the stability and control derivative estimates are obtained as part of the algorithm.
5. The minimum mean-square aircraft state variables (response variables) are obtained as an integral part of the algorithm. It is not required, however, that initial state estimates be supplied.

The following section gives a detailed derivation of the Maximum Likelihood Identification Method.

IV

MAXIMUM LIKELIHOOD (ML) IDENTIFICATION

The notion of the maximum likelihood estimate which was introduced into statistics by R. A. Fisher in 1906 is based on a relatively simple idea. Assume that the outcome Z of an experiment depends on an unknown parameter θ. We want to infer the best value of θ from the observation Z. One answer is to choose that value of θ which makes the observed value Z the most probable one to have occurred. This can be rigorously stated as: choose θ to maximize the conditional probability of Z, given a value of θ ; i.e.

$$\hat{\theta} = \max_{\theta} \; p(Z|\theta)$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ and $p(Z|\theta)$ is the conditional probability of Z, given θ. The same estimate is obtained by maximizing $\log \; p(Z|\theta)$ which is known as the likelihood function.

The above basic idea can be carried over to linear and nonlinear dynamic systems, with process and measurement noise, but the details of the application become quite involved. In practice, there are two major problems in obtaining ML estimates for dynamic systems. These are:

1. Deriving an expression for the likelihood function, and
2. Maximizing the likelihood function with respect to the unknown parameters.

These two problems are elaborated upon further. The likelihood function is the logarithm of the joint probability density of the observations given the parameters. If the observations are independent, the joint probability density function is easily written down since it is just the product of the

probability densities of each observation given the parameters. The derivation of the likelihood function becomes much more difficult when the observations are correlated. This is necessarily the case for dynamic systems with random inputs since the state at any time is correlated with the state at all the previous times. In the next section, it is shown how the likelihood function for a dynamic system can be derived in a simple form using a Kalman filter and the resulting white noise innovation sequence. This is shown schematically in Figure 4.1.

The second problem in obtaining ML estimates is a computational one. Generally, the likelihood function is highly nonlinear in terms of the parameters. For finite data lengths, it is also known to have several local maxima. In the case of dynamic systems, certain differential equation constraints have to be satisfied. The choice of a suitable search algorithm is very important for the successful application of ML identification.

The maximum likelihood identification method, as implemented by SCI, is an adaption and extension of the recent work of Astrom (Ref. 22), Kashyap (Ref. 23) and Mehra (Ref. 24). It is capable of solving the most general

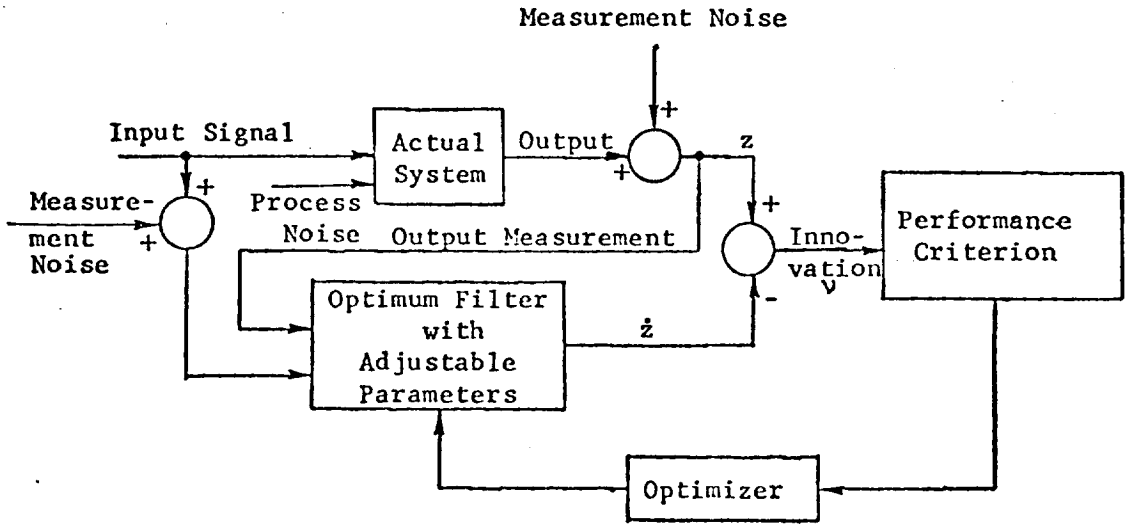


FIGURE 4.1 IMPLEMENTATION OF MAXIMUM LIKELIHOOD ESTIMATOR

identification problem, including (1) systems governed by non-linear differential equations of motion, (2) the presence of additive random process noise in the equations of motion, and (3) random disturbances corrupting the measurements of the system inputs and outputs. In this section the development of the theory justifying the use of the likelihood function as an optimizing criterion for identification is developed and the numerical algorithm for implementing the maximum likelihood identification method is outlined. The development is first given for linear systems and then extended to nonlinear systems.

4.1 Linear Systems:

Consider the linearized aircraft equations of motion

$$\dot{x} = Fx + Gu + \Gamma w \tag{4.1}$$

where

$x(t)$ = n x 1 state vector (p, q, r, u, v, w, etc.)
$u(t)$ = p x 1 input vector (δ_e, δ_a, δ_r)
$w(t)$ = q x 1 vector of random forcing functions

Let the measurement equations be

where $y(t) = Hx(t) + Du(t) + v(t)$
$y(t)$ = r x 1 output or measurement vector
$v(t)$ = r x 1 vector of random measurement errors (4.2)

and

$$E\{ w(t) \} = 0, \quad E\{ w(t)w^T(\tau) \} = Q\,\delta(t - \tau)$$

where $\delta(t - \tau)$ is the Dirac Delta function.

$$E\{ w(t)v^T(\tau) \} = 0$$

$$E\{ v(t) \} = 0, \quad E\{ v(t)v^T(\tau) \} = R\delta_{t,\tau}.$$

It is assumed that the structure of the model is known. The vector of unknown parameters in F, G, Γ, H, Q, R and x(o) are denoted by θ. Thus θ includes all the unknown stability and control derivatives, noise variances and the initial states.

If a sequence of observations y(1),.....,y(N) is made of the system state with noise, the maximum likelihood estimate of θ, following the idea stated earlier, is given by

$$\hat{\theta} = \max_{\theta} p(Y_N|\theta) \tag{4.3}$$

where $Y_N = \{y(1),....., y(N)\}$. With successive applications of Bayes Rule, an expression for $p(Y_N|\theta)$ can be derived as

$$\begin{aligned} p(Y_N/\theta) &= p(y(1),... y(N)/\theta) \\ &= p(y(N)|Y_{N-1}, \theta)p(Y_{N-1}|\theta) \\ &= p(y(N)|Y_{N-1}, \theta)p(y(N-1)|Y_{N-2}, \theta)p(Y_{N-2}|\theta) \end{aligned}$$

$$= \prod_{j=1}^{N} p(y(j)|Y_{j-1}, \theta)$$

Since the logarithm is a monotonic function, the maximum likelihood estimate can also be written as

$$\hat{\theta} = \max_{\theta} [\log p(Y_N|\theta)] = \max_{\theta}\left[\sum_{j=1}^{N} \log p(y(j)|Y_{j-1},\theta)\right] \tag{4.4}$$

where $\log p(Y_N|\theta)$ is the likelihood function.

If $x(o)$, $w(t)$ and $v(t)$ are normally distributed, $p(y(j)|Y_{j-1},\Theta)$ will also be normal and can be uniquely determined by computing the mean and covariance. Therefore define

$$E\{y(j)|Y_{j-1}, \Theta\} \triangleq \hat{y}(j|j-1) \tag{4.5}$$

and

$$\begin{aligned} \text{cov}\{y(j)|Y_{j-1},\Theta\} &= E\{(y(j) - \hat{y}(j|j-1))\ (y(j)-\hat{y}(j|j-1))^T\} \\ &\triangleq B(j|j-1) \end{aligned} \tag{4.6}$$

With these assumptions, the term $\log p\ (y(j)|Y_{j-1},\Theta)$ can be written as

$$\begin{aligned} \log p(y(j)|Y_{j-1},\Theta) = \text{Const.} - \frac{1}{2}(y(j)-\hat{y}(j|j-1))^T\ B^{-1}(j|j-1)(y(j)-\hat{y}(j|j-1)) \\ - \log |B(j|j-1)| \end{aligned} \tag{4.7}$$

The problem of determining the maximum likelihood estimate has now become one of finding a way of calculating the conditional mean, $\hat{y}(j|j-1)$, and the error covariance , $B(j|j-1)$. These quantities, however, are precisely the output of a Kalman Filter (Ref. 25) state estimator given Θ. This filter is designed to recursively process measurements one at a time, and, at each point produce the minimum variance state estimate based on all the data received up to that point.

The Kalman filter prediction and update equations can be derived as follows*:

*For a more rigorous derivation of the Kalman Filter, see either Kalman (Ref. 26) or Kailath (Ref. 27). Also note that the conditioning on Θ has been omitted from the equations to simply notation.

Initial Conditions: The Kalman filter is started with a priori state estimate $\hat{x}(0|0)$ and covariance $P(0|0)$.

The state prediction is done using the equations of motion and state update is done using the measurements.

Prediction Equations: At time (j-1), the Kalman filter has a state estimate $\hat{x}(j-1|j-1)$ and covariance $P(j-1|j-1)$. It is required to predict the state at time j. The resulting state estimate is denoted by $\hat{x}(j|j-1)$ and has a covariance of $P(j|j-1)$. The relationship between the updated and the predicted estimates can be obtained by taking conditional expectations on both sides of Eq. (4.1) and interchanging the operations of expectation and differentiation. This gives

$$\frac{d}{dt}\hat{x}(t|j-1) = F\hat{x}(t|j-1) + Gu(t) \qquad (4.8)$$

$$(j-1) \leq t \leq j$$

where the predicted value of the white noise w(t) based on previous information is equal to zero.

The covariance equation for the predicted estimate $\hat{x}(t/j-1)$ can be obtained by subtracting Eq. (4.8) from Eq. (4.1) and using the covariance propagation equation derived in Bryson and Ho (Ref. 25).

$$\frac{d}{dt}P(t|j-1) = F\,P(t|j-1) + P(t|j-1)F^T + \Gamma Q\Gamma^T \qquad (4.9)$$

Update Equations: The update equations for the Kalman filter can be derived using a well-known property of the conditional normal distributions (Ref. 28), viz.,

$$E\{\, a|b \,\} = \bar{a} + P_{ab}P_{bb}^{-1}(b - \bar{b}) \qquad (4.10)$$

$$\mathrm{cov}(a|b) = P_{aa} - P_{ab}P_{bb}^{-1}P_{ab}^T \qquad (4.11)$$

where a and b are normal random variables with

$$E\{a\} = \bar{a} \quad , \qquad E\{b\} = \bar{b}$$

$$\mathrm{cov}(a) = P_{aa} \quad , \qquad \mathrm{cov}(b) = P_{bb}$$

$$E\{ (a-\bar{a}) (b-\bar{b})^T \} = P_{ab}$$

Replacing a by $x(j)$ with mean $\bar{a} = \hat{x}(j|j-1)$ and covariance $P_{aa} = P(j|j-1)$ and replacing b by $y(j)$ with mean $\bar{b} = \hat{y}(j|j-1) = H\hat{x}(j|j-1)$ and covariances

$$P_{bb} = HP(j|j-1)\, H^T + R \tag{4.12}$$

$$P_{ab} = P(j|j-1)\, H^T \tag{4.13}$$

we obtain,

$$\hat{x}(j|j) = \hat{x}(j|j-1) + K(j)\,(y(j) - H\hat{x}(j|j-1)) \tag{4.14}$$

$$K(j) = P(j|j-1)H^T\,(HP(j|j-1)H^T + R)^{-1} \tag{4.15}$$

and

$$P(j|j) = (I - K(j)H)\, P(j|j-1) \tag{4.16}$$

The quantity $(y(j) - \hat{y}(j|j-1))$ represents the new information brought forth by the measurement $y(j)$. It is known as the "innovation" sequence and has been shown to be zero mean, Gaussian and white (Ref. 29) Denoting the innovations by $\nu(j)$, the likelihood function can be written as

$$\log p(Y_N|\Theta) = -\frac{1}{2} \sum_{j=1}^{N} \{ \nu^T(j) B^{-1}(j|j-1)\, \nu(j) + \log|B(j|j-1)| \} \tag{4.17}$$

where $B(j|j-1) = HP(j|j-1)H^T + R$ (4.18)

The maximum likelihood estimate $\hat{\Theta}$ is obtained by maximizing (4.17) with respect to Θ, subject to the constraints in equation (4.8)-(4.9), (4.14)-(4.16). This is a very difficult optimization problem. An approximation suggested in Ref. 24 simplifies the problem tremendously. It is assumed that the filter gain K(j) and covariance B(j|j-1) have reached constant values K and B. The vector Θ of unknown parameters is now defined to include (in addition to F,G) K and B instead of Q and R. Reference 57 gives a detailed derivation of the relation between K, B and Q, R. Then

$$\log p(Y_N|\theta) = -\frac{1}{2}\sum_{j=1}^{N}\{\nu^T(j)B^{-1}\nu(j) + \log|B|\} \tag{4.19}$$

Maximizing (4.19) over B, produces

$$\hat{B} = \frac{1}{N}\sum_{j=1}^{N}\nu(j|\hat{\alpha})\nu^T(j|\hat{\alpha}) \tag{4.20}$$

where $\hat{\alpha}$ is the ML estimate of unknowns of F, G and K. It is given by the root of the equation

$$\sum_{j=1}^{N}\nu^T(j)B^{-1}\frac{\partial\nu(j)}{\partial\alpha} = 0. \tag{4.21}$$

where $\frac{\partial\nu(j)}{\partial\alpha}$ is calculated from Eq. (4.8) - (4.18). The root of (4.21) is found by a Newton-Raphson iteration. Once $\hat{\alpha}$ is obtained, R and Q can be obtained from equations (4.9) and (4.18). In this way the non-linear constraints imposed by the equations (4.9), (4.15), (4.16) and (4.18) are avoided during optimization.

4.2 Nonlinear Systems

The approach to obtaining the maximum likelihood parameter estimates for nonlinear models is conceptually similar to that for linear models.

Consider a nonlinear dynamic system model of the form

$$\dot{x}(t) = f(x(t), \theta, u(t)) + \Gamma w(t) \tag{4.22}$$

$$y(t) = h(x(t)) + v(t) \tag{4.23}$$

where $f(\cdot)$ and $h(\cdot)$ are $n \times 1$ and $r \times 1$ vectors of nonlinear functions. Also, $w(t)$ and $v(t)$ are Gaussian white noise sequences with zero mean and covariances Q and R.

The evaluation of the exact maximum likelihood estimate involves the calculation of the conditional probability $p(y(j)|Y_{j-1},\theta)$ as in the linear model case. This would require an optimal nonlinear filter, which, to date, is computationally unfeasible since a complete description of $p(y(j)|Y_{j-1},\theta)$ requires computing all its moments. As a result, it is proposed to use an Extended Kalman Filter (Ref. 30) of the following form:

$$\dot{\hat{x}} = f(\hat{x}(t|j-1), \theta, u(t)) \tag{4.24}$$

$$\hat{x}(j|j) = \hat{x}(j|j-1) + K(j)\nu(j) \tag{4.25}$$

$$\nu(j) = y(j) - h(\hat{x}(j|j-1)) \tag{4.26}$$

The Kalman gain K(t) is calculated from equations (4.13)-(4.16) by using the time varying matrices H and F, defined by

$$H(t) = \left.\frac{\partial h}{\partial x}\right|_{x = \hat{x}(j|j-1)} \tag{4.27}$$

$$F(t) = \left.\frac{\partial f}{\partial x}\right|_{x = \hat{x}(j|j)} \tag{4.28}$$

Notice that the Extended Kalman filter linearizes the equations around the latest best estimate of the state. More advanced filters such as Second Order Filters, Single Stage Smoothing Filters, etc. (Ref. 31) can be used for state estimation, but for the aircraft parameter identification problem, when all the states and accelerations are being measured accurately, an Extended Kalman filter comes quite close to the optimal nonlinear filter in accuracy.

Kailath (Ref. 27) has shown that the density of the innovation $\nu(t)$ tends to a Gaussian density as the sampling rate is increased. Thus, for high sampling rates the likelihood function can again be written as

$$J \equiv \log(Y_N|\Theta) = -\frac{1}{2}\sum_{j=1}^{N} \nu^T(j)\, B^{-1}(j)\, \nu(j) + \log|B(j)| \qquad (4.29)$$

The validity of the above two assumptions viz: high sampling rates and accurate measurements should be checked in practice for each application of this method.

Remark:

The use of an Extended Kalman filter here is for state estimation only. It is also possible to use an Extended Kalman Filter for simultaneous state and parameter estimation (Refs. 21, 32). In the authors' opinion, this is not desirable since the uncertainties in the states are much smaller than the uncertainties in the parameters. Therefore, the assumptions of linearization which are valid for state estimation are generally not valid for parameter estimation in the aircraft parameter identification problem. Moreover, the Extended Kalman Filter for simultaneous estimation of the state and the parameters assumes knowledge of the a priori covariances which are unknown for the parameters. This is one of the reasons why an Extended Kalman filter typically gives unreliable confidence limits on the parameter estimates (Ref. 21). The maximum likelihood method described here will be shown to provide realistic estimates of confidence limits on a test case. It has also been found to converge in several cases where the Extended Kalman Filter for simultaneous state and parameter estimation failed to converge properly due to poor a priori values for the parameters.

4.3 Numerical Optimization Algorithm

The optimization algorithm described here for obtaining the maximum of the likelihood function is the Modified Newton-Raphson or Quasilinearization Method.* Determining an update to a set of parameter estimes θ, which will decrease the value of the likelihood function (cost), J, using this method involves computing two matrices: the gradient of the cost with respect to the unknown parameters, $\frac{\partial J}{\partial \theta}$, and the information matrix, M.

For the case of a nonlinear system with process noise, the likelihood function, J, is computed using an extended Kalman filter. The gradient and information matrix computation must therefore include at least the first order partials of the Kalman gain with respect to the parameters. With the nonlinear system model given by Equations (4.22)-(4.23), and the extended Kalman filter by Equations (4.24)-(4.28), the gradient of J with respect to θ is given by

$$\frac{\partial J}{\partial \theta} = \sum_{j=1}^{n} \nu^T(j)\, B^{-1}(j)\, \frac{\partial \nu(j)}{\partial \theta} - \frac{1}{2}\, \nu^T(j)\, B^{-1}(j)\, \frac{\partial B(j)}{\partial \theta}\, B^{-1}(j)\, \nu(j) + \frac{1}{2}\, \mathrm{tr}\left(B^{-1}(j) \frac{\partial B(j)}{\partial \theta}\right) \tag{4.30}$$

where $$\frac{\partial \nu(j)}{\partial \theta} = -H(t)\, \frac{\partial \hat{x}(j|j-1)}{\partial \theta} - \frac{\partial h}{\partial \theta} \tag{4.31}$$

and $$\frac{\partial B(j)}{\partial \theta_k} = \frac{\partial H(j)}{\partial \theta_k}\, P(j|j-1)\, H^T(j) + H(j)\, \frac{\partial P(j|j-1) H^T(j)}{\partial \theta_k} + H(j)\, P(j|j-1)\, \frac{\partial H^T(j)}{\partial \theta_k} + \frac{\partial R}{\partial \theta_k}$$

---

* Based on prior experience (Ref. 33), the convergence of other optimization algorithms, including conjugate gradient and Davidon method, has been found to be slower than that of the quasilinearization method.

The recursive equations for $\frac{\partial P(j|j-1)}{\partial \theta_k}$ are obtained directly from the Riccati equations, with the separation into update at the measurement times and prediction between measurement times again being made.

Prediction:

From Equation (4.9), the prediction equation for $(j-1) \leq t \leq j$ becomes

$$\frac{\partial \dot{P}(t|j-1)}{\partial \theta_k} = \frac{\partial F}{\partial \theta_k} P(t|j-1) + F \frac{\partial P(t|j-1)}{\partial \theta_k} + \frac{\partial P(t|j-1)}{\partial \theta_k} F^T$$

$$+ P(t|j-1) \frac{\partial F^T}{\partial \theta_k} + \frac{\partial \Gamma}{\partial \theta_k} Q\Gamma^T + \Gamma \frac{\partial Q}{\partial \theta_k} \Gamma^T + \Gamma Q \frac{\partial \Gamma^T}{\partial \theta_k} \qquad (4.32)$$

Update:

The update equation, at the j^{th} measurement point, is obtained from Equation (4.16)

$$\frac{\partial P(j|j)}{\partial \theta_k} = (I-K(j)H) \frac{\partial P(j|j-1)}{\partial \theta_k} - \frac{\partial k(j)}{\partial \theta_k} HP(j|j-1)$$

$$-K(j) \frac{\partial H}{\partial \theta_k} P(j|j-1) \qquad (4.33)$$

where

$$\frac{\partial K(j)}{\partial \theta_k} = \frac{\partial P(j|j-1)}{\partial \theta_k} H^T [HP(j|j-1) H^T + R]^{-1}$$

$$+ P(j|j-1) \frac{\partial H^T}{\partial \theta_k} [HP(j|j-1) H^T + R]^{-1}$$

$$- K(j) \left[\frac{\partial H}{\partial \theta_k} P(j|j-1) H^T + H \frac{\partial P(j|j-1)}{\partial \theta_k} H^T \right.$$

$$\left. + HP(j|j-1) \frac{\partial H^T}{\partial \theta_k} + \frac{\partial R}{\partial \theta_k} \right] (HP(j|j-1)H^T+R)^{-1} \qquad (4.34)$$

The recursive equations for the term $\frac{\partial \hat{x}(j|j-1)}{\partial \theta}$, defined as the sensitivity equations and appearing in Equation (4.31), are obtained from the update and prediction filter equations (4.24)-(4.25).

Prediction:

$$\frac{\partial \dot{\hat{x}}}{\partial \theta_k} = \frac{\partial f}{\partial \theta_k} + F(t) \frac{\partial \hat{x}}{\partial \theta_k} \; ; \quad j-1 \leq t < j \tag{4.35}$$

Update:

$$\frac{\partial \hat{x}(j|j)}{\partial \theta_k} = \frac{\partial \hat{x}(j|j-1)}{\partial \theta_k} + \frac{\partial K(j)}{\partial \theta_k} \nu(j) - K(j) \left[H(j) \frac{\partial \hat{x}(j|j-1)}{\partial \theta_k} + \frac{\partial h}{\partial \theta_k} \right] \tag{4.36}$$

These same sensitivity equations are used to compute the information matrix, which is given by

$$\begin{aligned} M_{k,\ell} \triangleq \frac{\partial^2 J}{\partial \theta_k \partial \theta_\ell} = \sum_{j=1}^{n} & \frac{\partial \nu^T(j)}{\partial \theta_k} B^{-1}(j) \frac{\partial \nu(j)}{\partial \theta_\ell} + \nu^T(j) B^{-1}(j) \frac{\partial B(j)}{\partial \theta_k} B^{-1}(j) \frac{\partial \nu}{\partial \theta_\ell}(j) \\ & - \nu^T(j) B^{-1}(j) \frac{\partial B(j)}{\partial \theta_\ell} B^{-1}(j) \frac{\partial \nu(j)}{\partial \theta_k} \\ & - \frac{1}{2} \operatorname{tr} \left\{ B^{-1}(j) \frac{\partial B(j)}{\partial \theta_\ell} B^{-1}(j) \frac{\partial B(j)}{\partial \theta_k} \right\} \end{aligned} \tag{4.37}$$

Note that second order partial terms and several first order partials (of matrix inverses) have been neglected. All the partial derivative terms appearing in Equation 4.37 can be obtained via Equation 4.31.

The update to the parameter estimates θ, or step size $\Delta\theta$, is then computed using the following equation

$$\Delta\theta = -M^{-1}\left(\frac{\partial J}{\partial \theta}\right)^{T} \tag{4.38}$$

Since M is the Fisher Information Matrix, M^{-1} provides the Cramér-Rao Lower Bound on the covariance of the θ estimates. The ML method approaches this lower bound asymptotically.

4.4 Relationship to Output Error and Equation Error Methods

As stated earlier, one of the principal advantages of the maximum likelihood method is that, under special circumstances, it reduces to the output error or equation error method, both of which have been widely used for extracting stability and control derivatives from flight test data.

For the case where there is no process noise present, i.e., $w(t) = 0$, the process noise covariance, $Q(t)$, is identically zero. With $P(0)$ either equal to zero, (if the initial state estimates are known) or small, this implies $P(t|t-1) \equiv 0$ for all t after some initial transient (see Equations (4.9)-(4.16)). The Kalman gain will then also be identically zero (Equation (4.15) and the innovations sequence reduces to

$$\nu(t) = y(t) - Hx(t)$$

for the linear case, and

$$\nu(t) = y(t) - h(x(t))$$

in the nonlinear case. In both cases $\nu(t)$ is exactly the output error. The only difference then between the maximum likelihood method and the more classical output error method is the choice of the weighting matrix. In the maximum likelihood method, it is given as R, the measurement noise

covariance, or its estimate. In the maximum likelihood method, is is chosen as

$$\hat{R} = \frac{1}{N} \sum_{j=1}^{N} \nu(j,\hat{\theta})\, \nu^T(j,\hat{\theta}).$$

For the case where no measurement noise exists, the measurement noise covariance R(t) is identically zero. For the case in which all the states and their time derivatives are measured, the likelihood function is the sum of squares of the equation error at sampling times. Thus, the ML estimates are identical to the equation error estimates.

4.5 Identifiability and Uniqueness Problems in Extraction of Stability and Control Derivatives

Although the maximum likelihood method discussed in the previous section represents one of the most advanced identification techniques developed to date, there still remain some basic problems associated with extracting stability and control derivatives from flight test data. Most of these problems can be classified under the heading of "identifiability," which is related to the degree of excitation for the particular modes of the system under investigation and the ability to identify the associated parameters. Identifiability also relates to whether the parameters themselves can be identified or whether they can only be identified as part of a linear combination. This section will discuss some of the symptoms and causes of identifiability problems, and a few of the methods which have been used to solve them.

4.5.1 Symptoms and Causes of Identifiability Problems

The most obvious symptoms of identifiability problems are physically nonrealizable parameter estimates and large associated error covariances. Either of these symptoms may arise for a number of different reasons. If the input sequence does not adequately excite some of the modes, or if the Stability Augmentation System is operating, thereby suppressing some of the

aircraft modes, the associated parameters may not be identifiable. If the model chosen to get the input-output data is inadequate, the parameters of that model may be forced to account for some major unmodeled effects. The estimated parameter values may, therefore, be quite different from what aerodynamic theory and previous results may indicate. If there are large, unaccounted for instrumentation errors or errors in the location of the c.g. and the sensors, again non-physical parameter values may result. Finally, such additional factors as too short a data length, local minima in the cost functional and poor initial parameter estimates may also result in non-physical parameter values.

Large error covariances principally result from poor input sequences and attempts at identifying too many parameters. The first factor reduces the sensitivity of the ouput to variations in some parameter values, and the second factor causes linear dependencies between parameter estimates. Since an extraneous parameter in the model does not, by definition, improve or degrade the fit to the observed data, its estimated value will be of no significance and the error covariance of the estimated value will be large.

Probably the most common identifiability problem encountered in processing flight data results from parameter dependencies. This may occur through a pair of parameters which always appear in the equations of motion together, as with C_{M_q} and $C_{M_{\dot{\alpha}}}$, or through a poor choice of inputs such that some of the aircraft response variables are linearly correlated, or it may occur through an overspecification of the number of parameters to be identified. In each case, the result of the dependencies is a nearly singular information matrix, which when inverted to obtain the step size in the parameter estimates, causes numerical problems.

Additional numerical problems associated with a nearly singular information matrix arise when the control input is expressible as a linear combination of the aircraft response variables. This matrix singularity results from the linear dependence between the partial of a response

variable with respect to a parameter in the input matrix and the partials with respect to a parameter of the dynamics matrix. Since the same singularity exists at each data point, the resulting information matrix will also be singular. A second input related problem arises when the input is of such a nature that the time histories of several of the aircraft response variables appear highly correlated. All the elements of the partial derivative of the output vector with respect to any one of the parameters will be the same, introducing a singularity.

4.5.2 Approaches to Identifiability Problems

Four different approaches have been used to alleviate identifiability problems. These are:

1. Fixing Parameters - The usual remedy for parameter dependencies has often been to fix some of the dependent parameters during identification. While this generally improves the numerical convergence the choice of a particular parameter to fix and the value at which it is fixed are generally not clear. Although it is possible to fix the parameters at the wind tunnel or theoretical (DATCOM) values, the estimated parameter values may depend upon these fixed values. In those instances where the wind tunnel values are inaccurate or DATCOM doesn't apply and no other a priori information is available, a better way of dealing with the parameter dependency is needed.

2. A priori Weighting - Whenever a priori values exist for certain parameters in a given model structure, they can be included in the maximum likelihood method by using a Bayesian formulation. If the a priori values have a Gaussian distribution, a quadratic term involving the weighted difference between the estimated parameter values and the a priori parameter values is added to the likelihood function. Depending on the weights given to these differences, it is possible to force any of the parameter values to the a priori ones. The a priori values for the

aircraft stability and control derivatives are usually derived from the wind tunnel estimates and theoretical calculations. The weights, which are the most subjective part of this technique, signify the confidence in the a priori values. An alternate procedure is to successively reduce the weights at each iteration, thereby discounting the dependence on a priori information. The main advantage of this procedure is numerical since the information matrix with a priori weighting is generally better conditioned than the one without it. This procedure is a special case of Tychonov Regularization used for solving an ill-conditioned set of equations (Ref. 34).

3. Constrained Optimization - If, from practical or theoretical considerations, a range of allowable values or relationships between the stability and control derivatives can be specified, they can be used as constraints on the parameter estimates to avoid non-physical estimates. Such a procedure would require a constrained optimization technique in lieu of the Newton-Raphson optimization method normally used (for the output error criterion). Including such parameter value constraints will most probably also reduce the convergence rate.

4. Rank-Deficient Solutions - Without any of the above remedies, the parameter identifiability problems will usually appear as a difficulty with inverting the information matrix and obtaining accurate parameter estimates and error covariances. This numerical problem can be related to the spread in the eigenvalues of the information matrix. A perfect dependency among the parameters should, strictly speaking, result in a zero eigenvalue. However, since round-off and other numerical errors prevent the matrix from being exactly singular, all the eigenvalues will be non-zero with a spread between the smallest and largest eigenvalue being many orders of magnitude. In such a case, it might be better to use a rank deficient solution for the inverse rather than

a full rank solution (Ref 35). That is, the inverse to the information matrix should be computed leaving out one or more of the smallest eigenvalues. Each eigenvalue which is left out relates to a singular direction in parameter space and, therefore, indicates a combination of parameters which cannot be identified uniquely. (see section 5.3.6)

The maximum likelihood identification program described below has options to use the above methods for solving identifiability problems. Further research in this important area is badly needed if identification programs are to be used on a routine basis for extracting stability and control derivatives. It should be mentioned that two other topics related to identifiability (one of which is discussed elsewhere in this report) are those of input design and model structure determination.

4.6 Maximum Likelihood Identification Program

This section describes the computer program that was designed to implement the maximum likelihood identification method for extracting stability and control derivatives from flight test data. Three options are provided for dealing with the identifiability problem: (1) a priori weighting, (2) fixing parameters at a priori values, and (3) rank-deficient solution for the information matrix inverse. At the outset of an identification run one of these three options is indicated (including the weighting matrix if a priori weighting is specified) and the program thereafter runs automatically. Step size cutting (in the event of a cost increase) and parameter bounding routines are always included in the algorithm, although they can be easily rendered inactive, if it is so desired.

The flowchart for the maximum likelihood program is shown in Fig. 4.2. The principal steps of the algorithm are all blocked out, omitting the numerical procedures used to compute such quantities as the solution to differential equations, matrix inverses, etc. The following paragraphs briefly outline the functions carried in each of the numbered blocks and how the logic progresses from block to block.

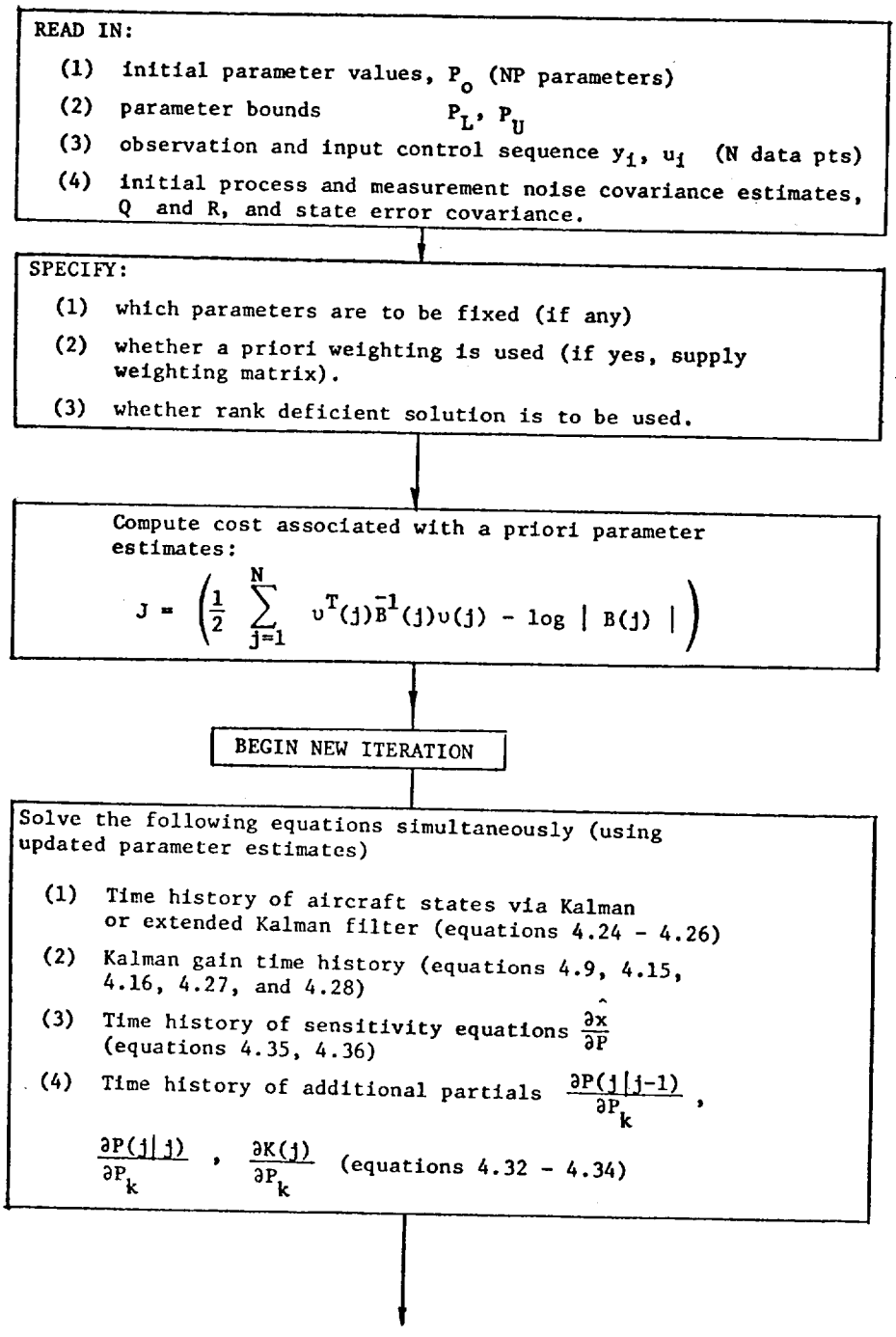


FIGURE 4.2 MAXIMUM LIKELIHOOD PROGRAM FLOWCHART

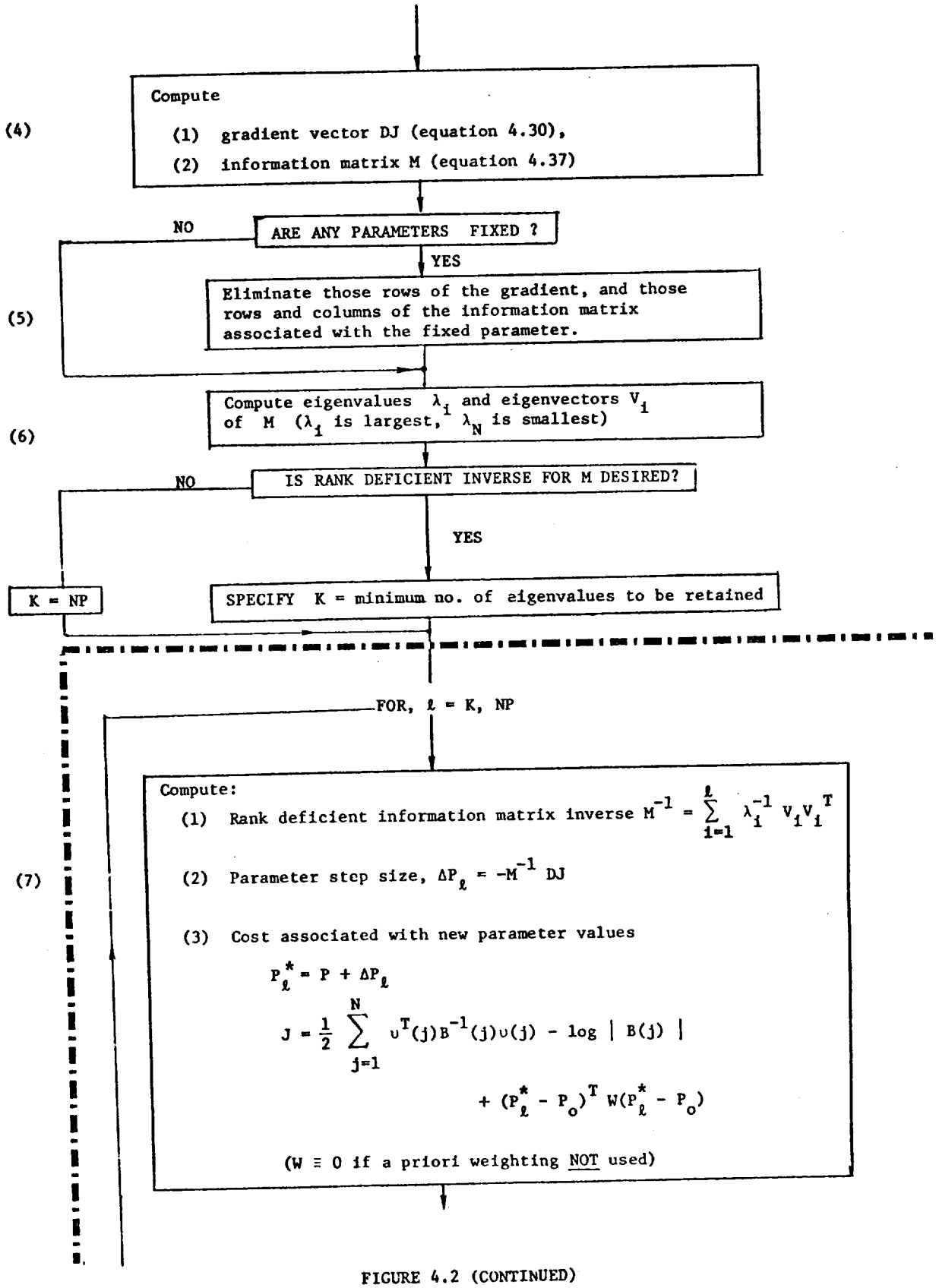


FIGURE 4.2 (CONTINUED)

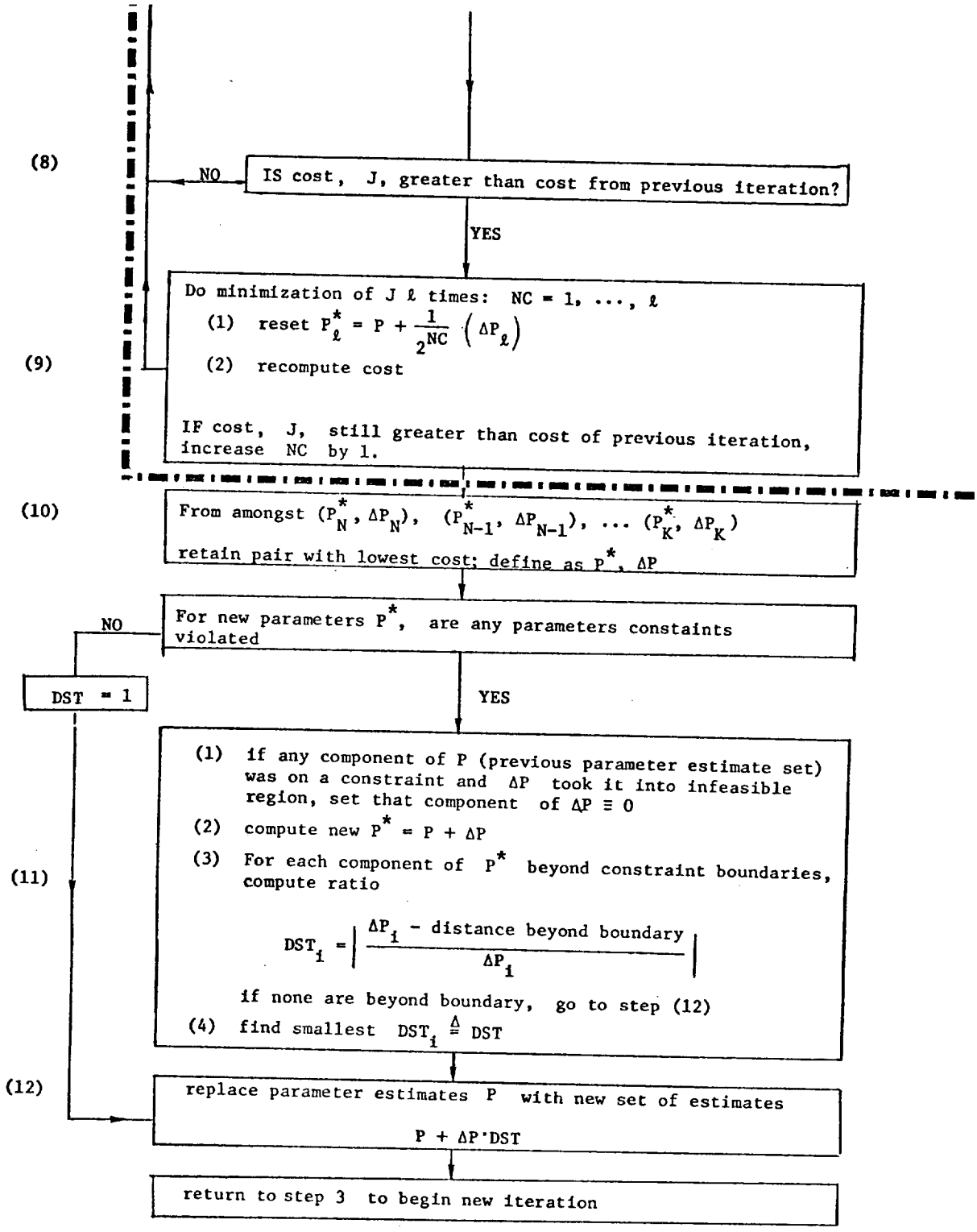


FIGURE 4.2 (CONTINUED)

The initialization procedure for the maximum likelihood identification method, indicated in blocks (1) and (2), consists of specifying a set of a priori parameter values, including the measurement and process noise covariances and, if desired, a set of upper and lower bounds for each parameter. The observations and input control time histories are then read in and stored. Since the maximum likelihood method is a batch processor, it will use the entire data record for each iteration. The initialization concludes by specifying which of the several options are to be used: (1) fixing parameters, (2) a priori weighting, (3) or rank deficient solution for the information matrix inverse.

With block (3) the first iteration begins. Using the equations given in Section 4.3 for the nonlinear system equations, the extended Kalman filter, the sensitivity functions and all the required partial derivatives, the time history for each of these quantities at each data point is computed. These differential equations can be solved using any one of a number of numerical techniques, e.g., Runge-Kutla. However, the majority of the computer time for each iteration will be consumed in solving these equations. When there is no process noise, Denery (Ref. 36) has shown that by using a transformation some of the sensitivity equations need not be evaluated, but rather can be expressed as a function of the others. The number of differential equations which need to be solved is thereby reduced.

In block (4) the time histories of the quantities computed in block (2) are combined to form the gradient, DJ, and the information matrix, M. Up to this point, all the computations can be performed considering a very complete and general set of parameters to be identified. This set may, however, be more general than needed for a particular application. For example, it may not be desirable to identify the rudder derivatives if there is no rudder input. If this is the case, the components of the information matrix and gradient due to the parameters which are not to be identified (thereby being considered fixed) must be removed. Following through the computation of DJ and M, this can easily be done by

simply deleting the rows of DJ and the rows and columns of M associated with the fixed parameters. This is performed in block (5).

Many of the computational problems associated with the Newton-Raphson optimization technique are involved with the large spread in the eigenvalues of M. Perhaps the most exact way of computing M^{-1} is therefore to use the eignvalue - eigenvector decomposition. This decomposition is performed in block (6). The eigenvalues and eigenvector are also required if a rank deficient inverse for M is desired.

The second program option consists of specifying if a rank dificient inverse is to be used, and if so, specifying, in addition, the minimum number of eigenvalues which are to be retained in computing M^{-1}. Note that if a rank deficient inverse is not desired, this minimum number is just set equal to the total number of parameters.

The logic for determining the rank deficient M is given in blocks (7) - (9). For each rank from the minimum to the full rank, the appropriate number of smallest eigenvalues are discarded and the information matrix inverse is computed. The associated parameter step is calculated and the likelihood function value is determined using the new set of parameter estimates. This involves computing the aircraft state and observation time histories and deriving the innovation sequence. Note that the third option enters in block (7) in the specification of whether a nonzero a priori weighting matrix is to be used.

Blocks (8) and (9) are concerned with the solution where the cost determined from the new parameter set is greater than the cost of the previous iteration. (If this is the first iteration, the previous cost is that associated with the a priori parameter estimates). When the new cost is higher, the parameter step size is cut in half and the cost reevaluated. If the new cost is still larger than the cost from the previous iteration, the step size is cut in half again. This same procedure is repeated a given number of times. The reason for this step size cutting is the nonquadratic nature of the likelihood surface.

This same step size cutting routine can be carried out for the rank deficient solution procedure. The end result is (k) sets of parameter estimates, each one resulting in some value of the likelihood function. In block (10), the set of parameter values and step sizes associated with the lowest cost is picked out and retained. The other sets of parameter values need not be saved.

The final option of the maximum likelihood program is to alter the parameter step size if any of the parameter constraints are violated. If this option is not desired, the parameter bounds are simply set to very large values. The routine for computing the optimal step size without exceeding the parameter constraints involves four calculations.* The first calculation checks the individual parameter values to see which ones are on a constraint. If the parameter step associated with any of these parameters results in violation of the constraint boundary, that step size is set equal to zero. In the second calculation, the new set of parameter estimates are computed, using the modified step size (some elements are zero).

In the third calculation, each component of the new set of parameter estimates is compared with the constraint boundaries. For any individual parameter value which is beyond the boundary, the absolute value of the ratio of the allowable parameter step to the actual parameter step is computed. This ratio is exactly the factor needed to have that particular parameter value fall on the constraint boundary. The smaller that factor, the farther beyond the constraint boundary the new parameter estimate would have been. In the last calculation, the smallest factor from among those computed for the individual parameter estimates is determined and retained.

---

* This procedure is based on the Generalized Reduced Gradient method of Abadie (Ref.37).

The final block (12) of the algorithm involves multiplying all the parameter step sizes by the smallest factor determined in block (11) or by 1, if the constrained otimization option was not chosen. If the option was used, only one additional parameter estimate will be on the constraint boundary. All other parameter estimates besides the ones with zero step sizes will be within the constraint boundaries.

The computation of a new step size for parameters marks the end of an iteration. To begin another iteration, these parameter values are used in the computations of block (3), and the cycle is restarted. The original cost now becomes the cost associated with these new parameter estimates.

V

RESULTS OF IDENTIFYING AIRCRAFT STABILITY AND CONTROL DERIVATIVES

This section discusses in detail, the experience and results of applying the maximum likelihood identification technique to simulated and real flight test data from three different aircraft. Included will be a discussion of the problems that were encountered and all the techniques that were used to alleviate them. Wherever possible, the cause of the problems is also spelled out along with possible implications for flight test procedures.

The first data that was used was from a computer simulation of X-22 VTOL aircraft. The aircraft model was highly non-linear and the data included process noise as well as measurement noise. Experiments were run with different input sequences and different measurement noise levels to investigate their effects on the parameter estimates. In all, 23 parameters were identified, excluding the measurement and process noise covariances, which were assumed known.

The second case involved actual flight data from an HL-10 lifting body. The digitized data, comprising approximately six and one-half seconds of flight, was supplied to SCI by NASA-Edwards FRC. A linear aircraft model was assumed in fitting the data and, in all, 20 parameters were identified, including the measurement noise covariance and the initial flight conditions. The data was assumed not to contain any wind gust (process noise) effects.

The third set of data, also supplied by Edwards FRC, was from an M2/F3 lifting body. This data, covering approximately eight seconds of the flight test, did contain wind gust effects and represented a test of the maximum likelihood technique in reducing flight data which had not been successfully reduced by the output error technique. In all

22 parameters were identified, including the measurement noise covariance, the parameters (time constant and driving noise covariance) of a wind gust model and the initial flight conditions.

5.1 X-22 Simulated Data

At the time this contract began, the maximum likelihood method had been applied to simulated X-22 data, containing the effects of gusts, with very promising results. However there remained several important problem areas which needed further investigation and improvements to be made to the existing program. This section will outline these problem areas, including the method of approach, results, and conclusions.

5.1.1 Generation of X-22 Simulated Data

The model of the longitudinal motion of the X-22 is given in Appendix A. These equations can be put in the nonlinear form

$$\dot{\underline{x}} = f(\underline{x}, c, p) + g(\underline{x}, p)\, v$$

where

$\underline{x} = [q, \Theta, u, w]^T$ is the 4-dimensional state vector ($q \triangleq$ pitch rate, $\Theta \triangleq$ pitch angle, $u \triangleq$ longitudinal velocity, $w \triangleq$ vertical velocity)

p is the 23 x 1 vector of unknown parameters (consisting of the coefficients of the polynominal expansion in u of the derivatives M_o, M_w, M_q, M_δ, X_o, X_w, X_δ, Z_o, Z_w, Z_δ)

c is the vector of deterministic control surface deflections and biases $= [1, \delta_{es}]^T$ ($\delta_{es} \triangleq$ elevator deflection)

v is a 3-dimensional white, Gaussian process noise with mean 0 and covariance Q.

The elements of the g matrix are obtained from the matrix of first partials $\frac{\partial f}{\partial x}$ and, therefore, the parameters and states appearing in f also appear in g.

The measurement equations are

$$\underline{z} = \begin{bmatrix} \underline{x} \\ \dot{q} \\ n_x \\ n_y \end{bmatrix} = \begin{bmatrix} \underline{x} \\ \dot{q} \\ \dot{u} + qw + g \sin\theta \\ \dot{w} - qu - g \cos\theta \end{bmatrix} + \begin{bmatrix} n_1 \\ \vdots \\ n_5 \\ n_6 \\ n_7 \end{bmatrix}$$

where

n_i , $i = 1,\ldots,7$ are independent, white, gaussian measurement noise samples with the properties $E\{\underline{n}_t\} = 0$ and $E\{\underline{n}_t \underline{n}_s\} = R\delta_{ts}$

Substituting for $\dot{u}$, $\dot{q}$, and $\dot{w}$, however, introduces, process noise into the measurement equations. The measurement equation can then be rewritten as

$$\underline{z} = \begin{bmatrix} \underline{x} \\ \\ f' \end{bmatrix} + \underline{n} + \begin{bmatrix} 0 \\ --- \\ g' \end{bmatrix} \underline{v}$$

where f' and g' are made up of specific rows of f and g, respectively. This gives rise to a correlation between the process noise and the measurement noise now consisting of the sum of the vector $\underline{n}$ and $\begin{bmatrix} 0 \\ --- \\ g' \end{bmatrix} \underline{v}$.

With the specification of an elevator deflection sequence, δ_{es}, and the process and measurement noise covariances, Q and R, the data, z, could be generated (using 4th-order Runga-Kutta integration of the nonlinear equations of motion). For each trial approximately 10 secs. of data was used, with a sampling rate of 20 per sec.

5.1.2 Program Description

The program that was initially used to extract the stability and control derivatives from the simulated data consisted of basically two parts. The first part was a least squares start-up routine (Ref. 13) which generated an initial estimate of the parameter values. This least squares technique is an equation error method which, in one pass through the data, obtains parameters estimates that minimize the following criterion

$$\min_{\hat{p}} \quad \left(\underline{\dot{x}}_i - f_i(\underline{x}, \hat{p})\right)^2, \quad i=i,..,4$$

for each derivative $\dot{x}_i$ which is measured or derived. Since, in the X-22 simulation, $\dot{q}$, n_x and n_y were measured, it was first necessary to express these quantities as linear functions of the parameters to be identified. From Appendix A it is possible to write $\dot{q}$ as a linear function of the parameters (polynomial coefficients) in the derivatives $M_o(u)$, $M_w(u)$, $M_q(u)$ and $M_{\delta_{es}}(u)$; n_x as a linear function of the parameters in $X_o(u)$, $X_w(u)$ and $X_{\delta_{es}}(u)$; and n_δ as a linear function of the parameters in $Z_o(u)$, $Z_w(u)$ and $Z_{\delta_{es}}(u)$. Since no parameter appears in more than one expression, a unique least squares estimate can be obtained for all the parameters.

The second part of the program was the maximum likelihood identification technique in a form designed to identify the parameters of a non-linear model when both the process and measurement noise covariances, Q and R, are known. The least squares parameter estimates were used as the initial conditions for the first iteration through the data

of the maximum likelihood routine. These estimates are updated with each iteration, until the algorithm converges. However, since one iteration through the data of the maximum likelihood technique required 1 minute of UNIVAC 1108 CPU time, only a few iterations were used.

5.1.3 Limitations of Previous Results

The first limited trial of the maximum likelihood identification technique applied to the problem of extracting aircraft stability and control derivatives was on simulated X-22 VTOL data supplied to SCI by Cornell Aeronautical Laboratory. The complete data set consisted of four cases; two without process noise and two with process noise. In each case a single step input was used to generate the data. For the no process noise cases 2A and 2C, satisfactory estimates were obtained for all parameters except for X_δ and Z_δ derivatives. For both the low process noise (2B) and moderate process noise (2D) cases, the errors in all the identified parameters were much larger. However, when a multistep input sequence, which supplied much more excitation, was used, the results for the process noise case improved greatly. It soon became apparent that the quality of the parameter estimates were very "input" dependent.

As outlined in Section 4.3 , the calculation of the update in the parameter estimates involves the computation of the gradient $\frac{\partial J}{\partial p}$, and the information matrix $\frac{\partial^2 J}{\partial p^2}$, where J is the likelihood function (Eq.4.29). These quantities in turn involve solving a differential equation for the sensitivity matrix $\frac{\partial \hat{x}}{\partial p}$ (i/i-1), where $\underline{\hat{x}}$ (i/i-1) is the output of a Kalman filter. Since the equation for the state estimate error covariance, P (i/i-1), does not reach steady state and involves the unknown parameters, the partial derivatives of P (i/i-1), and the Kalman gain W_i, with respect to p should be included in the computation of $\frac{\partial J}{\partial p}$ and $\frac{\partial^2 J}{\partial p^2}$.(See Appendix B) These were neglected in the earlier X-22

identification work. It was orginally thought that the lack of monotonic convergence of the identification algorithm could be attributed to these extra partials being neglected in the computation of the partial derivatives.

As noted earlier the accelerometer measurements introduce process noise into the measurement equation, thereby correlating the total effective measurement noise with the process noise. This correlation which effects the equations of the Kalman filter and, therefore, the computation of the sensitivity matrix, was not accounted for in the earlier application. There is an additional correlation between the g(.) function in the dynamics equation and the g´(.) function in the measurement equation since both are a function of the state estimate, $\hat{x}$ (i/i-1). It was originally thought that this might also have a significant effect on the parameter estimates and on the standard deviations supplied by the Cramer-Rao lower bound.

In both the data supplied by Cornell and generated at SCI the process noise and control were kept constant over an integration step. It was important to distinguish the cases where the process noise changed value before or after a measurement. In one case there would be a correlation between the measurement noise at a sampling point and the process noise during the preceeding integration step while in the other case the ocrrelation would be with the succeeding integration step. This difference, though subtle, is important.

All these areas were investigated with the objective of determining the effects, on the parameter estimates and standard deviations, of different modifications. The following description of the results is broken up into separate sections, each one involving a different area of investigation. Included in an accompanying table are the parameter estimates and standard deviations resulting from each change in the algorithm. These standard deviations are actually lower bounds on the actual values and are obtained from the diagonal elements of the matrix $\left(\frac{\partial^2 J}{\partial p^2}\right)^{-1}$. Also noted, in each case, is the number of iterations of the algorithm used in obtaining the results.

5.1.4 Comparison of Results with Single Step and Multi-Step Input Sequences

The initial processing of the data supplied to SCI by Cornel Aeronautical Labs resulted in unsatisfactory parameter estimates both for the low and moderate process noise cases, as shown in Columns 1 and 2 of Table 5.1. Since it was already known that the input sequence shown in Fig. 5.1, used to generate the Cornell data did not sufficiently

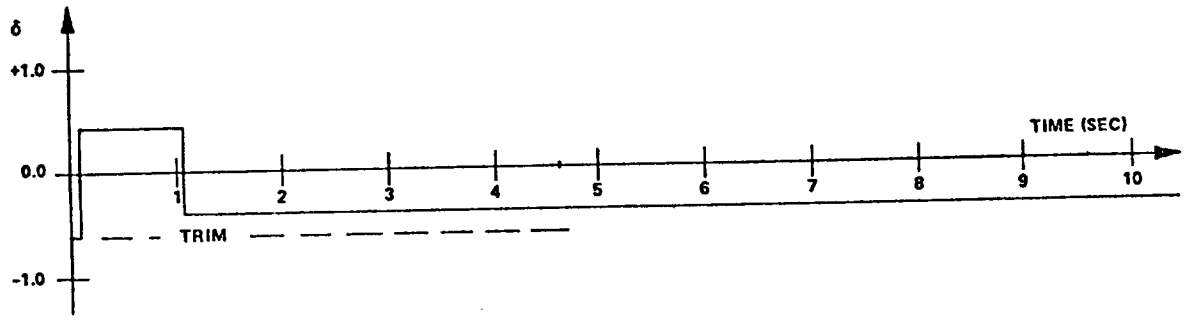


FIGURE 5.1 INPUT SEQUENCE USED IN GENERATING CORNELL DATA

excite all the modes of the system to allow adequate identification and since there was considerable uncertainty already as to how this data was generated, SCI programmed its own data generator using the equations of motion in Appendix A. The elevator deflection sequence used by SCI in generating the X-22 simulated flight data is shown in Fig. 5.2 and the process and measurement noise standard deviations are given in Table 5.2.

TABLE 5.1 X-22 IDENTIFICATION RESULTS

Parameters Cost J	1: Case 2D After Initial Changes 3.3127	2: Case 2B After Initial Changes 3.3745	3: Case 2B Using SCI Multistep 3.5388	4: Case 2B With Forward Correlation 3.5499	5: Case 2B With Constant g 3.5355	6: Case 2B With Constant g and Correlation 3.537
$M_o\begin{pmatrix}1\\u\\u^2\end{pmatrix}$ 28.95 −.176 −.000355	138.3 −1.632 .00429	72.46 −.835 .00211	28.04 −.159 −.000441	30.96 −.206 −.000284	31.47 −.209 −.000292	32.02 −.224 +.000174
$M_w\begin{pmatrix}1\\u\end{pmatrix}$ −.1 −.00317	−1.248 .00558	−.471 .000009	−.0947 −.00321	−.137 −.00274	−.148 −.00268	−.166 −.00235
$M_q\begin{pmatrix}1\\u\end{pmatrix}$ −.497 −.00103	−1.0721 .00434	−.995 .00278	−.489 −.00116	−.520 −.000814	−.500 −.000973	−.436 −.00182
$M_\delta\begin{pmatrix}1\\u\end{pmatrix}$ 18.66 .0669	−11.85 .311	4.45 .179	18.38 .0708	18.17 .0734	18.15 .0733	18.11 .0758
$X_o\begin{pmatrix}1\\u\\u^2\end{pmatrix}$ 18.3 −.0917 −.0003	35.80 −.352 .000659	23.01 −.160 −.0000539	18.88 −.103 −.00024	19.41 −.110 −.000228	17.87 −.0849 −.00033	17.97 −.0874 −.000316
$X_w\begin{pmatrix}1\\u\end{pmatrix}$ .2211 −.00159	.0182 −.000051	.174 −.00122	.220 −.00159	.2085 −.00143	.2233 −.00159	−.223 −.00160
$X_\delta\begin{pmatrix}1\\u\end{pmatrix}$ −.778 .0184	8.586 −.0521	−.976 .0202	−.691 .0171	−1.011 .0212	−.896 .0199	−.875 .0196
$Z_o\begin{pmatrix}1\\u\\u^2\end{pmatrix}$ −32.17 .91 −.007	45.95 −.1367 −.0037	−31.22 .511 .00568	−35.15 .969 −.00728	−27.08 .825 −.00669	−30.19 .879 −.00692	−28.03 .813 −.0066
$Z_w\begin{pmatrix}1\\u\end{pmatrix}$ −.2939 −.00287	−1.320 .00516	−.599 −.000399	−.272 −.00314	−.361 −.00213	−.3397 −.00233	−.356 −.00213
$Z_\delta\begin{pmatrix}1\\u\end{pmatrix}$ −.351 −.0167	20.6 −.122	−11.26 .106	−.673 .0200	−1.143 .0269	−1.136 .02697	−1.11 +.0271
No. of Iterations	3	2	2	2	2	2
Cost J'				.5794	.9158	.9061
Standard Deviations						
$M_o\begin{pmatrix}1\\u\\u^2\end{pmatrix}$	76.18 1.104 .00405	16.47 .254 .000996	2.150 .0352 .000145	2.249 .0371 .000154	2.583 .0418 .000170	3.501 .0573 .000235
$M_w\begin{pmatrix}1\\u\end{pmatrix}$	.695 .00581	.138 .00121	.0251 .000236	.0259 .000244	.0299 .000279	.0399 .000371
$M_q\begin{pmatrix}1\\u\end{pmatrix}$	.467 .00371	.183 .00135	.0139 .000152	.0146 .000159	.0144 .000157	.0157 .000173
$M_\delta\begin{pmatrix}1\\u\end{pmatrix}$	100.4 .7792	18.64 .145	.279 .00306	.288 .00315	.3359 .00361	.3368 .00356
$X_o\begin{pmatrix}1\\u\\u^2\end{pmatrix}$	20.35 .274 .000940	4.93 .0688 .000241	.873 .0132 .00005	.873 .0135 .0000531	.7890 .0127 .0000508	1.010 .0163 .0000660
$X_w\begin{pmatrix}1\\u\end{pmatrix}$	.173 .00142	.0422 .000351	.0105 .0000899	.0103 .0000893	.00919 .0000857	.0116 .000109
$X_\delta\begin{pmatrix}1\\u\end{pmatrix}$	41.04 .3173	7.686 .0596	.1139 .001134	.1132 .00114	.1073 .00116	.124 .00134
$Z_o\begin{pmatrix}1\\u\\u^2\end{pmatrix}$	96.71 1.276 .00428	20.78 .283 .000972	2.631 .0427 .000172	2.755 .04486 .000181	4.667 .0747 .000299	6.533 .1057 .000428
$Z_w\begin{pmatrix}1\\u\end{pmatrix}$	.779 .00657	.165 .00142	.0323 .000321	.0334 .000330	.0545 .000507	.0749 .000696
$Z_\delta\begin{pmatrix}1\\u\end{pmatrix}$	221.0 1.71	41.10 .3189	.388 .00451	.398 .0046	.636 .00688	.692 .00737

TABLE 5.1 CONTINUED

7 : Case 2B With Some Added Partials	8 : Case 2D With Constant g No Added Partials	9 : Case 2D With Constant g and Add Partial
3.5767	3.53684	3.6949
33.58	47.291	76.69
-.261	-.427	-1.049
-.0000359	+.000295	.00345
-.157	-.421	-.583
-.00247	.0000073	.00193
-.502	-.5088	-.321
-.00102	-.000826	-.00345
18.04	15.92	17.79
.0752	.1003	.0807
19.17	16.349	13.50
-.109	-.0593	-.0359
-.000221	-.000449	-.000449
.215	.228	.2764
-.00150	-.00159	-.00191
-.952	-1.402	-.661
.0207	.0261	.0181
-25.88	-14.18	29.94
.796	.636	-.179
-.00656	-.00619	-.00245
-.369	-.634	-.9659
-.00196	.000741	+.00389
-1.273	-4.618	-1.146
.0289	.0703	.0278
2	2	2
.5829	22.8934	23.167
	14.06	
	.2289	
	.000934	
	.1628	
	.00151	
	.0760	
	.000824	
	1.706	
	.0181	
	4.270	
	.0688	
	.000277	
	.0498	
	.000462	
	.544	
	.00581	
	25.27	
	.406	
	.001631	
	.2953	
	.00273	
	3.231	
	.03443	

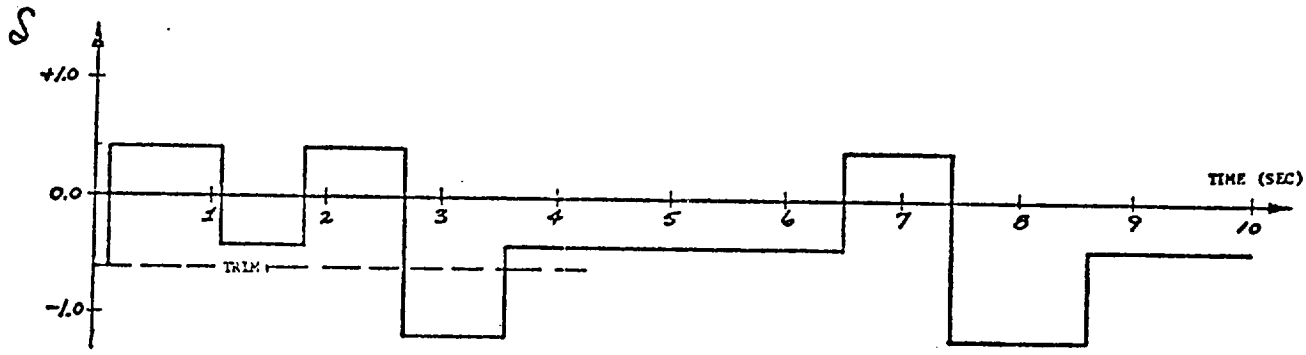


FIGURE 5.2 MULTISTEP INPUT

TABLE 5.2 STANDARD DEVIATION OF PROCESS AND MEASUREMENT NOISE

		Standard Deviation	
		Low	Moderate
gusts (process noise)	v_1	1.0 fps	5.0 fps
	v_2	1.0 fps	5.0 fps
	v_3	.2 deg/sec	1.0 deg/sec
measurement noise	n_u	0.5 fps	2.5 fps
	n_w	0.075 fps	.375 fps
	n_θ	.03 deg	.15 deg
	n_q	.01 deg/sec	.05 deg/sec
	n_{n_x}	.001 g	.005 g
	n_{n_z}	.005 g	.025 g
	$n_{\dot{q}}$	.0025 deg/sec^2	.0125 deg/sec^2

The initial state estimates used in the Kalman filter were set equal to the state measurements observed at the first data point, i.e., $\hat{x}_i(1/1) = z_i$. Since these measurements consist of the true state $\hat{x}_i(1)$ plus noise, the error covariance P (1/1) is given directly as $P(1/1) = E\{(\underline{x}(1) - \underline{\hat{x}}(1/1))(\underline{x}(1) - \underline{\hat{x}}(1/1))^T\} = \text{diag}\{R_{11}, R_{22}, R_{33}, R_{44}\}$ where R_{ii} is the i^{th} diagonal element of the measurement noise covariance matrix, R.

The effect of using the SCI data generation program with a multistep input sequence instead of a single step input sequence can be seen by comparing Cols. 2 and 3 of Table 5.1. The parameter estimates are greatly improved and the standard deviations are reduced. This enhanced ability to identify the parameters is attributed to the fact that the more varied the input sequence, the more the system modes are excited and the higher is the signal-to-noise ratio at the output.

It is important to realize that the results as shown were obtained for only one noise sequence, and therefore, although the parameter estimates improved considerably, the parameter estimates by themselves are not sufficient for comparison. Neither are the costs, themselves, since changes in the noise sequence will influence the costs. As will be seen in the later areas of investigation, some of the parameter estimates improved as the result of some change in the algorithm and some did not. This is almost always the case, and unless the individual relative effect of the parameters on the cost is known, it is very difficult to say on the basis of only a few of the parameter estimates improving, that the algorithm itself is improved by any change. The criterion that is more suitable for comparison is the standard deviation of the parameter estimates. A lower bound on the standard deviations is obtained from the inverse of the information matrix and this is adequate in many cases. However, if the differences in standard deviations are small, the Cramer-Rao lower bound may not reflect these differences.

5.1.5 Comparison of Forward and Backward Correlation

The next area of investigation involved the effects on the parameter estimates and standard deviations of the type of correlation between the input and output noise sequences. The original Cornell data specified that the process noise was kept constant over an integration step. This meant that whatever correlation existed at the measurement times (note that the measurements are taken at discrete instants) also existed throughout the entire integration interval. The SCI multistep data was first generated with the accelerations $\dot{q}$, n_x and n_y being calculated using the process noise from the previous integration step and the control, which was also held constant over an integration step, from the next integration step. Figure 5.3 below graphically shows when the values of v_i, the process noise, and δ_i, the control, were changed in relation to the measurement instances.

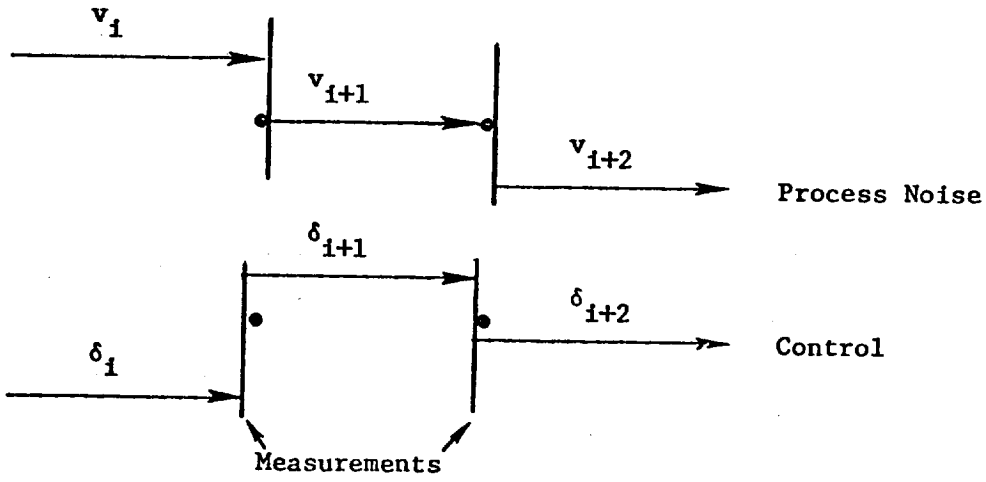


FIGURE 5.3 v AND δ ORIGINALLY USED IN THE CALCULATION OF $\dot{q}$, n_x, n_y
(DOTS INDICATE THE INTEGRATION TIME POINTS)

This seemed inconsistent with the way correlation between process and measurement noise is usually represented in state space models. For example, the discrete analog of the continuous time representation

$$\dot{x}(t) = Fx(t) + G\,v(t)$$

$$y(t) = Hx(t) + n(t)$$

is

$$x_{k+1} = \Phi\, x_k + \Gamma v_k$$

$$y_k = H\, x_k + n_k$$

where Φ,F and G,Γ are related by T, the sampling interval. In this model, v_k and n_k are correlated, i.e. the process noise during (t_k, t_{k+1}) is correlated with the measurement noise at t_k. The correlation between n_k and v_k effects x_{k+1}, not x_k. Similarly, x_{k+1} is calculated using δ_k (the control at time t_k). In continuous time this means that the correlation between $v(t)$ and n_k must be during the interval between t_k and t_{k+1}. In addition, the derivative $\dot{x}(t)$ at time t_k must be calculated using the control that existed between t_{k-1} and t_k. Therefore, $\dot{q}$, n_x and n_y should be calculated using the process noise that will exist in the next integration step and the control that existed in the previous integration step, as in Fig. 5.4.

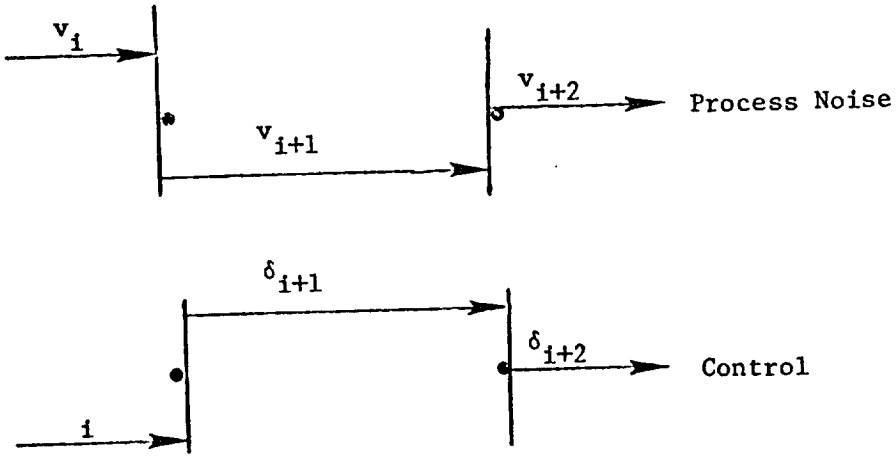


FIGURE 5.4 v and δ USED IN CALCULATING OF $\dot{q}$, n_x, n_y AFTER CHANGE

Note that the correlation between measurement and process noise has changed from a backward correlation to a forward correlation, i.e. the measurement noise at the sampling instant is now correlated with the process noise during the succeeding integration interval. This forward correlation, if unmodeled with effect the (forward) Kalman filter operation while the backward correlation will effect a second (backwards), smoothing run through the data. In an actual application with continuous time dynamics and measurements, this problem will not arise. However, if discrete measurements are recorded, the type of correlation that exists will be an issue, and for data generated by a physical system the forward correlation is the more natural.

The effect of using the forward correlation in the data generation and then identifying the parameters, although not accounting for this forward correlation in the Kalman filter, is shown in Cols. 3 and 4 of Table 5.1 (for the low process noise case, 2B). Almost all the parameter estimates have degraded, offset by an accompanying slight increase in standard deviation. These results indicate that if forward correlation (the type normally used in computer models of discrete systems) is not modeled, it can have a detrimental effect on the quality of the parameter estimates.

5.1.6 Additional Performance Index

In Col. 4 of Table 5.1 a new performance measure is introduced, labelled J'. This represents the unweighted mean square error in estimating the output, given by

$$J' = \frac{1}{2N} \sum_{i=1}^{N} [z_i - h(\hat{x}_{i/i-1}, \hat{p})]^T [z_i - h(\hat{x}_{i/i-1}, \hat{p})]$$

J' was introduced as another means of comparing the results of the different runs and differs from J by the fact that changes in the Kalman filter only effect it through the state estimates $\hat{x}_{i/i-1}$. It was not substituted for J in the identification algorithm since it does not weight

the various state estimates and is not the likelihood function. As was pointed out earlier, the weighting matrix used in J was a function of the Kalman filter covariance and therefore varied as different changes or additions were made in the filter.

Experience with the algorithm and the two costs, J and J', indicated that J' was more sensitive to the parameter estimates than J, which always appeared to be in the range 3.5 to 4. However, this is to be expected since, if the parameter estimates are bad, the state estimates will likewise be bad and the associated state estimate error covariances will be large. The weighted residuals, which are inversely proportional to the state estimate error covariance, may therefore change very little. J', on the other hand, has no weighting, and therefore reflects the absolute accuracy of the state estimates.

5.1.7 Accounting for Correlation Between Process and Measurement Noise

The initial attempts to account for the correlation between the measurement and process noise were not successful either in reducing the cost J or in improving the parameter estimates. It was decided that part of this problem was due to the fact that the noise term $g(x_i)\cdot v_i$ appearing both in the system equations and the measurement equations, depends on the state x_i and gives a long-term correlation. For this reason, it was decided to alter the system equations to include a constant g matrix, calculated from the initial control values, the nominal state values and the actual parameter values.

The identification program was first run without accounting for the process noise-measurement noise correlation, in order to get a new standard for comparison. As shown in Column 5 of Table 5.1, many of the parameter estimates were worsened and the J' cost increased slightly. The fact that the J cost decreased slightly can be attributed to the fact that the weighting matrix is a function of the g matrix. Also the worsened parameter estimates were not,in all cases, offset by increased standard deviations.

The improved accuracy of the parameter estimates with the non-constant g matrix can be attributed to the fact that the parameters of the g matrix account for the system gain factors and are therefore easily identifiable. Since, in the original system model, the same parameters appeared in the f and g matrices, the parameter estimates were overall improved.

It was agreed that, although the constant g assumption was a large change from the original problem, it did not represent a departure from reality. As can be seen from the system equations, the g matrix was originally constructed from the linearized f $(\cdot,\cdot)$ matrix, the motivation being that the process noise would then enter the dynamical equation linearly.

With the constant g assumption, the correlation between process and measurement noise was accounted for by adding the following terms (indicated by $\boxed{\quad}$) to the identification algorithm. (See Section 4.3)

Defining $S_i = E\{v_i^T n_i\}$ to be the measurement noise and process noise correlation at the i^{th} measurement time:

Kalman filter state prediction:

$$\dot{\hat{x}} = f(\hat{x}) + \boxed{g\, S_{i-1}\, R^{-1}[z_{i-1} - h(\hat{x}_{i-1/i-1})]}$$

Kalman filter covariance prediction:

$$P_{i/i-1} = \Phi_{i-1}\, P_{i-1/k-1}\, \Phi_{i-1}^T + g\, Q\, g^T(\Delta T)^2 \boxed{- g\, S_{i-1}\, R^{-1}\, S_{i-1}^T\, g^T(\Delta T)^2}$$

$$\text{where } \Phi_{i-1} = I + \left.\frac{\partial f}{\partial x}\right|_{x=\hat{x}_{i-1/i-1}} \Delta T - \boxed{g\, S_{i-1}\, R^{-1} \left[\left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}_{i-1/i-1}} \frac{\partial \hat{x}}{\partial p} + \frac{\partial h}{\partial p}\right]}$$

Sensitivy equations:

$$\frac{d}{dt}\left[\frac{\partial \hat{x}}{\partial p}\right] = \frac{\partial f}{\partial x}\frac{\partial \hat{x}}{\partial p} + \frac{\partial f}{\partial p} - \boxed{g\, S_{i-1}\, R^{-1}\left[\left.\frac{\partial h}{\partial x}\right|_{x=\hat{x}_{i-1/i-1}} \frac{\partial \hat{x}}{\partial p} + \frac{\partial h}{\partial p}\right]} \quad t_{i-1} < t < t_i$$

The resulting parameter estimates and cost are given in Col. 6 of Table 5.1. Comparing these results with those of Col. 5 it is seen that, while some parameter estimates improved, others did not. The overall cost J remained the same, while J' decreased slightly. An important point is that the standard deviations for almost all the parameters increased. This implies that, with the initial set of least squares parameter estimates as good as they are, the inclusion of the terms accounting for the input/output noise correlation does not gain much by way of the parameter estimates. However, with these correlation terms included, the standard deviations come out to be more realistic, in view of the differences between the actual and estimated parameter values. All this is not to say, however, that for a less exact set of initial parameter estimates, the correlation terms won't improve the algorithm performance.

5.1.8 Inclusion of Additional Partial Derivatives

There were two principle motives for adding the additional first order partial derivative terms of the state covariance matrix to the identification algorithm. The first was that they would be required if Q and R were to be identified, since they both appear explicitly in the equations for the state estimate error covariance. The second motive was that the cost J, instead of monotonically decreasing with each iteration, was oscillating.

A possible cause of this was that the gradient direction was being calculated incorrectly due to the fact that the neglected partial derivatives have an appreciable effect on the gradient of the likelihood function in the vicinity of the minimizing set of parameter estimates.* It was necessary, therefore, to investigate the importance of these extra partial terms in identifying the system paramters.

The inclusion of the additional first order partial derivatives into the identification algorithm presented special problems due to the nature of the X-22 model. Since the g matrix is a function of both the states and of the parameters, the derivative of the $g\,Q\,g^T$ term (appearing in the covariance equations) with respect to p must be included. This is a particularly lengthy computation. A first attempt to include all the additional partials except those of the $g\,Q\,g^T$ term is shown in Col. 7 of Table 5.1. Comparing this with Col. 4 (since constant g was not assumed), it is seen that both costs J and J' increased slightly and some of the parameter estimates, themselves, are slightly degraded.

Two changes were then decided upon. The first was that case 2D (moderate process noise) would be used instead of 2B (low process noise). The second was that the g would again be modeled as constant. The first change was motivated by the desire to see larger variations in the costs. More process noise would make the initial least squares parameter estimates worse and therefore the effect of the additional partials would, potentially, be the greatest. The second change was motivated by the fact that with the constant g assumption, partials of the $g\,Q\,g^T$ terms are identically zero.

---

* A second possible cause was too large a step size, which was corrected by halving the step size along the calculated gradient direction whenever the cost J, increased.

Column 8 of Table 5.1 gives the parameter estimates, cost and standard deviations for the 2D case without the added partial terms, and with the constant g assumptions. Column 9 gives the same results with the additional partials included in the algorithm. Once again the cost, J, increased slightly ($\approx$ 5%). This can be attributed to the fact that the convergence of the algorithm with the added partial terms may be slower. The standard deviations of these parameter estimates (for both Cols. 7 and 9) are not given since their calculation requires another full iteration of the algorithm. Since the added partials quadruple the run time per iteration, it was decided not to calculate these values. The important point is that, considering the slight variation in J, even for this worst case, and the possible benefits of the added partials in terms of the vastly increased run time, it is not necessary to include these added partials in the identification algorithm.

5.1.9 Aerodynamic Derivative Estimates

As was noted at the beginning of this section, the aerodynamic derivatives themselves were not identified. Rather, the coeffients of first or second order polynomial expansions in the longitudinal velocity, u, of these derivatives were identified. Using these identified coefficients, it was then possible to reconstruct the time histories of the total derivatives, and compare these estimates with the actual values. These comparisons are shown in Fig. 5.5 for the 2B data and the original model structure. The fits to most of the derivatives was good. This indicated that, although some of the estimates of the polynomial coefficients had relatively large uncertainties,their influence on determining the total derivative behavior was small.

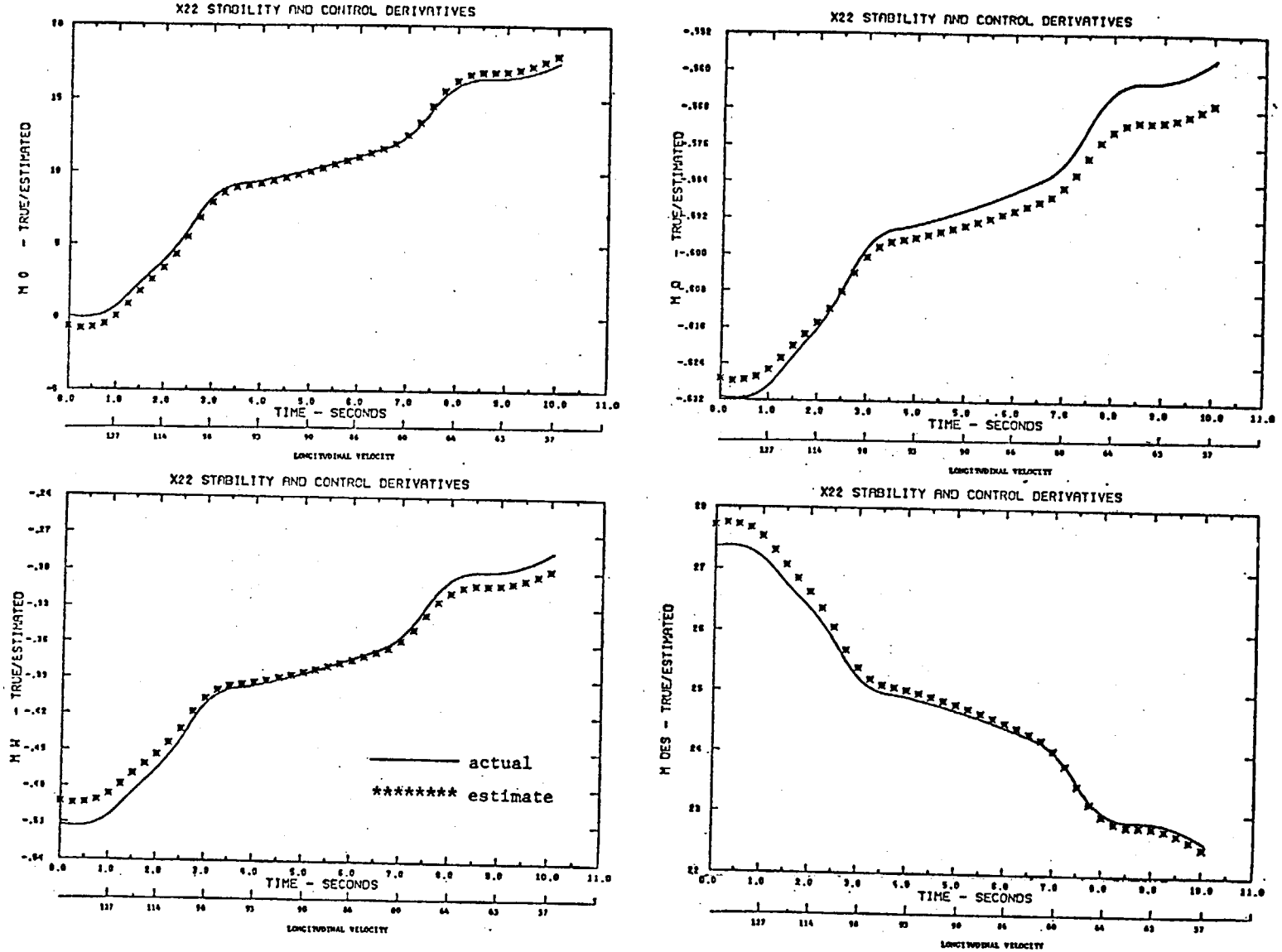


FIGURE 5.5 X-22 ESTIMATED AND ACTUAL (SIMULATED) STABILITY AND CONTROL DERIVATIVE TIME HISTORIES

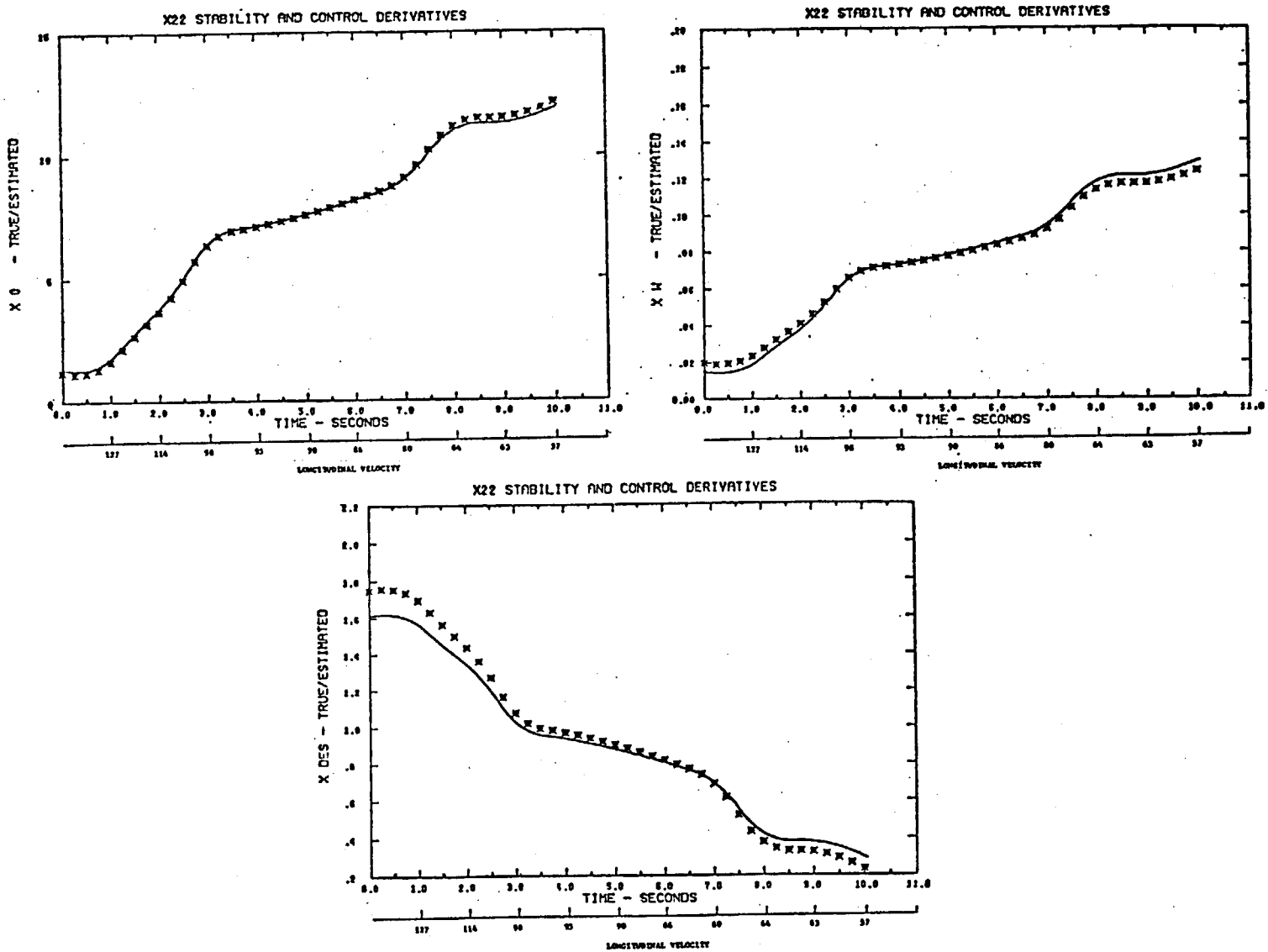


FIGURE 5.5 (CONT.) X-22 ESTIMATED AND ACTUAL (SIMULATED) STABILITY AND CONTROL DERIVATIVE TIME HISTORIES

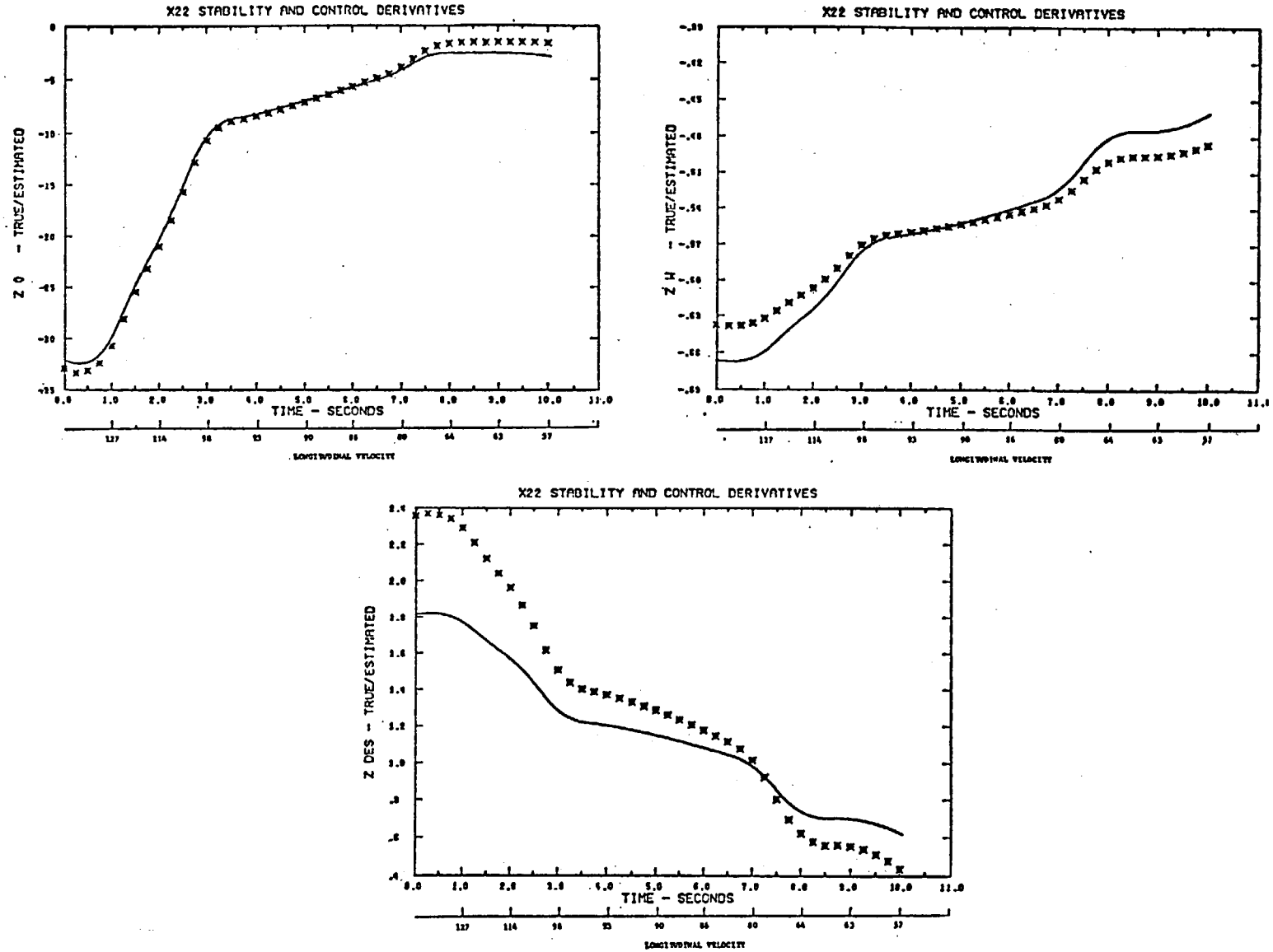


FIGURE 5.5 (CONT.) X-22 ESTIMATED AND ACTUAL (SIMULATED) STABILITY AND CONTROL DERIVATIVE TIME HISTORIES

5.2 HL-10 Flight Test Data

This data was used mainly for checking and validating the maximum likelihood program. Flight data for the HL-10 lifting body was supplied to SCI by NASA FRC (Flight Research Center) Edwards along with as much information about the flight condition as was available at the time. FRC also supplied SCI the results of their stability and control derivative extraction program along with the specific measurement and a priori parameter weights that were used. The HL-10 data did not contain gust effects and a linear model for the lateral dynamics was used.

It should be noted for the case of unknown measurement and process noise covariances an additional term, $N\ell n(\det(HP_{ss}H^t + R))$ is added to the cost criterion, where P_{ss} is the steady-state Kalman filter error covariance matrix. This is because the weighted mean square error will always have a constant value of $\frac{N}{2}$ x (no. of states) since the weighting matrix is the sample covariance.

5.2.1 Dynamical Equations of Motion and Observation Equations

The linearized lateral equations of motion, including the effect of the wind gusts, are:

$$C\underbrace{\begin{bmatrix}\dot{p}\\ \dot{r}\\ \dot{\beta}\\ \dot{\phi}\end{bmatrix}}_{\dot{x}} = \underbrace{\begin{bmatrix} L_p & L_r & L_\beta & 0\\ N_p & N_r & N_\beta & 0\\ Y_p+\sin\alpha & Y_r-\cos\alpha & Y_\beta & \frac{g\cos\theta}{V}\\ 1 & \tan\theta & 0 & 0\end{bmatrix}}_{F}\underbrace{\begin{bmatrix}p\\ r\\ \beta\\ \phi\end{bmatrix}}_{x}$$

$$+\underbrace{\begin{bmatrix} L_{\delta_a} & L_{\delta_r} & L_o\\ N_{\delta_a} & N_{\delta_r} & N_o\\ Y_{\delta_a} & Y_{\delta_r} & Y_o\\ 0 & 0 & 0\end{bmatrix}}_{G}\underbrace{\begin{bmatrix}\delta_a\\ \delta_r\\ 1\end{bmatrix}}_{u} + \underbrace{\begin{bmatrix}L_\beta\\ N_\beta\\ Y_\beta\\ 0\end{bmatrix}}_{\Gamma}\beta_n \qquad (5.1)$$

where p is roll rate (°/sec)*
r is yaw rate (°/sec)
β is sideslip angle (°)
ϕ is roll angle (°)
δ_a is aileron deflection (°)
δ_r is rudder deflection (°)
β_n is wind gust in equivalent sideslip angle (°)
C is a transformation matrix

$$C = \begin{bmatrix} 1 & -\frac{I_{xz}}{I_x} & 0 & 0\\ -\frac{I_{xz}}{I_z} & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$$

* All the quantities are in the body axes system.

For both the HL-10 and the M2/F3 flight data, (which is discussed in Section 5.3), α , the angle of attack, γ, the flight path angle, and V, the velocity, were assumed constant over the data record and their values were supplied. The same was true of I_{xz}, I_x and I_z. In most cases, Y_p and Y_r were assumed to be zero, which implied that there were nominally 20 parameters to identify (7 in F, 9 in G and 4 initial conditions), excluding the wind gusts and biases in measurements.

The observation equations, again assuming the existence of gusts, are given below:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & Y_\beta & 0 & 0 \end{bmatrix}}_{H} \begin{bmatrix} p \\ r \\ \beta \\ \phi \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ Y_{\delta_a} & Y_{\delta_r} & Y_o \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} \delta_a \\ \delta_r \\ 1 \end{bmatrix}}_{u}$$

$$+ \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ Y_\beta \end{bmatrix}}_{\Gamma} \beta_n + \underbrace{\begin{bmatrix} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \end{bmatrix}}_{\underline{n}} \tag{5.2}$$

where y_5 is the lateral acceleration

n_i, $i = 1,\ldots,5$ are independent white noise Gaussian measurement errors.

Notice that β_n always appears with β when gusts are assumed to exist.

In the processing of the HL-10 data, $\beta_n=0$, thereby leaving the measurement noise, $\underline{n}$, as the only random noise source. The Kalman filter equations reduce to the original state equations and the Maximum Likelihood Method is essentially similar to the Output Error Method with the exception of the weighting matrix.

5.2.2 Characteristics of HL-10, Flight 19-2

The HL-10 data was assumed to contain no gusts and was, therefore, processed by using the generalized output error criterion. The results, however, were somewhat unexpected and, in all, eight different runs, each with a variation on the original output error run, were made to solve the problems which were encountered. It was later learned that these same problems had been encountered by FRC. Each of the different runs are described in this section, along with the objectives, observations and conclusions particular to each.

Along with the observations and control time histories (see Fig. 5.6) wind tunnel derived parameter estimates were supplied to SCI by FRC. There were 327 data points at a sampling rate of 50 per second, for a total of 6.54 seconds of data. The angle of attack of this flight was 16.8^{o} and the mach number was 1.22.

It was evident from the data that a substantial amount of clipping and quantization had occurred during data collection. An accurate model for the observations would include the dynamics of the instrumentation system, but no information was available. One effect of not including a model of the instrumentation dynamics and quantization effects would be to have correlated nongaussian residuals, in each of the observations.

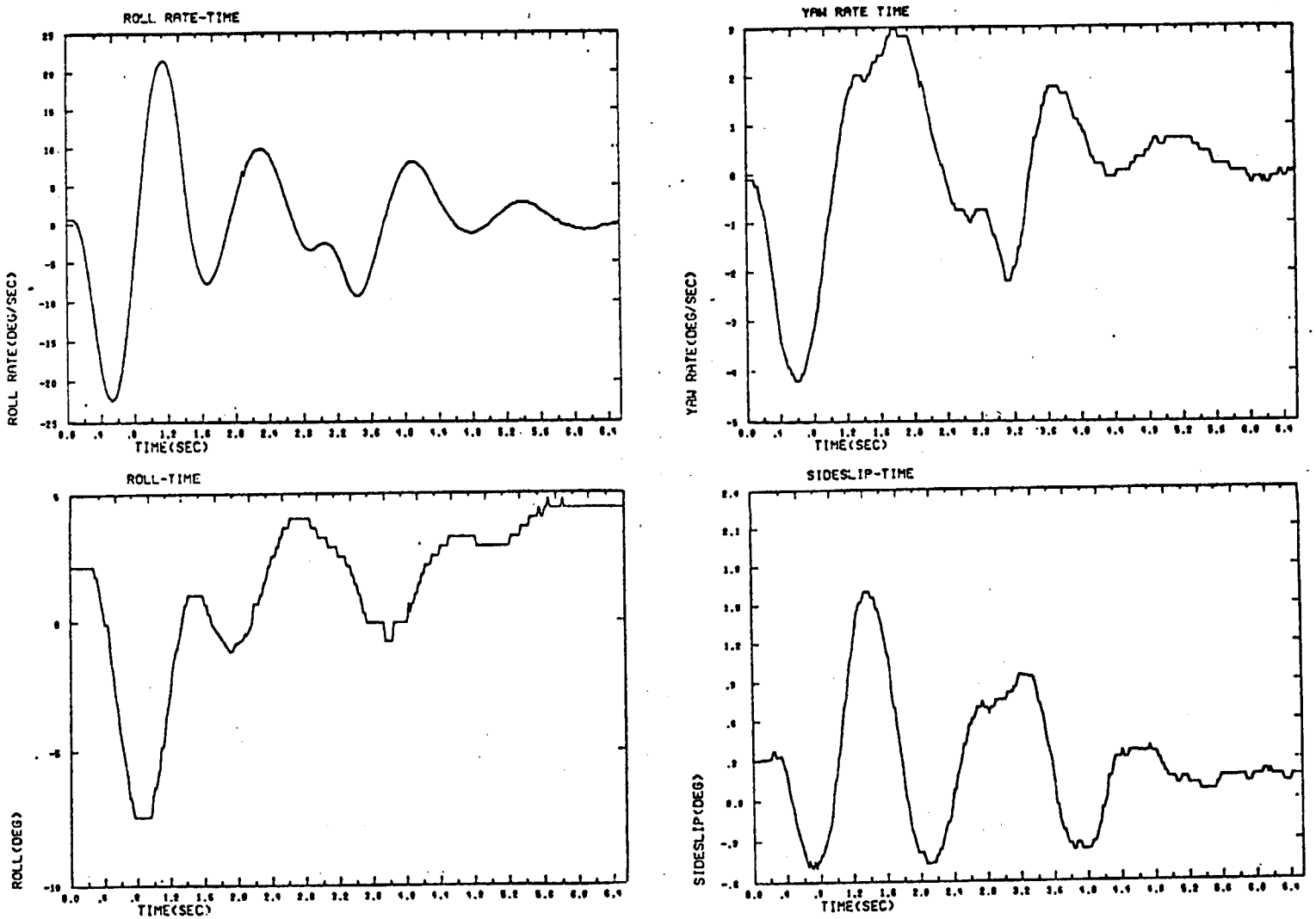


FIGURE 5.6 HL-10 OBSERVED DATA AND CONTROL SEQUENCE TIME HISTORIES

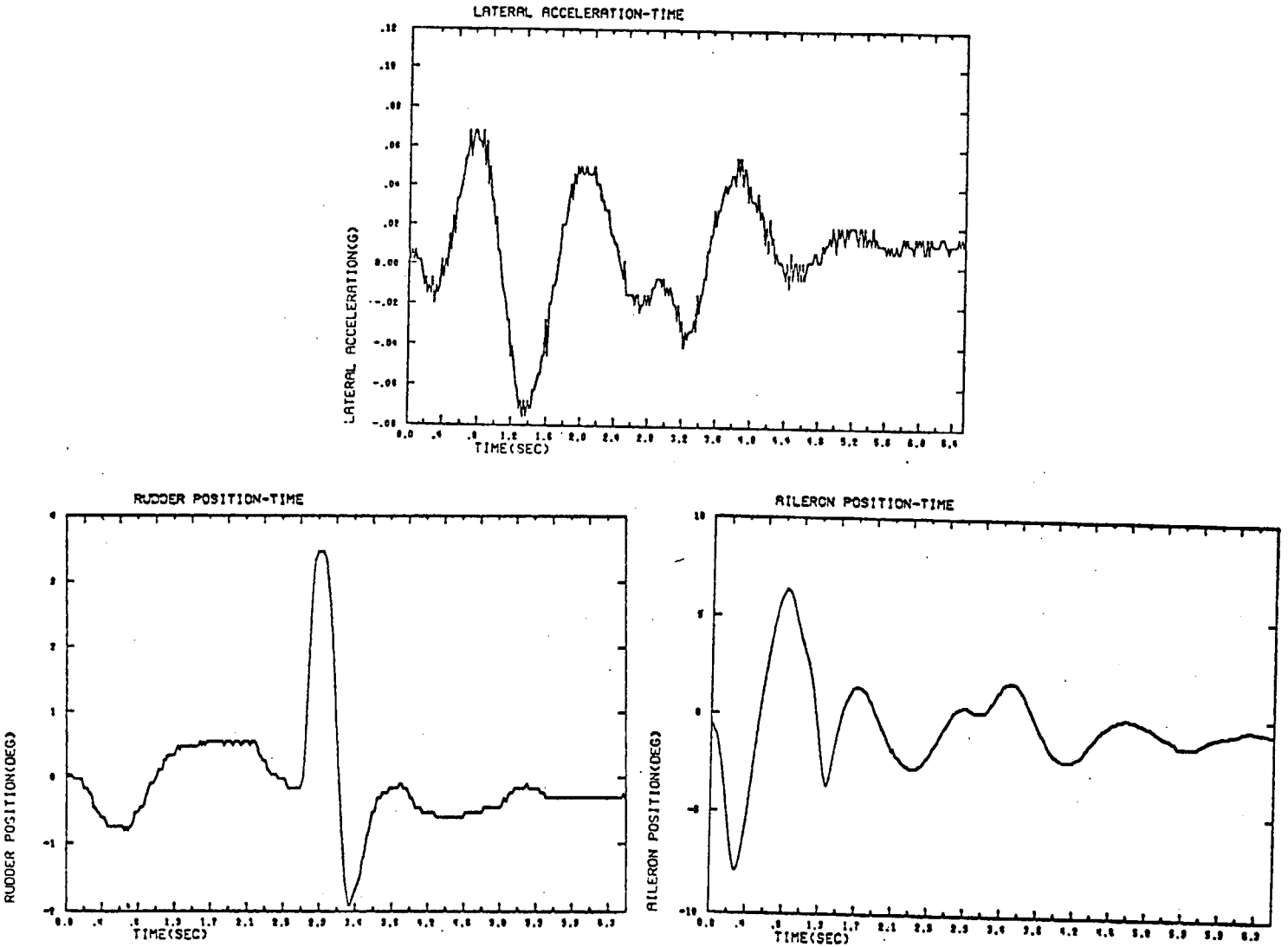


FIGURE 5.6 (CONT.) HL-10 OBSERVED DATA AND CONTROL SEQUENCE TIME HISTORIES

5.2.3 Results of Flight 19-2

The first processing of the HL-10 data was with the maximum likelihood identification program in an output error mode. As such, it differed from the FRC Newton-Raphson identification program in only one way, the weightings of the fit errors. FRC uses a constant weighting matrix, based on an idea of the instrumentation accuracies. The SCI approach estimates the measurement noise covariance matrix and weights the fit errors by the inverse of this matrix. It is shown in Section 4.4 that the ML estimate of the measurement noise covariance matrix is given by the sample covariance of the observations. Edwards also uses a priori weighting on the difference between the parameter estimates and the wind tunnel estimates, which was not initially included in the SCI maximum likelihood program.

The time histories of the five response variables along with the estimated values obtained after 11 iterations of the data are given in Figure 5.7. Included also in Figure 5.8 is the fit error in the p and r observations. A comparison of the parameter values themselves and the wind tunnel values is given in Table 5.3 along with the associates standard deviations in the parameter estimates. The parameter values obtained from the FRC output error method with fixed weights after 7 iterations and a priori weighting are also shown in Table 5.3, along with the associated confidence bounds.

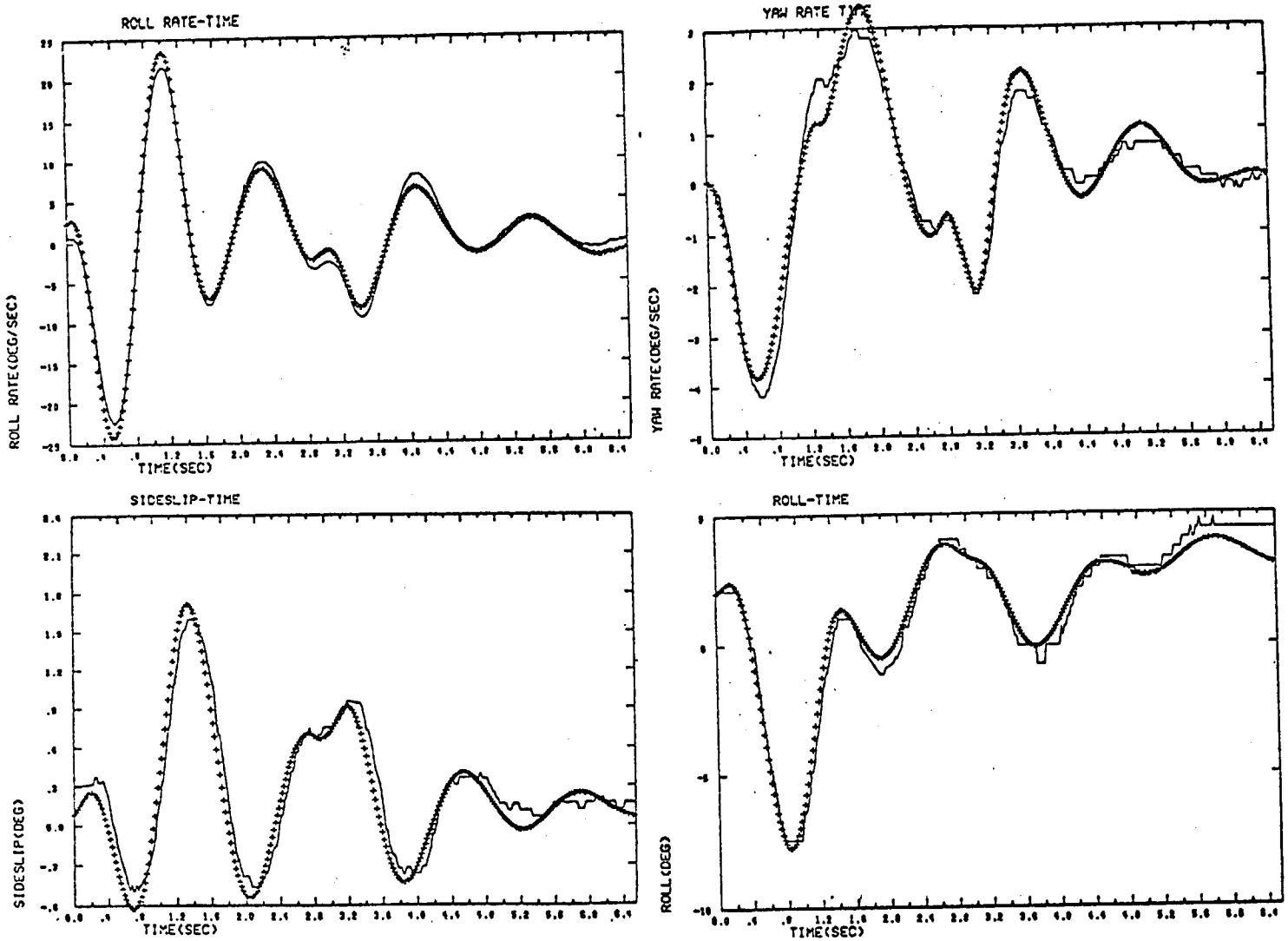
As Figure 5.7 shows, the fits in all the observations were very good. However, as is often the case, the fit error alone does not indicate an acceptable set of parameter values. The two major problems that appeared were that (1) the signs of the L_p, L_r, N_p and N_r derivatives had all changed from those of the wind tunnel values* and (2) the fit error in

* In this investigation, it was assumed that the wind-tunnel and theoretical values had correct signs. This may not necessarily be the case for lifting bodies flown at transonic speeds due to limitations of wind-tunnel testing and theoretical calculations. The question of how much confidence can really be placed in these values has not been resolved.

TABLE 5.3 HL-10 PARAMETER ESTIMATES AND STANDARD DEVIATIONS

Parameter	Wind Tunnel and Theoretical Value	FRC Values with a priori weighting (with confidence bounds)	Max. Lik. Estimates (with St'd. dev.)	Max. Lik. Estimates with Y_p and Y_r (with St'd dev.)	Max. Lik. Estimates With a priori Weighting (with St'd dev.)	Max. Lik. Estimates With a priori weighting and Bias (with St'd dev.)
L_p	-0.3435	-0.3436 (0.196)	0.915 (0.025)	0.395 (0.022)	-0.295 (0.0114)	-0.271 (0.0114)
L_r	0.2723	1.188 (0.0196)	-1.363 (0.138)	0.0671 (0.111)	1.574 (0.0728)	1.349 (0.0747)
L_β	-30.75	-52.073 (1.18)	-56.489 (0.305)	-46.94 (0.417)	-47.179 (0.285)	-50.32 (0.330)
N_p	0.0245	0.0326 (0.0157)	-0.160 (0.00429)	-0.187 (0.00370)	0.0380 (0.00352)	0.0550 (0.00323)
N_r	-0.1290	-0.1114 (0.0157)	0.432 (0.0187)	0.548 (0.0152)	-0.111 (0.0114)	-0.0896 (0.0108)
N_β	6.8411	7.0496 (0.118)	8.523 (0.719)	7.474 (0.0978)	6.292 (0.0574)	6.845 (0.0570)
Y_p	-0.0617	-0.0584 (0.00392)	—	0.329* (0.00307)	0.335* (0.00192)	0.310 (0.00192)
Y_r	-0.0120	-0.0122 (0)	—	-1.091* (0.0122)	-1.064* (0.00754)	-1.019 (0.00762)
Y_β	-0.0916	-0.0855 (0.00392)	-1.471 (0.0202)	-0.1458 (0.0161)	-0.0949 (0.00383)	-0.0918 (0.00382)
$L_{\delta a}$	11.2464	11.996 (1.18)	12.415 (0.0731)	12.494 (0.0775)	12.124 (0.0586)	12.282 (0.631)
$L_{\delta a}$	5.665	5.877 (1.18)	6.288 (0.173)	6.544 (0.148)	6.252 (0.111)	6.429 (0.114)
L_o	-	-	21.03 (0.302)	17.341 (0.222)	0.247 (0.117)	-0.0137 (0.0951)
$N_{\delta a}$	0.8135	1.456 (0.118)	1.262 (0.0136)	1.245 (0.0201)	1.404 (0.0131)	1.357 (0.123)
$N_{\delta r}$	-3.617	-3.178 (0.118)	-3.186 (0.033)	-3.313 (0.0326)	-3.257 (0.0229)	-3.194 (0.0212)
N_o	-	- -	-1.633 (0.0479)	-1.206 (0.0401)	1.561 (0.0205)	1.562 (0.0163)
$Y_{\delta o}$	-0.00180	-0.0018 (0)	0.0623 (0.00390)	.0515 (0.00373	-0.0231 (0.00435)	-0.0482 (0.00443)
$Y_{\delta r}$	0.0111	-0.00427 (0)	0.919 (0.00992)	0.0629 (0.00895)	0.0717 (0.00782)	0.519 (0.00768)
Y_o	-		0.513 (0.00898)	0.412 (0.00778)	-0.751 (0.0112)	-0.639 (0.00829)
$P_{initial}$			1.76 (0.158)	3.325 (0.146)	2.225 (0.116)	2.348 (0.117)
$R_{initial}$			0.117 (0.322)	0.0251 (0.0263)	-0.710 (0.021)	-0.599 (0.0201)
$\beta_{initial}$			0.398 (0.0188)	0.1305 (0.0152)	-0.226 (0.0125)	-0.219 (0.0111)
$\phi_{initial}$			1.939 (0.0697)	1.767 (0.0511)	-1.748 (0.0663)	-1.429 (0.0571)
Likelihood Function Value			-2243	-2359	-1552	-2264

*The identified quantities are $Y_p + \sin\alpha$, $Y_r - \cos\alpha$



——— OBSERVED DATA

+ + + ESTIMATE

Fig. 5.7 HL-10 OBSERVED DATA AND ESTIMATES:
OUTPUT ERROR

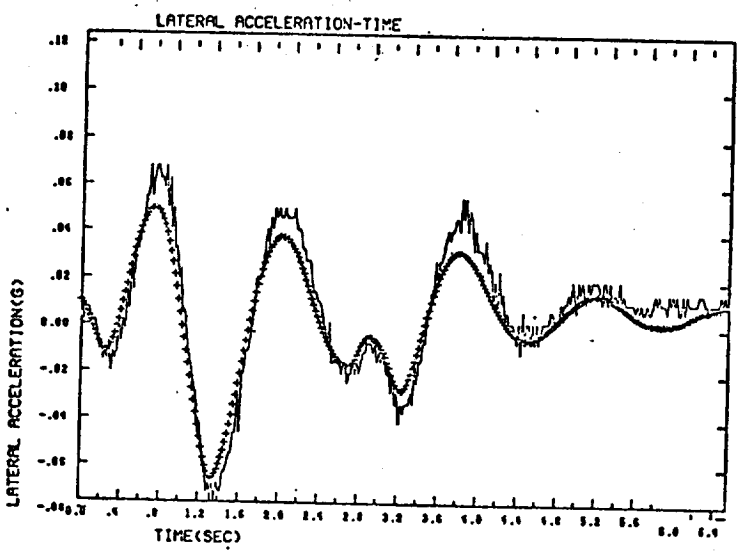


Fig. 5.7 HL-10 OBSERVED DATA AND ESTIMATES - OUTPUT ERROR

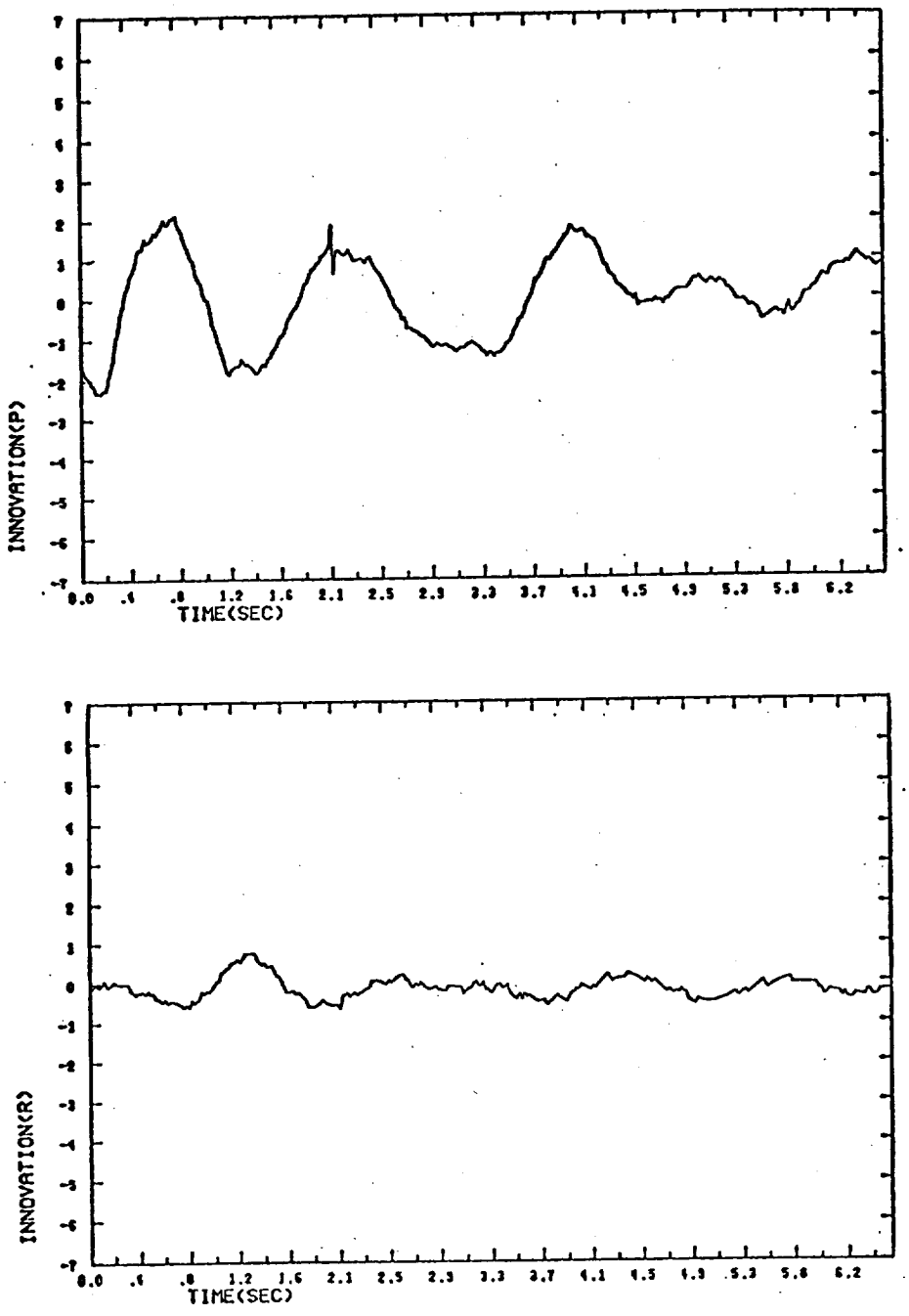


Fig. 5.8 HL-10 FIT ERRORS IN p AND r MEASUREMENTS - OUTPUT ERROR

the observations (although only p and r are shown) exhibited a sinusoidal characteristic. Although the precise reasons for these problems are not known, there are several contributory factors.

The first factor is that the parameters with the opposite signs are weak parameters -- that is, they cannot be accurately identifed from flight test data. This is the approach FRC has taken with the L_r and N_p parameters and in the a priori weighting, they weighed their wind tunnel values very strongly. There is also an identifiability problem because the stability augmentation system (SAS) was used on this flight. Such a system would tend to suppress certain modes of the system, while emphasizing others. The parameters of the suppressed mode are, therefore, hard to identify.

A second factor could be that the linearized dynamics are not accurate enough for the flight conditions of this data. Also, there may be coupling between the longitudinal and lateral modes, which is not included in the model.

A third factor, which may account for the sinusoidal characteristic in the fit error, is that due to the instrumentation dynamics, the measurement noise is actually correlated. This hypothesis could be verified by reprocessing the data with the measurement noise modeled by a second order linear system. Final verification of these possibilities would have to be based on processing additional data under similar flight conditions.

In many instances, the solution to such problems is to adjoin to the likelihood function other measures of performance, usually indicating some a priori knowledge of the parameter values. Several examples of these are the a priori weighting and constrained parameter values which are discussed later. A less direct approach, of reducing the dimensionality of the parameter space, is also presented.

5.2.4 Output-Error with Y_p and Y_r Identified

The remaining series of runs were all aimed at solving the problems encountered with the first processing of the HL-10 data. First, the Y_p and Y_r derivatives were considered as two additional parameters to be identified.

In the previous run, both Y_p and Y_r were considered zero. However, by examining the equations of motion, $\ddot{p}$ can be expressed as a function of $(Y_p + \sin \alpha)$ and $\ddot{r}$ can be expressed as a function of $(Y_r - \cos \alpha)r$. This would introduce previously neglected second order effects into the estimate of p and r, and possibly account for the sinusoidal characteristic in the fit error.

The results indicated that the fit in each of the observations was about the same as in the straight output error case although, as shown in Table 5.2, L_r does have the same sign as the wind tunnel value and L_p is less positive. However, on the other hand, both N_p and N_r are worse. In addition, the sinusoidal characteristic of the fit error remained, diminished only slightly.

5.2.5 Output Error With Constrained Parameter Values

Since it seemed clear that the opposite signs on the four parameters were a result of trying to minimize the fit error, and until additional

terms were added to the model, these signs were likely to remain opposite, the next run constrained the values of L_p and N_r to remain negative. This would answer the questions of whether there was a set of parameter values which would minimize the cost criterion (although not globally) with the indicated parameters having the same sign as the wind tunnel values. If these values remained on the constraints, no such minimizing set exists.

The results of this run were that the L_p and N_r values remained on the constraints and the L_r and N_p values again had the opposite signs. In addition, the fit in the observations was drastically degraded. Only the fit on r was of equal quality as in the two previous trials.

5.2.6 Output Error With Different Initial Conditions

One remaining possible cause of the changed values could be the presence of local minima having the opposite signs on L_p, L_r, N_p and N_r. The next run used initial parameter estimates of L_p and N_r which were more negative than the wind tunnel values, while keeping the control derivatives the same. The results of this run were that the signs of all four parameters (L_p, L_r, B_p, N_r) were again reversed and, if more iterations had been performed, the final parameter values would have, more than likely, been equal to the initial set of output error values.

Although the results from many sets of initial parameter values would be necessary to conclusively determine if the values from the initial output were truly the global minimum, it appears that this might be the case. If the signs on the four parameters are to be the same as the wind tunnel values, an additional cost must be put on the difference between the parameter estimates and the wind tunnel values. This is precisely the reason for the "a priori weighting" mentioned earlier.

5.2.7 Output Error With A Priori Weighting

The values for the parameter weights used in this run were obtained

directly from the FRC's runs supplied to SCI. Figure 5.9 shows the time histories of the observations and the resulting estimates (except for ϕ) for this weighting. It is clear that the fits to the observations, except for p, have been degraded. However, as shown in Table 5.3, the values of the four indicated parameters have the same sign as the wind tunnel values.

It was found, however, that except for p and r there was appreciable bias in the fits to the observed data. Not accounting for this bias in the computation of the measurement noise covariance will cause incorrect weights to be assinged to the different observation residuals. This will effect both the computation of the gradient and the information matrix, resulting in incorrect parameter step sizes. Another run was made with the sample bias of the observation residuals computed, at each iteration, and accounted for in the sample covariance calculation. The fits to the observed data for this second processing of the data with a priori weighting are shown in Figure 5.10. The fits to r, β and ϕ are much improved over the previous case. Only the fit to the lateral acceleration data has not improved. As shown in Table 5.3, many of the parameter estimates for the run are closer to the original wind tunnel values than for the previous run, without considering the biases. It, therefore, appears that when using a priori weighting, consideration must be given to the possibility of having biased residuals, which must be used in computing the sample covariance.

A final processing of the data with a priori weighting and including the identification of the output biases was made with the additional feature of retaining only the diagonal elements of the sample covariance for the estimation of the instruction noise variances. All the off-diagonal terms were set to zero. The rationale for this was that each of the measuring instruments on board the aircraft operate independently and therefore the errors would be uncorrelated. The fit to the observed data did not improve over the previous run and many of the parameters were now

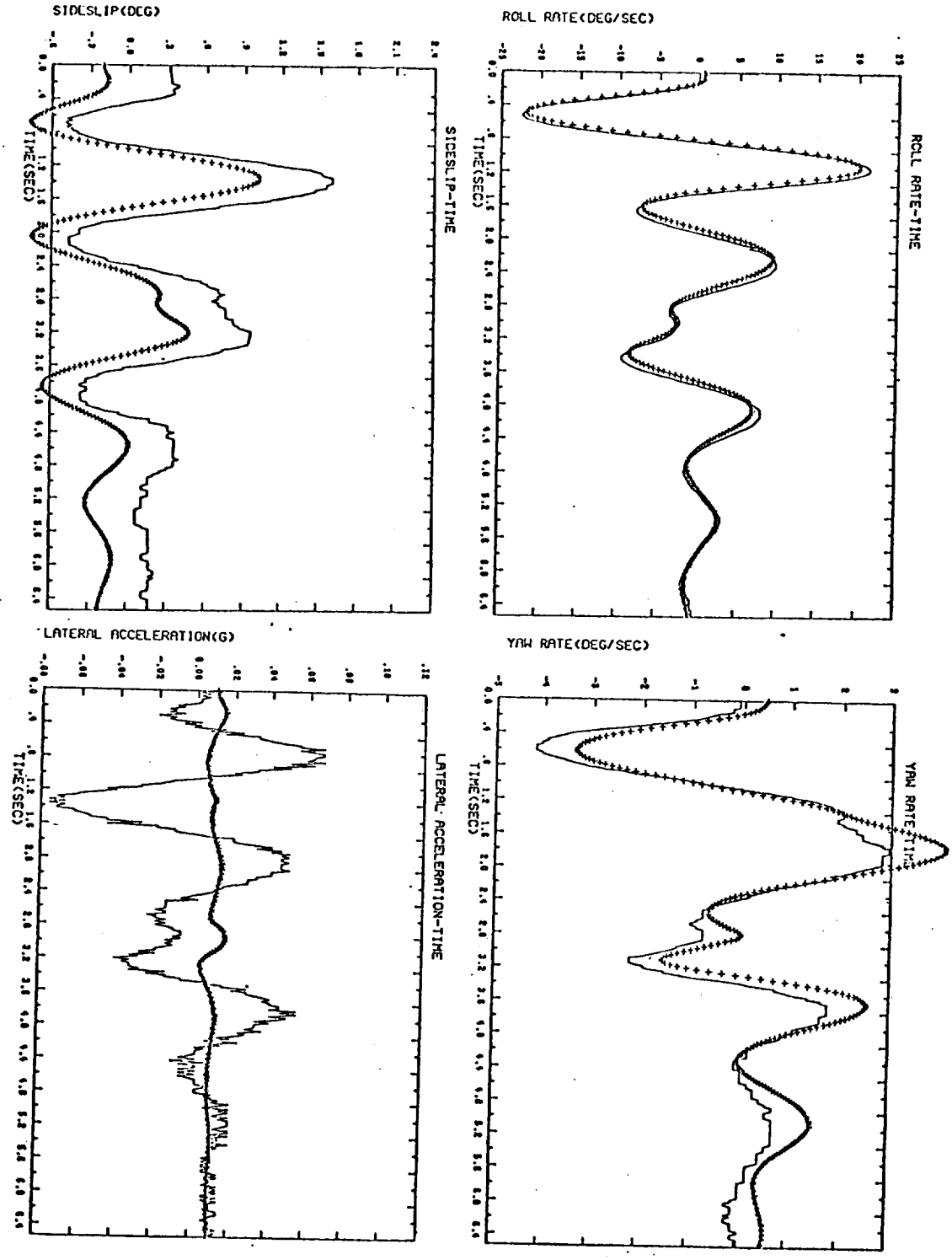


Fig. 5.9 HL-10 OBSERVED DATA AND ESTIMATES: OUTPUT ERROR WITH A PRIORI WEIGHTING

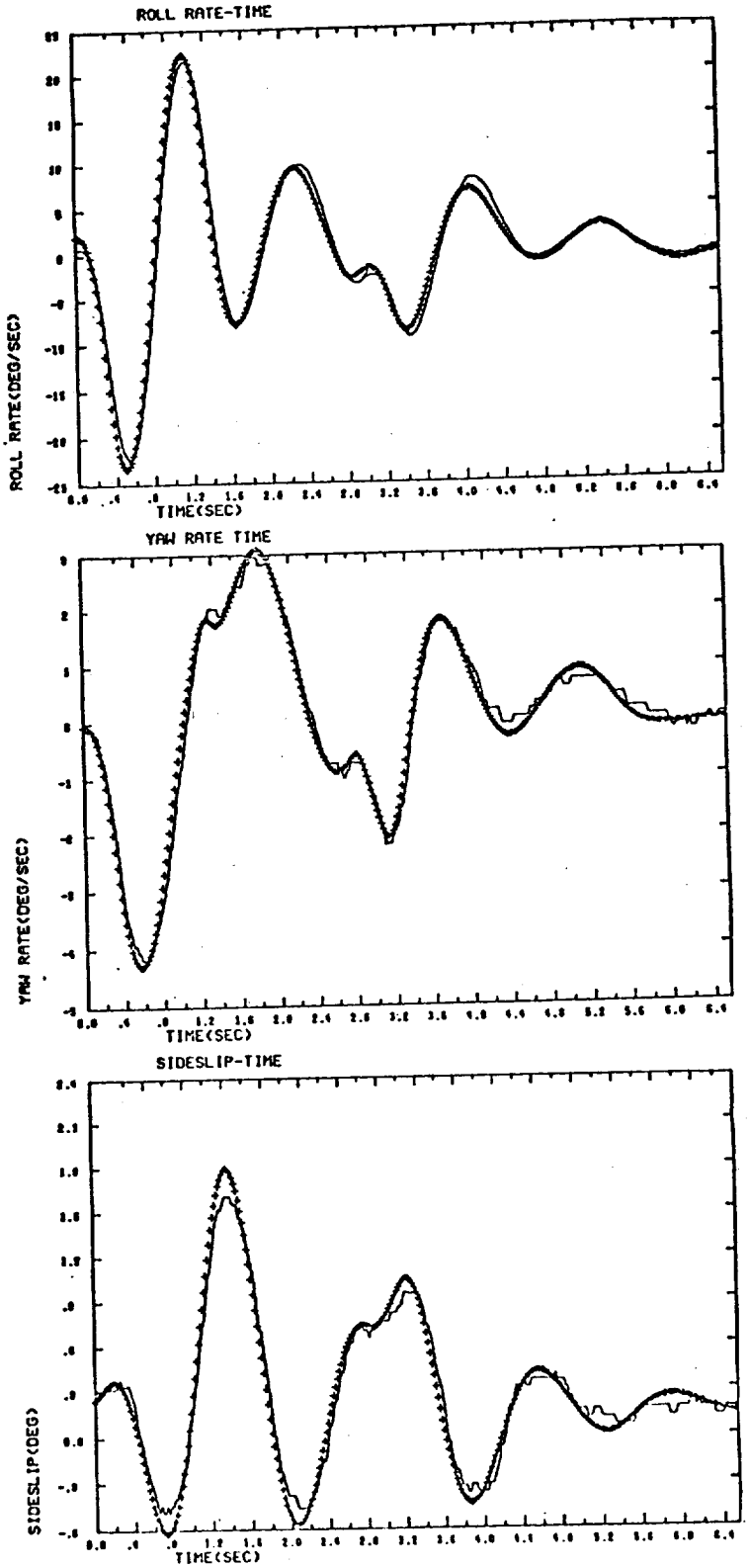


Fig. 5.10 HL-10 : OUPUT ERROR WITH A PRIORI WEIGHTING AND BIASES

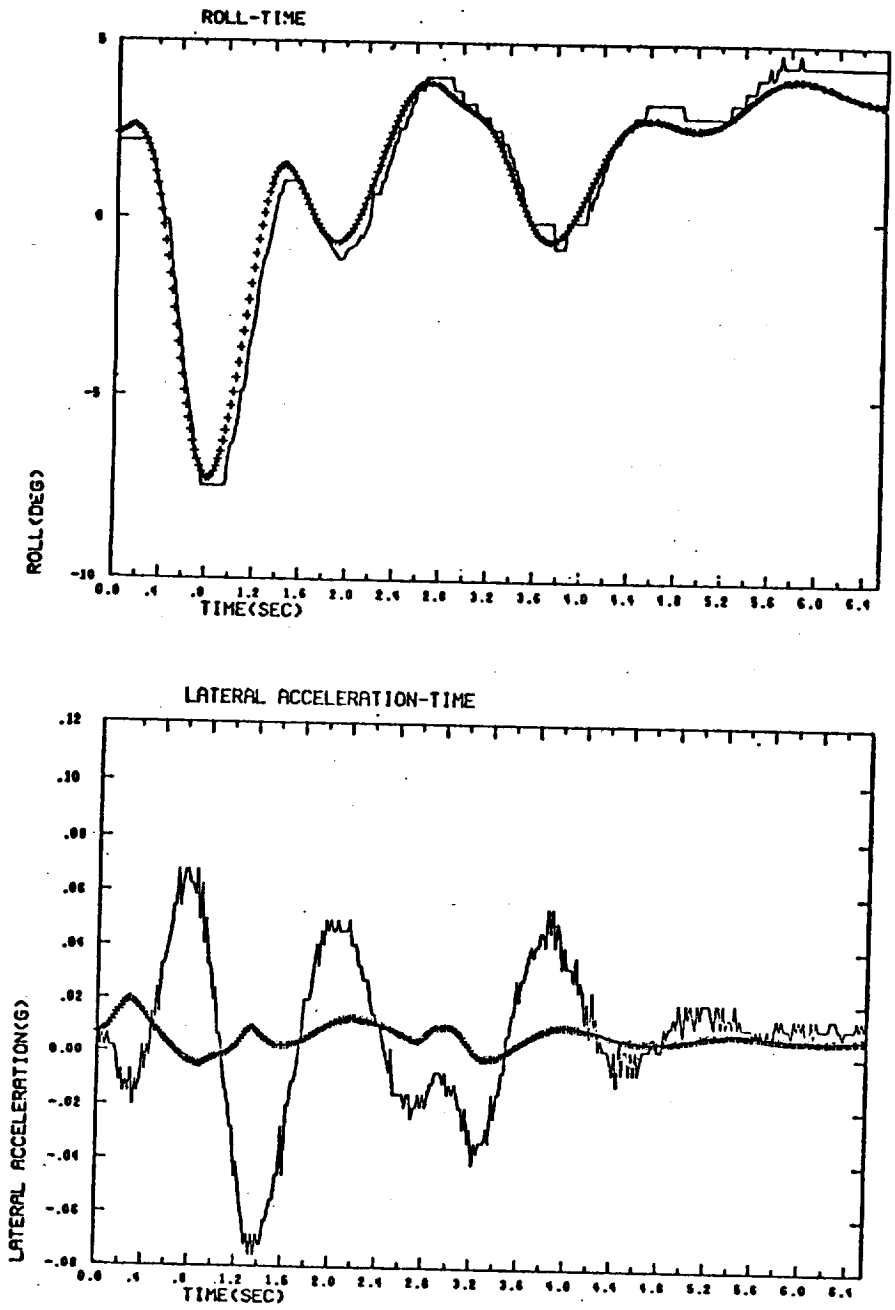


Fig. 5.10 HL-10 : OUTPUT ERROR WITH A PRIORI WEIGHTING AND BIASES

farther removed from the wind tunnel values.

No additional processings with varied a priori weights were made, since sufficient data by which these variations could be justified were not available.

5.2.8 Parameter Estimates Used for Prediction

It has been often stated that using a set of estimates for the stability and control derivatives to predict the measurements from a flight test, under similar conditions and with similar instruments as the one used to identify the derivatives, would be the most valid test of the accuracy of the parameter estimates. Since another set of flight data for the HL-10, under similar conditions, was not available, an experiment was run in which only the first 227 points of data were used to identify the parameter estimates and these results were used to predict the final 100 points (2 seconds) of data. The identification algorithm which was used included the a priori weighting and the identification of the output biases. As the fits to the observed data, given in Figure 5.11, indicate, there is some divergence at the end, especially for r. However, the divergence in β and ϕ was anticipated since the observed data suffers heavily from clipping during the final 1 second. The fit to the lateral acceleration was as good as might be expected considering the fit to the first 227 points.

5.3 M2/F3 Flight Test Data

The data supplied to Systems Control, Inc. on flight No. 21, case 6 of the M2/F3 lifting body is shown in Figure 5.12. The influence of wind gusts is evident in the time histories of the sideslip angle and the lateral acceleration. Referring to Section 5.2.1, the wind gusts were assumed to enter the dynamical equations of motion in exactly the same manner as the sideslip angle β . Nothing was known, a priori, about the

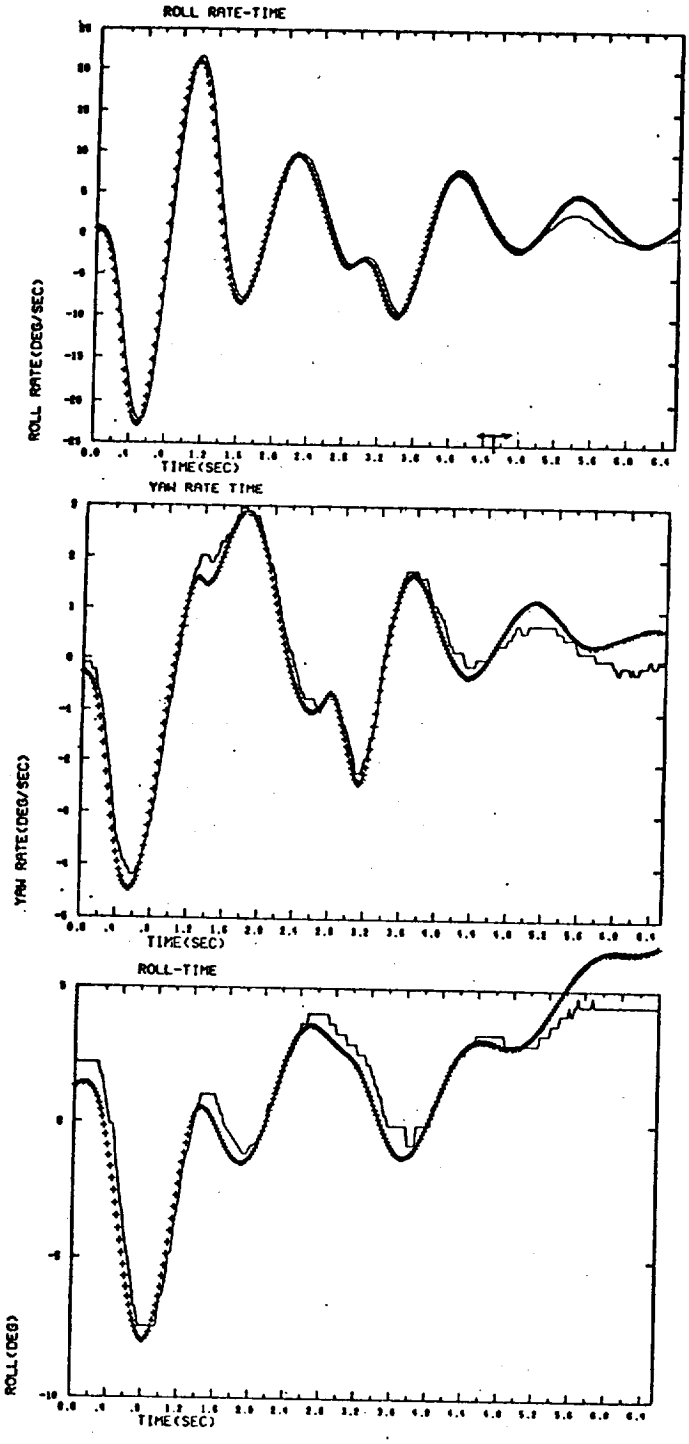


FIGURE 5.11 HL-10: PREDICTION OF FINAL 2 SECONDS OF DATA

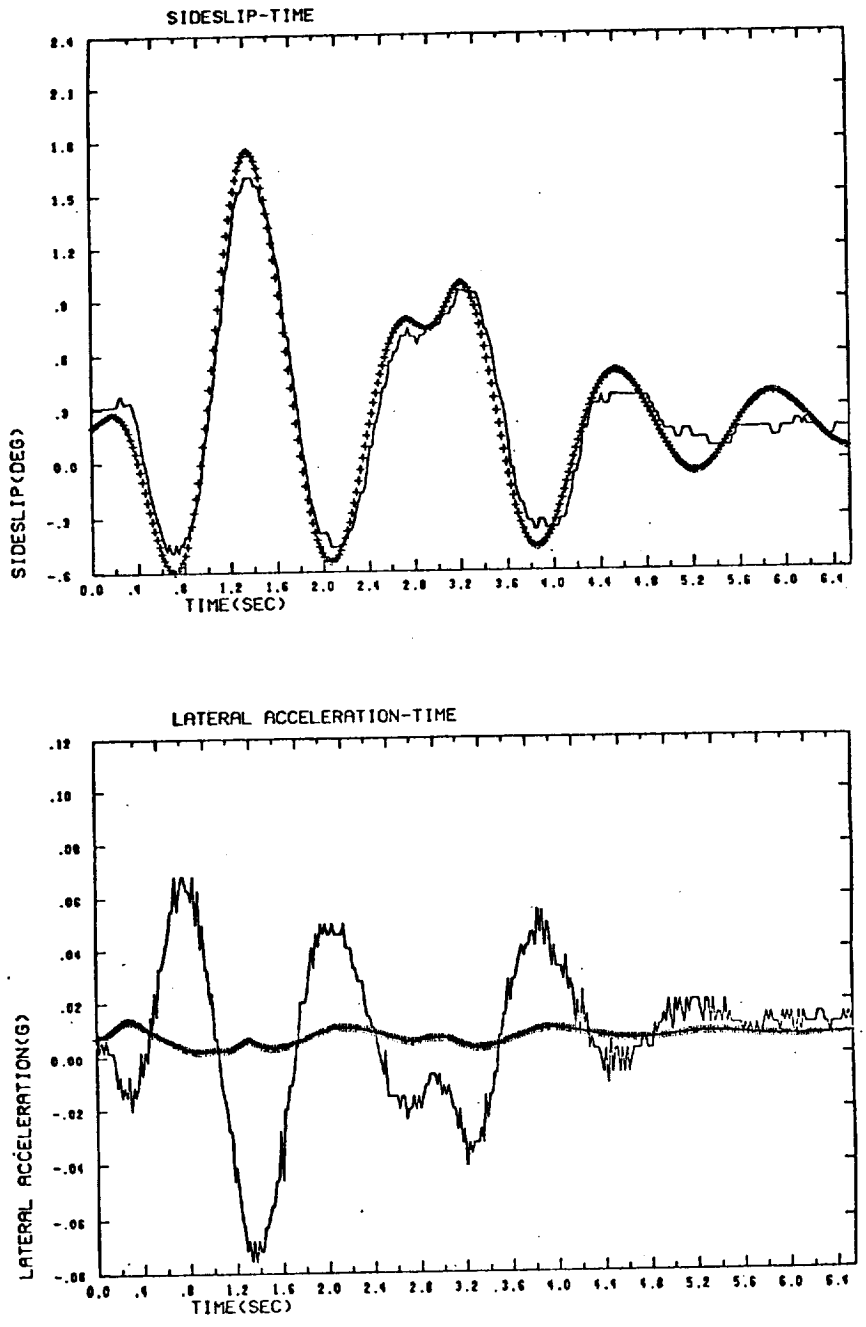


FIGURE 5.11 (CONT.) HL-10: PREDICTION OF FINAL 2 SECONDS OF DATA

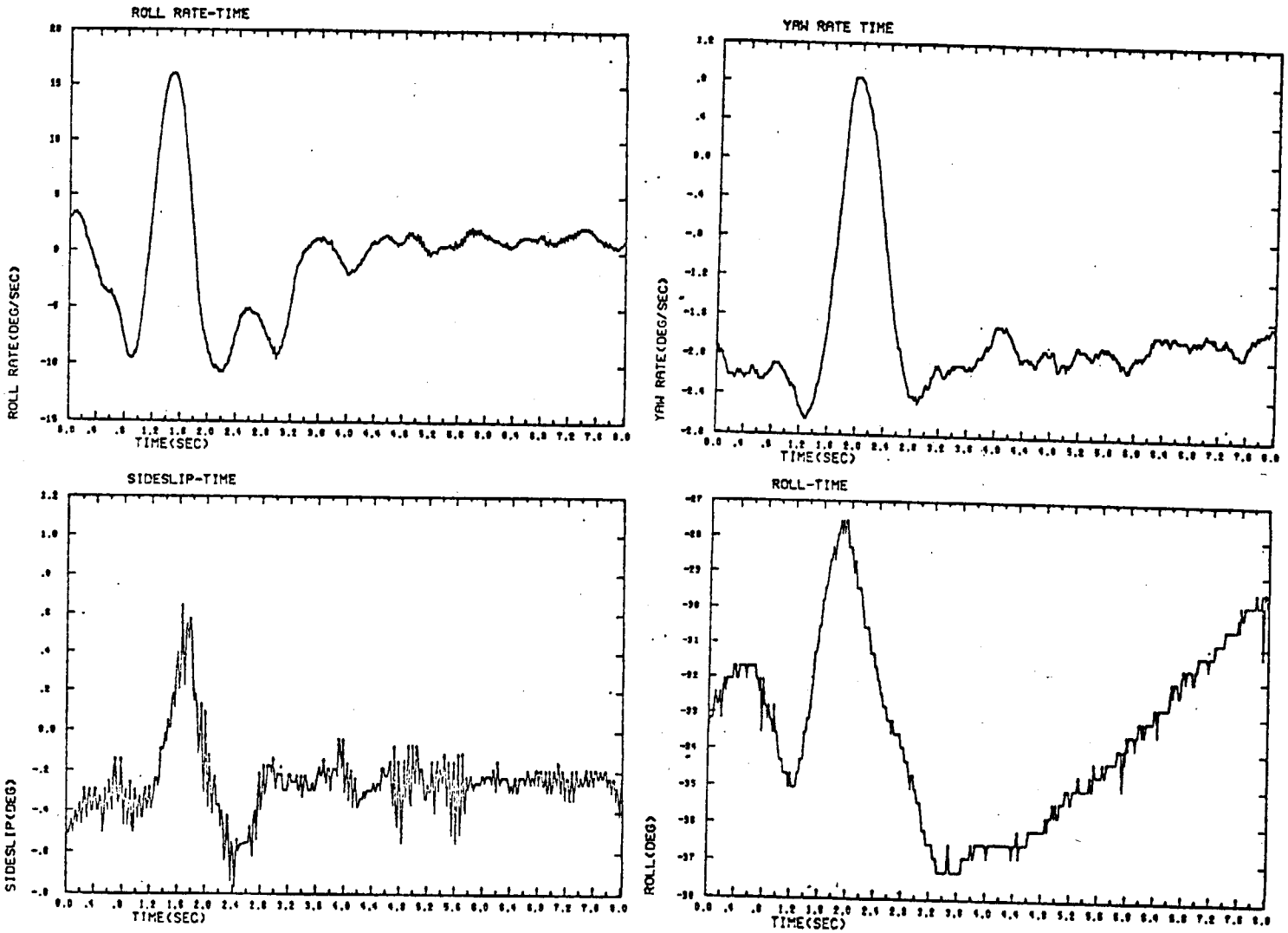


FIGURE 5.12 M2/F3: OBSERVED DATA AND CONTROL SEQUENCE TIME HISTORIES

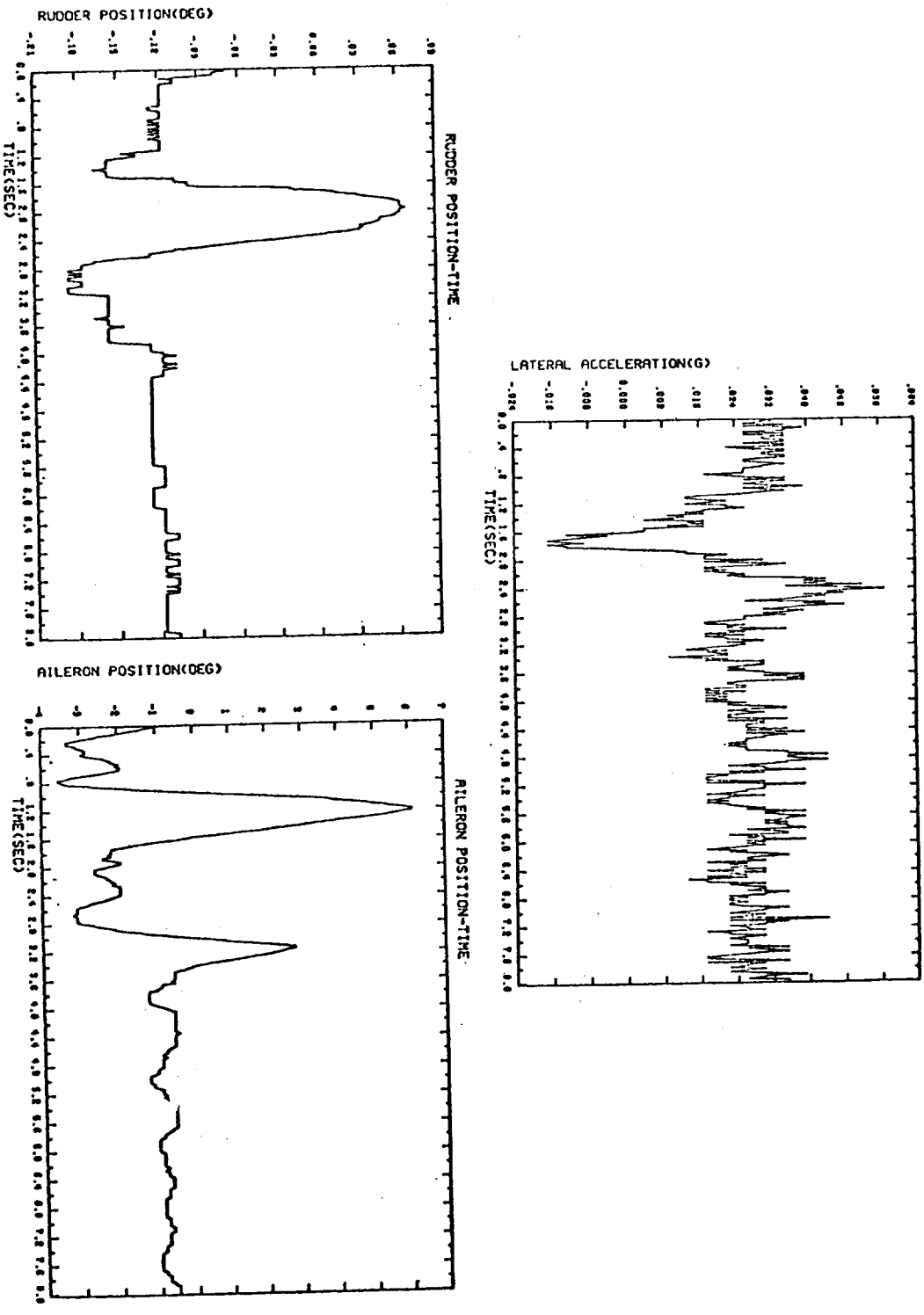


FIGURE 5.12 (CONT.) M2/F3: OBSERVED DATA AND CONTROL SEQUENCE TIME HISTORIES

statistics (correlation time, mean square value) of the wind gusts. The only information supplied SCI was that the output error program used by FRC had failed to match the time histories adequately. A total of 401 data points were supplied, representing 8.02 seconds of data. Once again, the effects of the instrumentation (quantization, clipping) were ignored as were the dynamics of the boom which measures sideslip angle (β vane). It was also interesting to note that the flight conditions were appreciably different than for the HL-10. The angle of attack was only 1.57° and the Mach number was .468.

Seven separate runs were made with the M2/F3 data, indicating a succession of possible model representations for the equations of motion. Since neither the measurement noise nor wind gust statistics were known a priori, these were included, where called for, in the list of parameters to be identified, along with the stability and control derivatives and initial conditions.

5.3.1 Output Error - No Wind Gusts Included

The maximum likelihood algorithm, in the output error mode, with the wind gusts assumed zero, was first used in trying to process the M2/F3 flight data. It was intended that from such a run, it would become apparent where the wind gusts were having the most impact and also the results would serve as a standard against which the identification algorithm performance with the wind gusts included, could be measured.

The time histories of the fit in each of the five measurements are given in Figures 5.13. As these figures indicate, the worst fits were obtained on the sideslip angle and lateral acceleration measurements, although none of the fits were as good as with the HL-10 data. These results also indicated that the model for including the wind gusts, suggested in Section 5.3.3 is appropriate, since the measurements involving the sideslip angle show the most random fluctuation when compared to the data from the HL-10 flight.

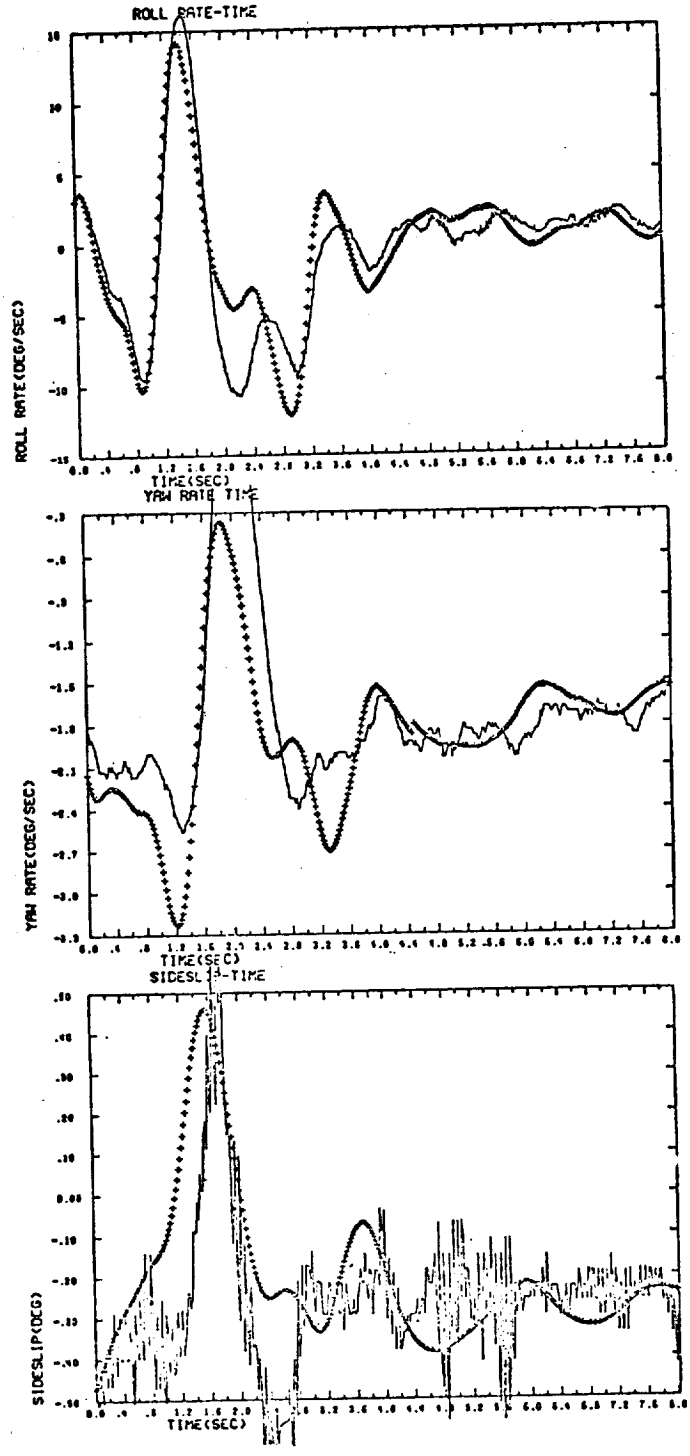


FIGURE 5.13 M2/F3: OBSERVATIONS AND ESTIMATES - OUTPUT ERROR

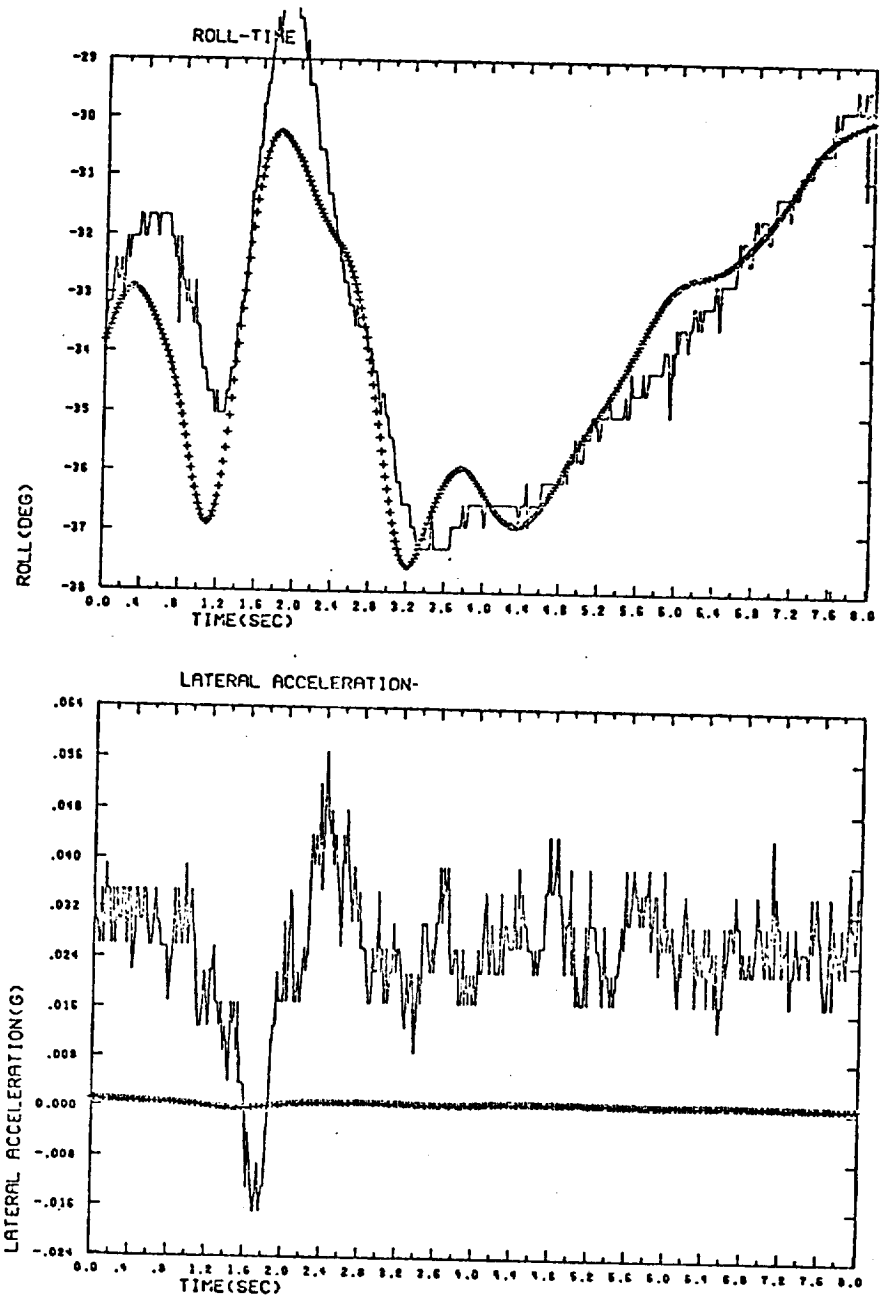


Fig. 5.13(cont'd) M2/F3: OBSERVATIONS AND ESTIMATES -OUTPUT ERROR

The parameter values obtained for this processing of the data are given in Table 5.4.

5.3.2 Perfect Measurement of Sideslip Angle

For this processing of the M2/F3 data, it was assumed that the measurement noise on the sideslip angle measurement is much smaller than the gust noise. With this assumption there is a perfect correlation between the process noise and the sideslip angle measurement noise, both being β_n. The state and sideslip angle measurement equations now appear as

$$C\dot{x} = Fx + Gu + \Gamma\beta_n$$

$$y_3 = \beta + \beta_n + n_\beta \cong \beta + \beta_n$$

The Kalman filter for the complete four state, five output model must account for the perfect β_n correlation. The most direct method for doing this is to first construct an equivalent four state model which is uncorrelated with the sideslip angle measurement. This is done by adding the quantity $y_3-\beta-\beta_n$, which has value zero, to the dynamics, i.e.,

$$C\dot{x} = Fx + Gu + \Gamma\beta_n + \theta\ (y_3-\beta-\beta_n)$$

and solving for θ such that $E\left\{(\Gamma\beta_n - \theta\beta_n)\ \beta_n^T\right\} = 0$. $\theta = \Gamma$ is seen to be the solution and the equivalent model has the resulting from

$$C\dot{x} = Fx + Gu + \Gamma(y_3-\beta)$$

This equation is in the form of the Kalman filter, and it can further be shown that is the exact Kalman gain. Since is the third column of F, the dependence of $\dot{x}$ on β is eliminated and the equations of motion and the measurement equations can be rewritten as

TABLE 5.4 M2/F3 PARAMETER ESTIMATES AND STANDARD DEVIATIONS

Parameter	Wind Tunnel & Theo-retical	Max. Lik. Estimate -output error mode (with St'd dev.)		Max. Lik Estimate Assuming perf. β Meas. (with St'd dev.)		Max. lik. Estimate Directly ident. of β_n (with St'd dev.)		Max. lik. Estimates With a priori weighting (with St'd dev.)		Max. lik. estimates with dependent params. fixed. fixed. (with St'd dev.)
L_p	-0.4673	-1.548	(0.0935)	0.679	(0.035)	-1.779	(0.214)	-0.461	(0.0182)	*
L_r	0.8878	2.008	(1.187)	10.49	(0.547)	25.46	(1.908)	4.154	(0.140)	*
L_β	-75.140	-54.49	(2.45)	-97.79	(1.615)	-135.38	(2.238)	-67.95	(1.02)	*
N_p	.0802	.102	(0.006)	-.0203	(0.00393	-.142	(0.0147)	.00475	(0.00349)	*
N_r	-.6876	-.0307	(0.078)	-1.675	(0.0590)	1.628	(0.199)	-.764	(0.0134)	*
N_β	7.5342	2.876	(0.136)	7.324	(0.152)	-9.890	(0.349)	6.763	(.0876)	4.435 (.113)
Y_p	*									*
Y_r	*									*
Y_β	-.2001	-.0476	(0.125)	-1.249	(.0597)	-1.466	(0.0386)	-.202	(0.00392)	-1.36 (.0594)
$L_{\delta a}$	14.04	14.82	(0.301)	9.804	(0.109)	16.022	(0.3017)	10.96	(0.161)	9.66 (.169)
$L_{\delta r}$	10.03	73.97	(8.59)	-109.28	(5.519)	-157.130	(17.88)	-42.18	(3.13)	*
L_o	0	11.14	(1.828)	-10.46	(0.328)	.145	(2.11)	-.572	(0.115)	-9.004 (.141)
$N_{\delta a}$	.83	.596	(.0223)	.719	(.0104)	2.128	(0.0456)	.762	(0.111)	.756 (.0134)
$N_{\delta r}$	-4.06	-12.874	(0.578)	6.844	(0.643)	-11.754	(1.467)	-4.37	(0.106)	*
N_o	0	-.345	(0.121)	.177	(0.0357)	.427	(0.198)	-.233	(0.0433)	-.00239 (.0320)
$Y_{\delta a}$	0	-.00033	(0.0151)	-.0363	(0.00669)	-.0125	(0.00689)	-.0847	(0.00867)	-.0275 (.00634)
$Y_{\delta r}$	0	.0301	(0.363)	-.874	(0.222)	-.926	(0.227)	-1.932	(0.286)	*
Y_o	0	.0179	(0.354)	.378	(0.0299)	.295	(0.0313	-.0974	(0.378)	.456 (.0189)
ϕ_{bias}				-.281	(0.0531)	2.936	(0.0629)	-6.01	(0.0933)	-.667 (.108)
$p_{initial}$		3.807	(0.521)	.359	(0.188)	1.657	(0.852)	4.846	(0.296)	-3.125 (.239)
$r_{initial}$		-2.262	(0.0785)	-1.66	(0.0280)	-1.604	(0.0556)	-2.22	(0.0453)	-1.061 (.0597)
$\beta_{initial}$		-.558	(.0251)	*		- .565	(.0279)	*		*
$\phi_{initial}$		-34.44	(.175)	-32.69	(.158)	-33.630	(.0796)	-31.91	(.223)	-31.52 (.224)
a						-44.147	(.856)			
q						6.231	(.107)			
$\beta_{\eta_{initial}}$						2.937	(.256)			
Likelihood Function Value		-1502		-2237		-2038		-1122		-1051

TABLE 5.4 (CONT'D)

Parameter	Max. Lik. with Rank Deficient Solution	
L_p	-.531	(.0189)
L_r	4.268	(.144)
L_β	-103.35	(.105)
N_p	.0397	(.00682)
N_r	-.989	(.0672)
N_β	7.568	(.306)
Y_p	*	
Y_r	*	
Y_β	-1.19	(.0590)
$L_{\delta a}$	10.25	(.0845)
$L_{\delta r}$	-5.539	(.0257)
L_o	-10.89	(.280)
$N_{\delta a}$	.561	(.0254)
$N_{\delta r}$	-.512	(.651)
N_o	.587	(.0833)
$Y_{\delta a}$	-.0360	(.00660)
$Y_{\delta r}$	-.737	(.219)
Y_o	.408	(.0296)
ϕ_{bias}	-.164	(.0428)
$P_{initial}$	-1.029	(.139)
$r_{initial}$	-2.054	(.0552)
$\phi_{initial}$	-31.576	(.0705)
Likelihood Function Value	-1689	

$$C' = \begin{bmatrix} \dot{\hat{p}} \\ \dot{\hat{r}} \\ \dot{\hat{\phi}} \end{bmatrix} = \begin{bmatrix} L_p & L_r & 0 \\ N_p & N_r & 0 \\ 1 & \tan\theta & 0 \end{bmatrix} \begin{bmatrix} \hat{p} \\ \hat{r} \\ \hat{\phi} \end{bmatrix} + \begin{bmatrix} L_{\delta_a} & L_{\delta_r} & L_o & L_\beta \\ N_{\delta_a} & N_{\delta_r} & N_o & N_\beta \\ 0 & 0 & \phi_o & 0 \end{bmatrix} \begin{bmatrix} \delta_a \\ \delta_r \\ 1 \\ y_3 \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} p \\ r \\ \phi \\ Y_\beta \cdot y_3 + Y_{\delta_a}\delta_a + Y_{\delta_r}\cdot\delta_r + Y_o \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ n_4 \\ n_5 \end{bmatrix}$$

where y_3 is treated as a deterministic control. The order of the dynamical system has been reduced to 3 and the number of measurements to 4.

Once a complete set of parameters has been obtained for this reduced order system, the time history of β_n can be recovered. This is important since the identification of the statistics of the wind gusts is also possible using identification. Sideslip angle estimate $\hat{\beta}$ can be found by substituting the parameter values of the three state model into the original four state model and solving for its time history. Then subtracting from the sideslip angle measurement gives the time history of $\beta_n + n_\beta$ (n_β was originally assumed small). A first-order linear model of the form

$$\dot{\beta}_n = a\beta_n + v_\beta$$

$$(y_3 - \hat{\beta}) = \beta_n + n_\beta$$

where v_β is the process noise with covariance q, can be fit to this data and the time constant a, the process noise covariance q, and the covariance of the measurement noise n_β identified.

The time histories of the fit to the four observations (not including sideslip angle) are given in Figs. 5.14. The parameter estimates along with the estimates of the process noise covariance and the (reciprocal of the) time constant for the wind gust are given in Table 5.3. The time history of the wind gust β_n (including the neglible measurement noise) is shown in Fig. 5.15.

The fit in each of the four measurements is very good, although time histories of the fit error indicate that there is still the same sinusoidal variation, especially in p, that was observed in the HL-10 fit errors. Only the fit error in the lateral acceleration, a_y, approached being white noise, which is the indication of the best possible fit. The value of the covariance of the noise on the sideslip angle measurement was almost two orders of magnitude smaller than the process noise covariance which supports the original assumption of this run. One surprising result,

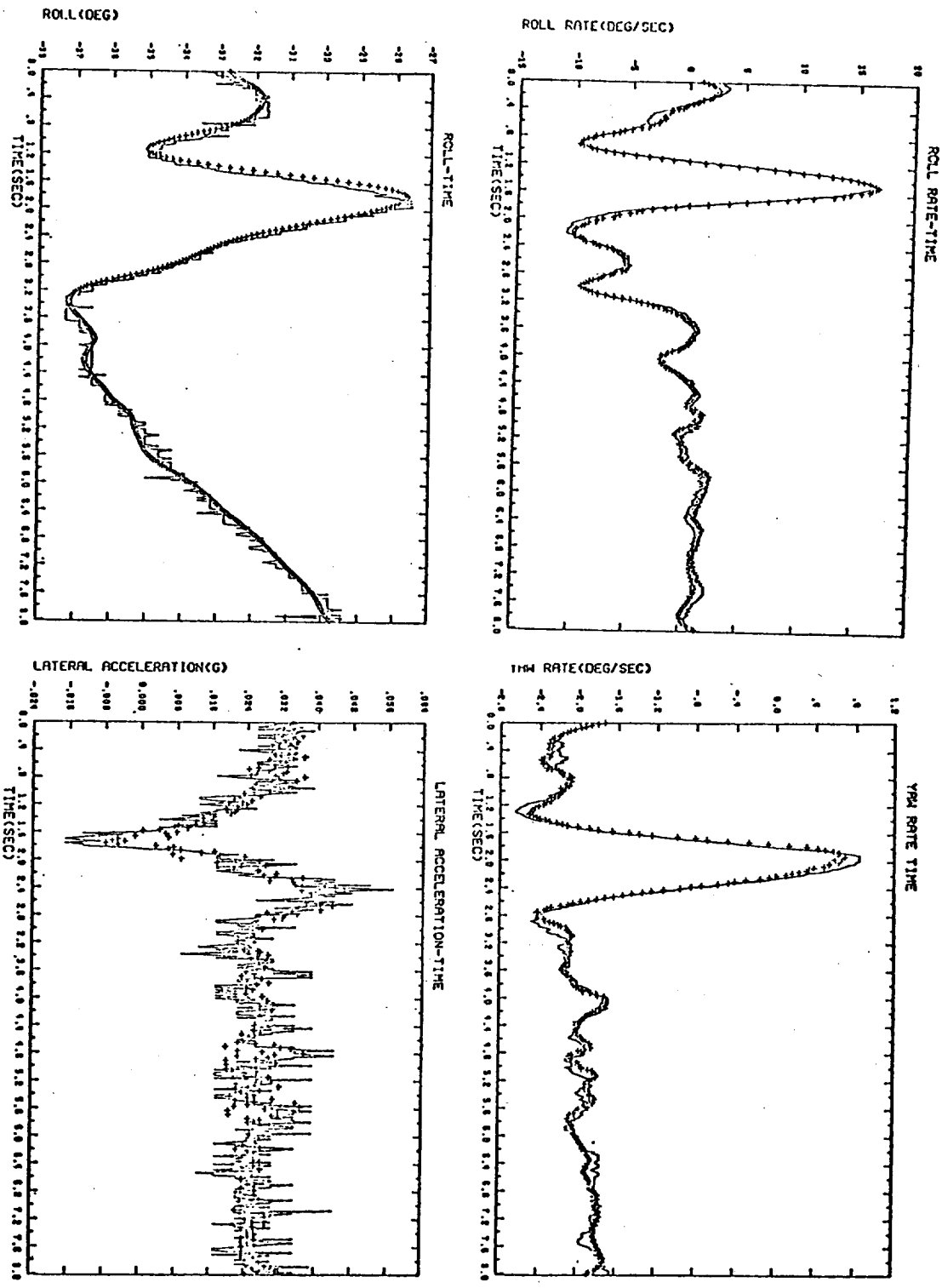


Fig. 5.14 M2/F3 : OBSERVATIONS AND ESTIMATES - KALMAN FILTER WITH $z_\beta = \beta + \beta_N$

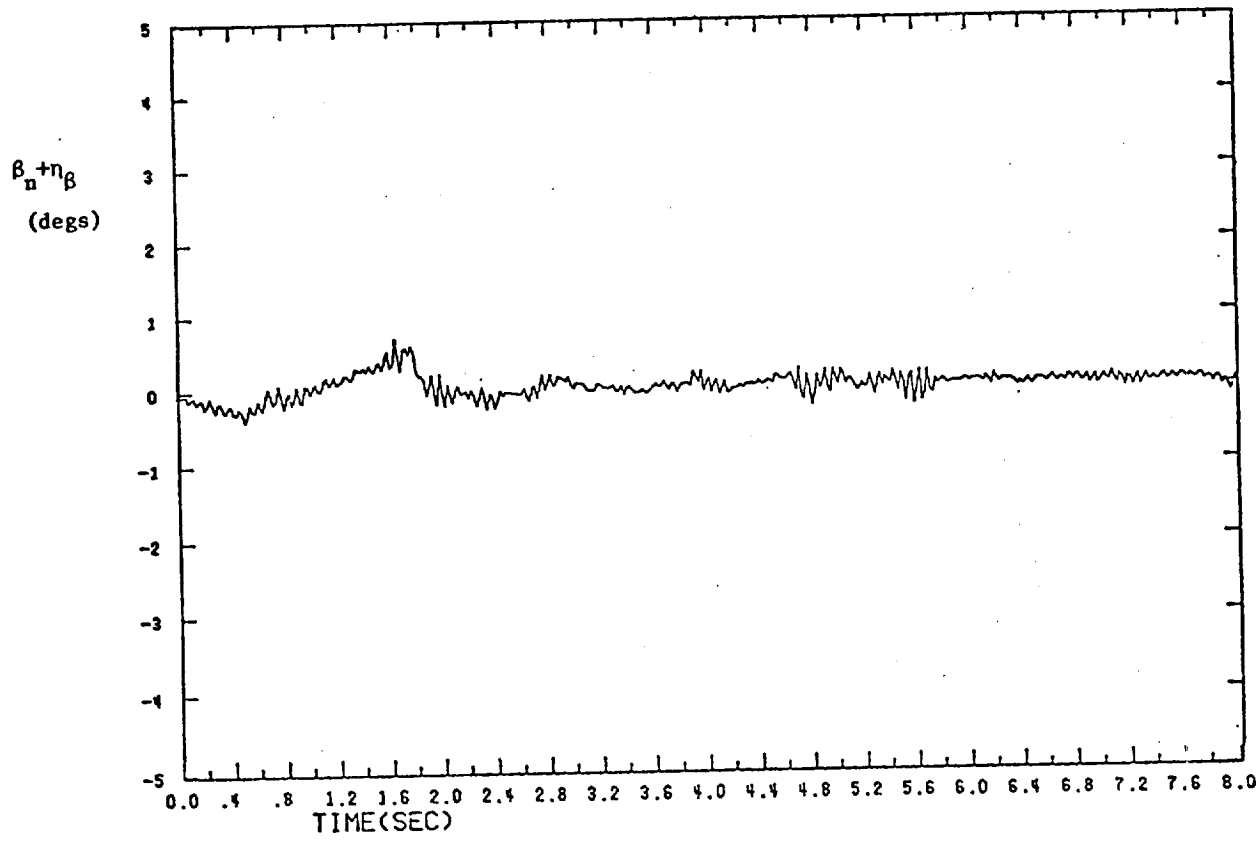


Fig. 5.15 TIME HISTORY of $\beta_n + \eta_\beta$

however, is that the L_p, N_{δ_r}, and L_{δ_r} parameters changed sign from the wind tunnel value.

There appeared to be two principle reasons for the N_{δ_r} and L_{δ_r} parameters having the wrong signs. The first was that the magnitude of the aileron variation was much larger than the rudder variation, unlike in the HL-10 case. Since the effect of the controls is additive in determining $\dot{p}$ and $\dot{r}$, there is an identifiability problem with respect to N_{δ_r} and L_{δ_r}. This is substantiated by the small values of the terms of the sensitivity matrix corresponding to the N_{δ_r} and L_{δ_r} parameters.

The second factor contributing to the incorrect signs is the operation of the yaw damper. This causes a feedback loop which activates the rudder as a result of yaw rate. The time histories of r and δ_r appear in phase, therefore, in the M2/F3 time histories. In such a situation, unless the control δ_r is modeled as a linear combination of the states, there is a uniquesness problem as to whether the actual aircraft dynamics on the feedback loop is being identified. With the yaw rate and δ_r in phase, the N_{δ_r} parameter, at least, will appear with a positive sign.

The problem of incorrect signs was of major concern and was the motivating factor for many of the remaining processing of the M2/F3 data. The experience with the HL-10 data indicated that constraining those parameters with wrong signs to have the same signs as the wind tunnel values would not correct the problem. The solution had to lie either in a more complete aircraft model or in dealing directly with the numerical problems causing the incorrect signs.

5.3.3 Wind Gusts Included: Direct Identification of Process Noise Covariance and Time Constant of Correlated Gusts

For the third processing of the M2/F3, the gusts were included directly in the dynamical and measurement equations requiring that a full Kalman filter be used in the maximum likelihood identification algorithm in order to obtain the sensitivities. For this processing of the data, the model of the wind gusts obtained from the previous run was used

$$\dot{\beta}_n = a\beta + v_\beta \tag{5.3}$$

where a is the reciprocal of the time constant and v_β is an unknown disturbance with covariance q. Replacing β_n in the original system equations by equation (5.3) results in

$$C\begin{bmatrix}\dot{p}\\ \dot{r}\\ \dot{\beta}\\ \dot{\phi}\\ \dot{\beta}_n\end{bmatrix} = \begin{bmatrix} L_p & L_r & L_\beta & 0 & 0\\ N_p & N_r & N_\beta & 0 & 0\\ Y_p + \sin\alpha & Y_r + \cos\alpha & Y_\beta & \frac{g\cos\theta}{v} & 0\\ 1 & \tan\theta & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & a\end{bmatrix}\begin{bmatrix}p\\ r\\ \beta\\ \phi\\ \beta_n\end{bmatrix} + \begin{bmatrix} L_{\delta_a} & L_{\delta_r} & L_o\\ N_{\delta_a} & N_{\delta_r} & N_o\\ Y_{\delta_a} & Y_{\delta_r} & Y_o\\ 0 & 0 & \Phi_o\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\delta_a\\ \delta_r\\ 1\end{bmatrix} + \begin{bmatrix}0\\ 0\\ 0\\ 0\\ 1\end{bmatrix} v_\beta$$

Note that an additional bias term Φ_o has been added to the $\dot{\phi}$ equation. With t he inclusion of the covariance of v_β and the β_n state, there are now 24 parameters to identify: 17 stability and control derivatives (Y_p and Y_r are assumed zero), 5 initial conditions, q and a.

The difficulty involved in setting up the identification algorithm in such a case is that both the measurement and process noise covariances, R and q respectively, are unknown. However, both R and q are needed in establishing the Kalman filter gain, which is assumed to be in steady state. To begin the identification, therefore, some initial estimates of both q and R are necessary. R is assumed to be the diagonal elements of the sample covariance matrix obtained from the output error method. An initial value for q is obtained from the results of the previous run.

Once the initial iteration is completed, the value of q is updated like any other parameter and the measurement noise covariance, R, is obtained from the sample covariance. This last fact is derived from the property that with $\hat{x}_{i|i-1}$ defined as the Kalman estimate of the state at time t_i given data up to time t_{i-1},

$$\lim_{N\to\infty} \frac{1}{2}\sum_{i=1}^{N} (y_i - H\hat{x}_{i|i-1} - Du_i)\,(y_i - H\hat{x}_{i|i-1} - Du_i)^T \doteq HP_{ss}H^T + R$$

where P_{ss} is the steady state error covariance matrix, obtained from solving a discrete Ricatti equation. The above expression is only approximate for finite data lengths.

The time histories of the observations and the estimates are given in Fig. 5.16, and the final parameter values are given in Table 5.4. Although the fits to the p, β, ϕ and a_y measurements obtained from this run improved over those obtained from the output error method, they are not totally acceptable. It is interesting to note that the signs of the parameters L_p and L_r have retained the same sign as the wind tunnel values.

The time histories in Fig. 5.16 also indicate that most of the fits to the observed data are biased. Inclusion of measurement biases in the list of parameters to be identified did not, however, improve the performance.

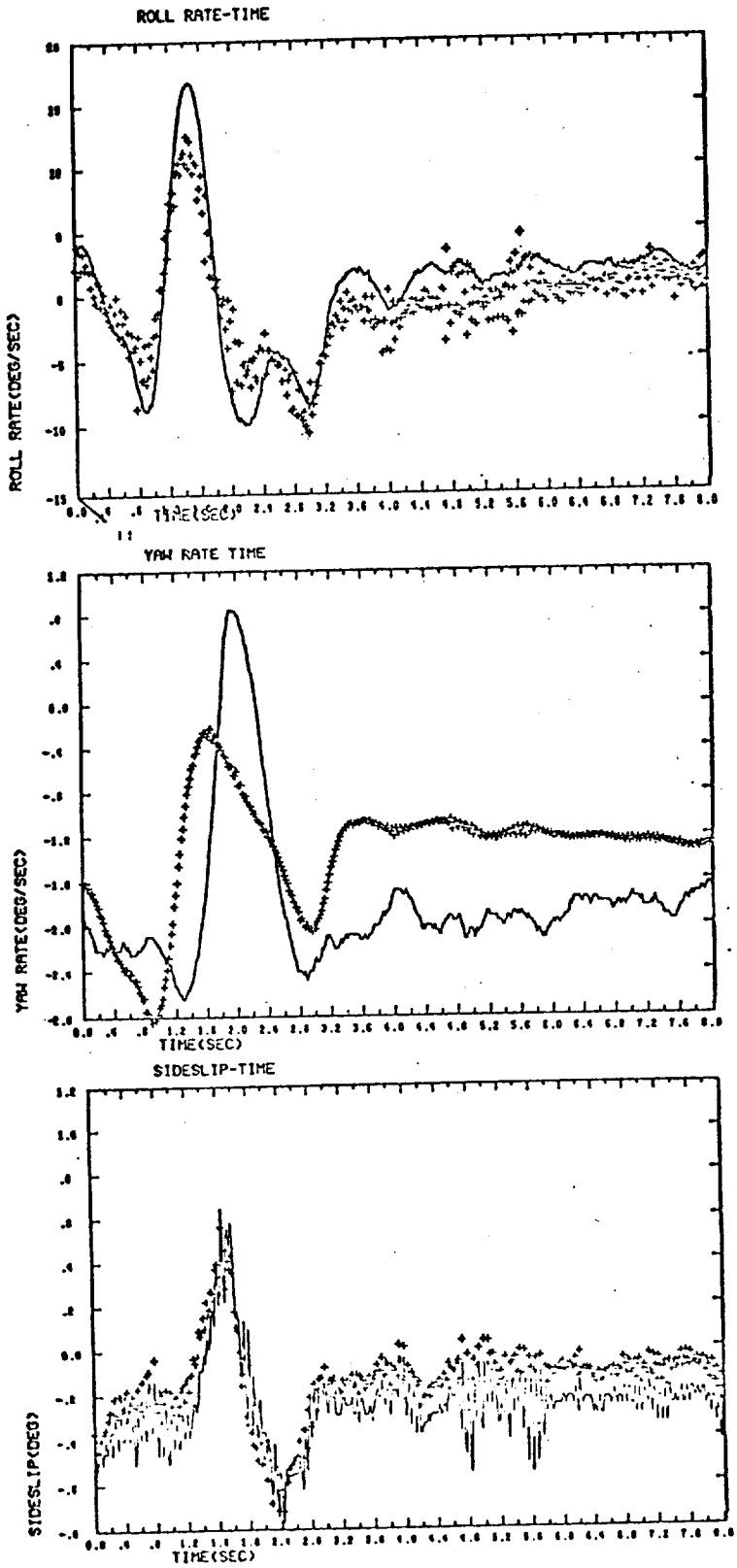


FIG. 5.16. M2/F3: DIRECT IDENTIFICATION OF WIND GUST MODEL

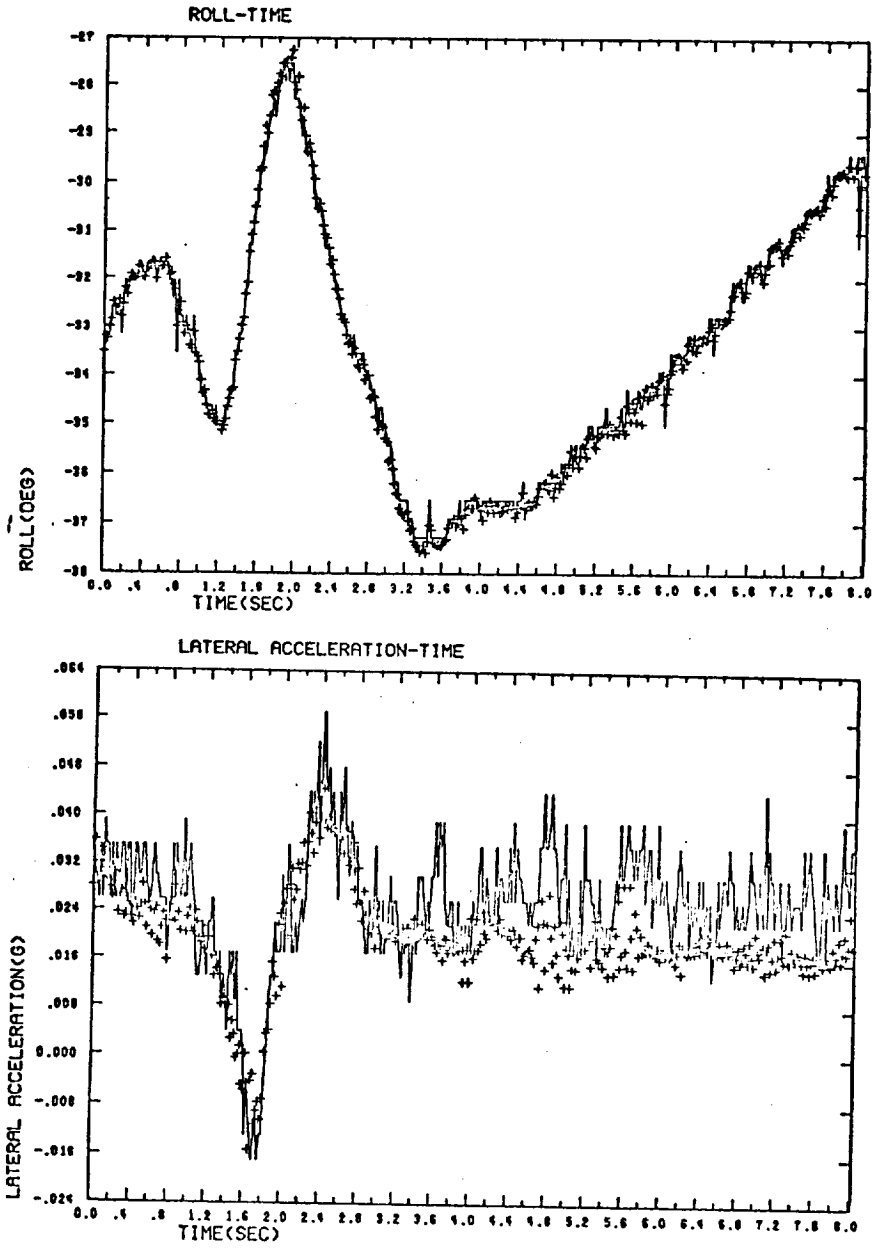


FIG. 5.16 (CONT'D): DIRECT IDENTIFICATION OF WIND GUST MODEL

The reason the maximum likelihood algorithm with the Kalman filter cannot reduce the fit error any further is basically numerical. What has occurred is that the diagonal element of the measurement noise covariance matrix, R, associated with the measurement of the sideslip angle has become very small when compared to the covariance of the wind gust disturbance. Indeed, the wind gust itself is practically white. As a result, the measurement noise cannot be distinguished from the wind gust. The alternative was to restructure the model so that the measurements of $\beta + \beta_n$, the total sideslip angle are perfect. This is precisely what was done in the previous processing of the data.

5.3.4 Three State Model With A Priori Weighting

The results of the two previous processings of the M2/F3 data indicated that the assumption of perfect measurements of the sideslip angle was reasonable and produced the best fits to the data. However, as stated earlier, the three state model resulted in wrong signs for many of the parameters. The first processing of the data in an attempt to correct these incorrect signs used the priori weighting technique. The same weights as for the HL-10 data were used for the M2/F3. Measurement biases were included in the list of parameters to be identified since the use of a priori weighting on the HL-10 data indicated the need for bias estimation.

As shown in Fig. 5.17 the fits to the observed data resulting from a priori weighting were quite poor, especially in roll angle and lateral acceleration, although the parameter estimates themselves, as given in Table 5.4, were quite close to the wind tunnel values. It is interesting to note, however, that L_{δ_r} still has a wrong sign.

5.3.5 Three State Model With Fixed Parameters

The basic causes of the incorrect signs for some of the parameters are that either the sensitivity of the output to changes in that parameter are small, as indicated by a relatively small diagonal element in the information matrix, or that there is a correlation, with respect to the sensitivity, between two or more of the parameters being identified. Such a situation would be indicated by an off-diagonal element of the normalized information matrix being close to ± 1. If this were the case the correlated parameters could not be individually identified. Both these problems existed with the M2/F3 data.

One technique which has been used for treating both these problems is to fix one or more of a set of parameters that are correlated. The results of the identification run with fixed parameters indicated (i) the convergence of the algorithm to the final set of parameters estimates is more rapid and monotonic than when all the parameters are being identified, and (ii), the final fit to the observed data is degraded to a certain degree. This latter characteristic is due to the fact that the number of degrees of freedom (equal to the number of parameters to be identified) for fitting the observed data has been reduced. In comparing the value of the likelihood function or cost for two cases with different numbers of parameters being identified (measurement noise covariance R being identified in both cases) the comparison should be made between a corrected cost, given by

$$f(N,k)\,\ell n|R|$$

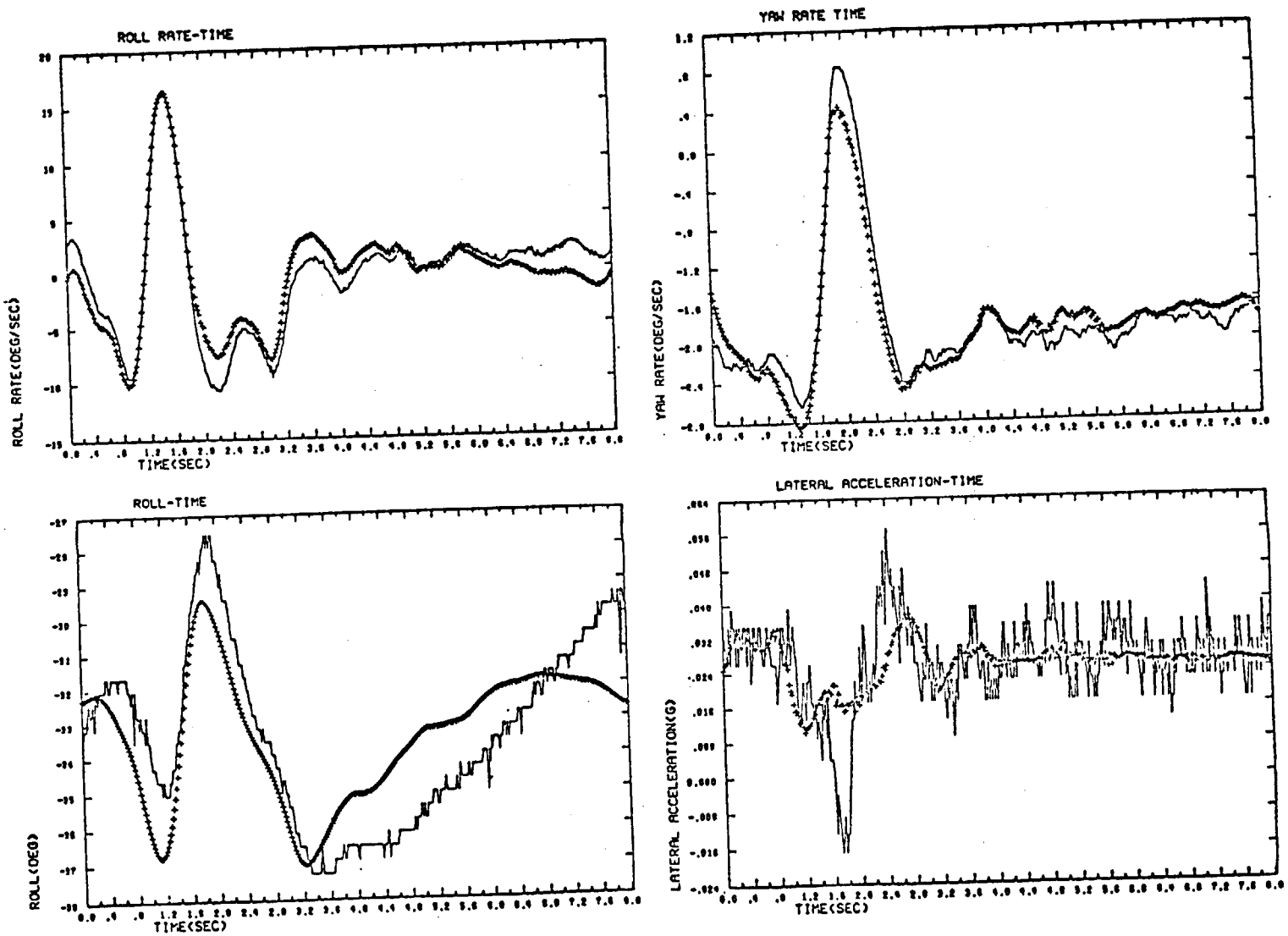


FIG. 5.17 M2/F3 TIME HISTORIES WITH A PRIORI WEIGHTING + + + estimate

where N is the number of data points, and k is the number of parameters being identified and f(N,k) is a monotonically increasing function of k which depends upon the objective of identification. For certain types of systems, Akaike (Ref. 38) has shown that the one-step ahead final prediction error using the identified model is given by

$$J = \frac{N-k}{N+k} \ln |R|$$

A plot of J vs k for a typical model is shown in Fig. 5.18. The important thing to note is that the predictive qualities of a model do not improve monotonically with the number of parameters, even though the fit error decreases monotonically with the number of parameters

It was indicated earlier that there was very little variation in the rudder during the M2/F3 flight, which would make identification of the δ_r derivatives very difficult. In fact, the identified values of the L_{δ_r} and N_{δ_r} derivatives with the three state model were physically unreasonable, being opposite in sign from the wind tunnel values. The first processing of the M2/F3 data in this set of runs was therefore made with the rudder derivatives fixed at the wind tunnel values, with measurement noise biases being included in theunknown parameter set. The results indicated a strong correlation between L_p and N_r and almost all the other parameters, and a fairly poor fit to the data. Fixing the same parameters but including a priori weighting did not improve the performance.

The second processing of the data in this series included fixed L_p and N_r derivatives as well as fixed L_{δ_r} and N_{δ_r}. The results showed only a slight improvement over the previous processing, and a strong correlation still existed between L_r and L_p; N_p and N_β; and

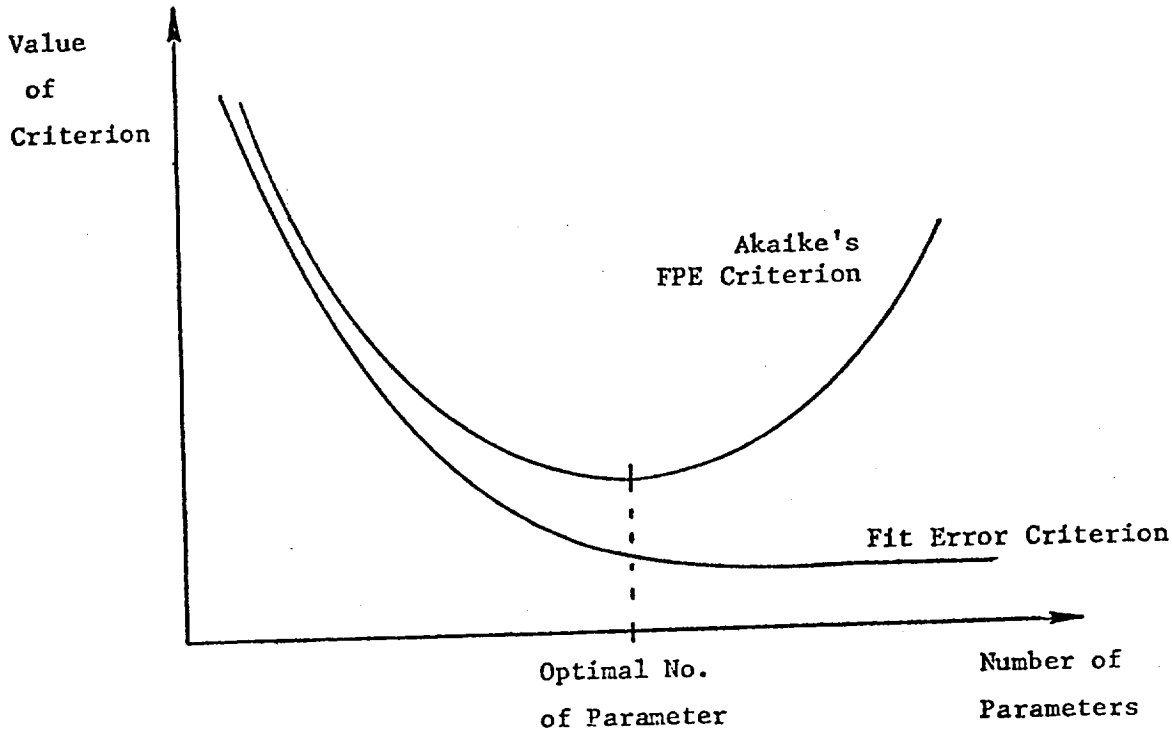


FIG. 5.18 PERFORMANCE CRITERION AS A FUNCTION OF THE NUMBERS OF MODEL PARAMETERS

N_p and y_β. It appeared, at this point, that a problem of identifiability was developing between the measurement biases and the initial conditions. The next processing of the data therefore included fixed L_p, N_r, δ_r derivatives and initial conditions. The fits to the observed data were still very bad and correlation still existed between L_r and L_β, L_r and L_o, and N_P and N_o.

After several more experiments it was decided that L_p, L_r, L_β, N_p, N_r and the δ_r derivatives should all be fixed and the initial conditions be identified instead of the measurement biases. The results, shown in Fig. 5.19, were the best fits to the observed data obtained with the technique of fixing parameters at the wind tunnel values. The values of the parameters which were identified were all of the same sign as the wind tunnel values, as shown in Table 5.4.

5.3.6 Three State Model With Rank Deficient Solution

The results of the data processing, using the technique of fixing selected parameters of a correlated set at fixed values, showed that the convergence rate was improved due to better conditioning of the information matrix. The basic reasons why the parameter fixing technique does not always work are: (i) the correlation is usually not simply between pairs of parameters, but may involve the entire set of unknown parameters, and (ii) it is not usually possible to correctly choose a set of parameters that should be fixed and the values at which they should be fixed. It was decided at this point to investigate, more fully, the problem of possible correlation between more than just pairs of unknown parameters. The solution of this type of dependency problem is to find the directions in parameter space corresponding to combinations of parameters which cannot be identified. A perfect dependency among the parameters would, strictly speaking, result in a zero eigenvalue of the information matrix, causing it to be singular. However, since round-off and other numerical

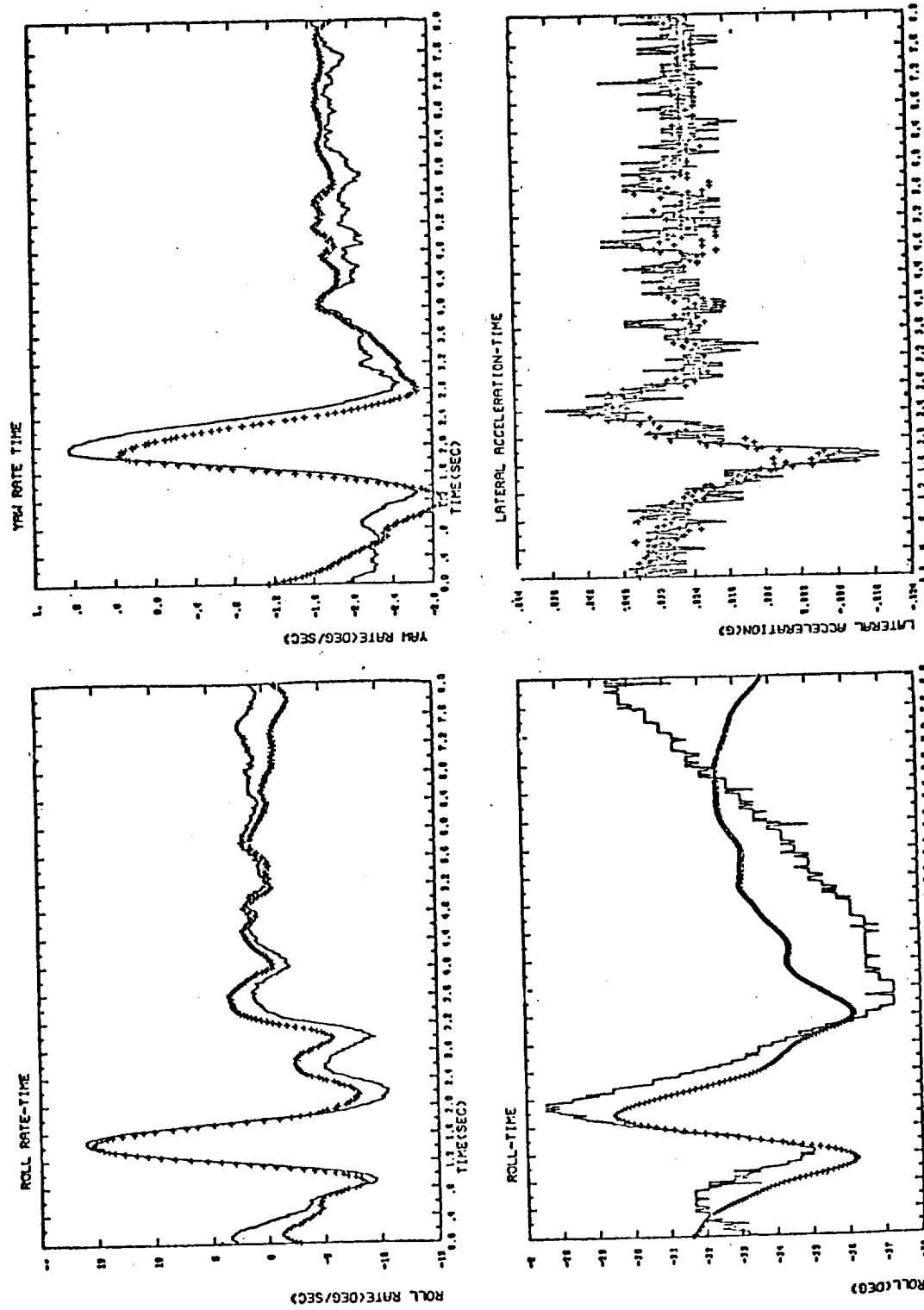


FIG. 5.19 M2/F3 TIME HISTORIES WITH DEPENDENT PARAMETERS AT FIXED VALUES

errors prevent the information matrix from being exactly singular, all the eigenvalues will be non-zero with a spread between the smallest and largest eigenvalue being many orders of magnitude. In such a case it is better to use a rank deficient solution for the inverse rather than a full rank solution. That is, the inverse to the information matrix should be computed leaving out one or more of the smallest eigenvalues.

The reason for neglecting the smallest eigenvalues can be explained in terms of the parameter step. The eigenvalues of the information matrix are the dimensions of the uncertainty ellipsoid, associated with the parameter estimates, in parameter space. The smaller eigenvalues indicating a larger dimension and therefore more uncertainty. Since the LxL information matrix M can be expressed in terms of its eigenvalues and eigenvectors, λ_i and V_i, i = 1, . . , L

$$M = \sum_{L=1}^{L} \lambda_i V_i V_i^T$$

the parameter step is given by (see Section 4.3)

$$\Delta p = M^{-1} (DJ) = \sum_{L=1}^{L} \lambda_i^{-1} V_i V_i^T$$

where DJ is the gradient of the likelihood function. Therefore, the smaller eigenvalues also contribute the largest proportions to the parameter step. This implies that the largest components of the parameter step are in the direction of the most parameter uncertainty. Therefore, the information matrix inverse is computed neglecting a certain number of the smaller eigenvalues, i.e.

$$M^{-1} = \sum_{i=1}^{L-K} \lambda_i^{-1} V_i V_i^T$$

Each eigenvalue which is left out relates to a singular direction in parameter space, and, therefore indicates a combination of parameters which cannot be identified uniquely. Rather than fix the value of one or more of the parameters, as was necessary with the a priori weighting technique, the rank deficient solution fixes combinations of parameters corresponding to nearly zero eigenvalues. Thus the dimension of the space in which the set of

parameter values that minimizes the cost are sought, is reduced. It is important to realize, however, that at each iteration the values for all the specified unknown parameters are assigned updated estimates.

The number of eignevalues to be neglected depends on which order rank deficient solution produces a parameter step resulting in a set of parameters with the lowest associated cost. The procedures is as follows. Starting from some inimimum number of eigenvalues (10 in the HL-10 case), a reduced rank inverse is computed, the parameter step is determined and the associated cost evaluated. One more eignvalue is then added and the procedure repeated. This same thing is done until all the eigenvalues are added in, with the last inverse being a full rank inverse. The same procedure, starting from a minimum number of eigenvalues and progressing to the full rank, is repeated every iteration.

The fits to the observed data are shown in Fig. 5.20. Comparing these with Fig. 5.14, it can be seen that the fits are only slightly degraded. The fits to roll angle and lateral acceleration are much better than those obtained for either a priori weighting or fixing of correlated parameters. The parameter values obtained for this third order rank deficient solution (3 eigenvalues neglected) are given in Table 5.4. Although several of the parameter still have opposite signs from the wind tunnel values, many of them are much closer to the wind tunnel values and are more reasonable than the full-rank, 3 state parameter estimates. Some, such as N_{δ_r}, now have the correct sign from physical considerations, where before they did not. It is clear that further development work on this rank-deficient solution approach will improve the estimates even more.

It should be mentioned that the basic identifiability problem in the M2/F3 data is due to the stability augmentation system (SAS) providing all the sudden movement. The methods described above (viz. a priori weighting, fixing parameters and rank-deficient solutions) are indirect means of handling this problem. It would be more exact to model the relevent characteristics of SAS in order to alleviate the problems.

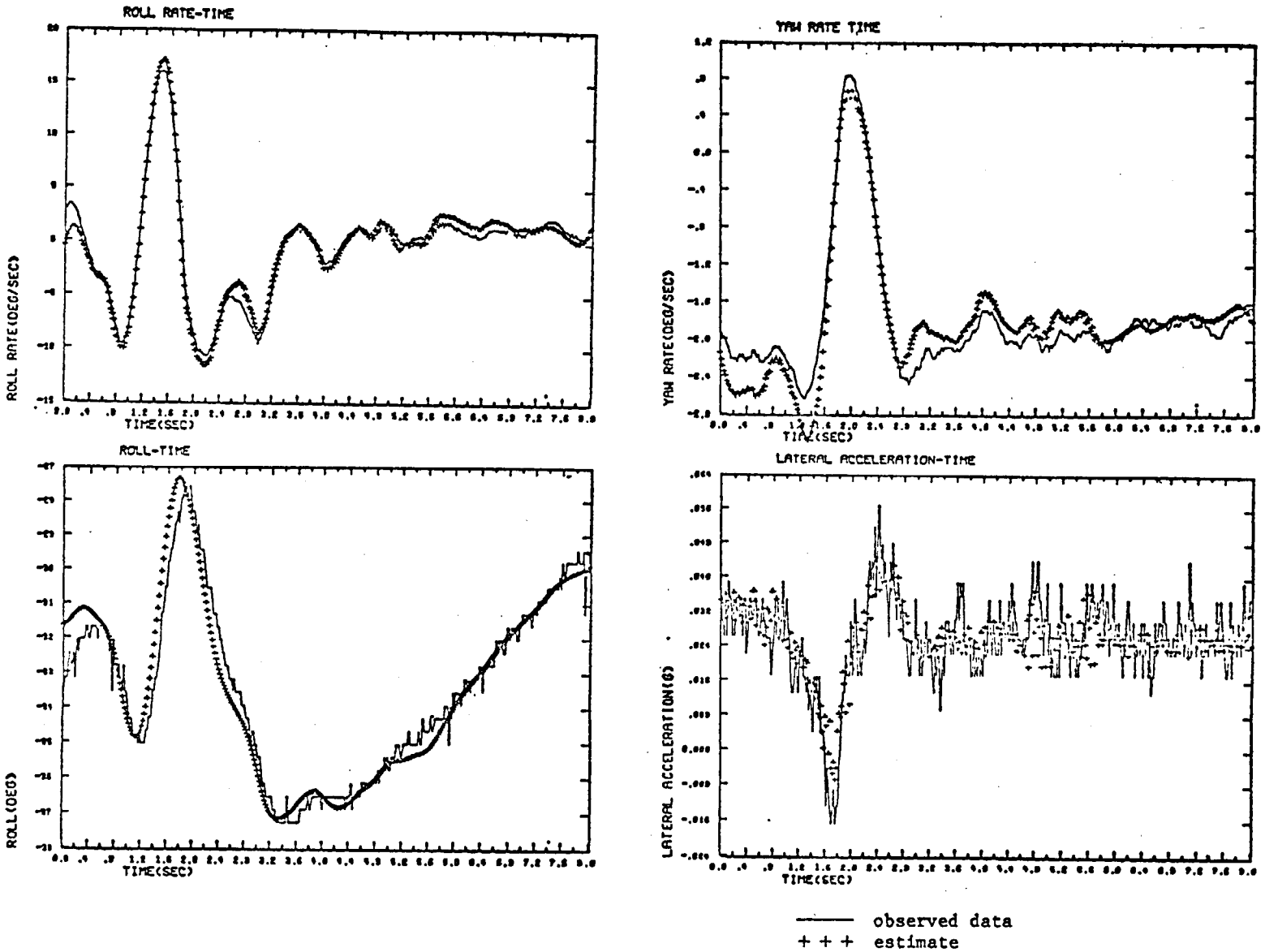


FIGURE 5.20 M2/F3 TIME HISTORIES WITH RANK DEFICIENT SOLUTION

VI

BACKGROUND FOR LINEAR SYSTEM INPUT DESIGN

The importance of choosing appropriate inputs (i.e., control surface deflections) for exciting specific modes of an aircraft or executing specific maneuvers has been recognized for a long time. Several considerations which enter into the selection of inputs for an aircraft are:

1. Pilot Acceptability - The inputs should be capable of being implemented easily by a pilot and the resulting response of the aircraft should not endanger pilot safety.
2. Parameter Sensitivity - The measured response of the aircraft should be sensitive to the parameters that are being identified. This is necessary for obtaining good estimates of the parameters from the flight test data during the inverse computation or the identification process.
3. Instrumentation Limitations - The dynamic range of the instruments and their signal-to-noise characteristics impose limitations on the types and magnitudes of aircraft maneuvers. The relationship between input design and instrumentation specification has been emphasized in (Ref. 40).
4. Derivative Extraction Method - In the past, the choice of control inputs has often been dictated by the desire to use a particular method for derivative extraction. For example, sinusoidal inputs were used initially to obtain the transfer function of an aircraft at specified frequencies (Ref. 40). However, it was soon realized that this was very expensive in terms of the total flight test time required to obtain the aircraft stability and control derivatives.(Ref. 40) Next, the step and the doublet type of inputs were used and specialized

methods such as Prony's Method (Ref. 41) and the Time Vector Method(Ref. 1) were devised to extract derivatives. With the more powerful digital techniques available today such as the Newton-Raphson (Ref. 8) and the Maximum Likelihood Methods, (Ref. 33) arbitrary inputs can be handled and it is no longer necessary to limit the inputs for the success of the derivative extraction method.

5. Modeling Assumptions - The six-degree-of-freedom equations of motion and the nonlinear aerodynamic model for an aircraft contain a large number of parameters (over 200). The simultaneous estimation of all these parameters from a single maneuver is not attempted since this would lead to nonuniqueness and identifibility problems. Generally, linearized decoupled equations of motion are used for the extraction of longitudinal and lateral stability and control derivatives. The inputs selected for exciting these modes should be such that the assumptions of linearity and decoupling are not violated. The inputs currently in use are mostly of the doublet type. The resulting aircraft response is an impulse-type of response about a given trim condition. Generally no attempt is made to optimize the frequency, the shape, or the timing of the impulses in order to make the aircraft response sensitive to the parameters that are being identified.

The motivation for the present study comes from a simulation of the X-22 VTOL Aircraft perfomed in 1970, and described in Section 5.1.4. The multistep input gave parameter estimates which are an order of magnitude more accurate than the estimates obtained using the Cornell input. At about the same time, one of

the authors developed a general theory of optimal input design based on the techniques of modern control theory. (Ref. 42) In the following section, the salient features of the theory and the computation aspects of input design are presented. The results of applying it to the C-8 aircraft parameter identification problem are considered in section VIII.

6.1 Related Work on Input Design in System Identification

The importance of input selection for system identification has also been recognized for a long time, though a unified mathematical treatment has emerged only recently. Some of the earlier attempts at input design were based on frequency domain methods and engineering judgment. A large amount of literature exists on Pseudo Random Binary Sequence (PRBS) inputs which have been found to provide improved identification for a large number of systems. (Ref. 43-45) However, PRBS inputs use very little information about the known properties of the system. Since in a number of physical systems some a priori information is available about the modes of the system (e.g., short period mode, phugoid mode, etc. of an aircraft's longitudinal motion), one can use this information to design inputs for identifying these modes more precisely.

The work described in this report is most closely related to that of Aoki and Staley, (Ref. 46) Levadi, (Ref. 47) Nahi and Wallis, (Ref. 48) and Levin (Ref. 49) on input signal design for system identification and to that of McAulay (Ref. 50) and Esposito (Ref. 51) for signal synthesis. Aoki and Staley (Ref. 46) consider single-input, single-output discrete-time systems. Levadi's results (Ref. 47) are only applicable to the case in which the unknown parameters enter linearly in the system impulse response. Levine's results (Ref. 49) are applicable when linear regression is used to estimate the unknown parameters.

The results presented here are applicable to multi-input, multi-output, continuous time systems. The computational algorithms proposed are new and and have not been used earlier for input design purposes.

VII

THEORY OF INPUT DESIGN FOR LINEAR SYSTEM IDENTIFICATION*

The problem of input design for linear system identification is formulated here as an optimal control problem. The performance criterion used is the sensitivity of the system response to the unknown parameters. The other criteria for input design such as pilot acceptability, instrumentation characteristics and state deviations described in Section VI are considered indirectly through an energy or power type of constraint on the input and through modifications of the final results. It is assumed that an output error or maximum likelihood method which can handle arbitrary inputs is used for derivative extraction. In these methods, the measured response of the system z(t) is expanded in terms of the system parameters as follows:

$$z(t) = y(\theta,t) + n(t) \tag{7.1}$$

where $y(t,\theta)$ is the true system response, θ is an Nxl vector of unknown parameters and n(t) represents measurement noise. Let θ_0 be the best a priori guess of θ. By a Taylor series expansion of Equation (7.1),

$$z(t) = y(\theta_o, t) + \nabla_\theta y(\theta_o, t)\,(\theta - \theta_o) + n(t) + \text{higher order terms} \tag{7.2}$$

where $\nabla\theta = \left(\frac{\partial}{\partial\theta_1}, \frac{\partial}{\partial\theta_2}, \ldots \frac{\partial}{\partial\theta_N}\right)$

In the output error method, the step $(\theta-\theta_0)$ is determined by a least squares solution of Eq. (7.2) over the time interval [0,T]. It is intuitively clear that the sensitivity function $\nabla_\theta\, y(\theta_0,t)$ must be sufficiently large to achieve high accuracy in determining the parameter step $(\theta-\theta_0)$. Furthermore, the sensitivity functions must not be linearly dependent over $0 \leq t \leq T$ for a unique solution of $(\theta-\theta_0)$.

---

* The theoretical aspects of this work were supported in part under AFOSR contract No. F44620-71-C-0077. See "Dual Control and Identification Methods for Avionic Systems - Part II, Optimal Input Design for Linear System Identification".

For input design, we now define a scalar performance index in terms of the sensitivity functions. For a single parameter, a suitable performance index is the weighted mean square value of the sensitivity function over the time interval [0,T].

$$J_1 = \sum_{i=1}^{P} \frac{1}{\sigma_i^2} \int_0^T \left(\frac{\partial y_i}{\partial \theta_1} \right)^2 dt \tag{7.3}$$

where σ_i^2 is the variance of noise in the measurement y_i. The performance index J_1 favors the measurements which are more accurate and makes them more sensitive to θ_1 compared with the mesurements that are less accurate. It can also be shown that J_1 represents the information in the measured response about the parameter θ_1.

For the multiparameter case, the choice of a scalar performance index is much more complicated due to the conflicting nature of the individual sensitivity functions. One possible choice is a weighted sum of the individual sensitivity measures.

$$J = \sum_{j=1}^{N} w_j \; J_j \tag{7.4}$$

where the weighting coefficients w_j reflect the relative importance of parameters. It is clear that the criterion (7.4) can also be written as the weighted trace of the Fisher Information Matrix M.

$$J = \text{tr} \; (WM) \tag{7.5}$$

where $W = \begin{bmatrix} w_1 & & 0 \\ & w_2 & \\ & \ddots & \\ 0 & & w_N \end{bmatrix}$ (7.6)

and $M = \int_0^T \left(\frac{\partial y}{\partial \theta}\right)^T R^{-1} \frac{\partial y}{\partial \theta} \, dt$ (7.7)

$R = \begin{bmatrix} \sigma^2 & & 0 \\ & \sigma_2^2 & \\ & \ddots & \\ 0 & & \sigma_N^2 \end{bmatrix}$ (7.8)

Other criteria such as the weighted trace of M^{-1}, the smallest eigenvalue of M or the determinant of M lead to nonlinear optimization problems which are much more difficult to solve.

An alternative method for choosing a criterion function is to consider the objective of identification. For example, if the identified parameters are to be used for control system design, the inputs should be selected so as to minimize the control error. Similarly, if the parameters are used for response prediction, the input design should be based on minimizing the mean square response error. However, these criteria generally lead to optimization problems which are mathematically intractable or extremely difficult to solve. It is also felt that the advantage gained by solving these difficult problems may be more than offset by the basic uncertainty about the initial parameter values θ_0. The input design process described here should be viewed as a sequential design process in which the inputs are selected based on the current best knowledge of the parameters.

7.1 Problem Formulation

Consider the linearized aircraft equations of motion

$$\dot{x} = Fx + Gu \tag{7.9}$$

where x is nxl state vector and u is a mxl control vector. F and G are n x n and n x m matrices of unknown parameters (stability and control derivatives). The measurements are denoted by a p x 1 vector z(t) which is contaminated with white noise n(t)

$$z(t) = Hx(t) + n(t). \tag{7.10}$$

H is a p x n matrix and n(t) is a zero mean Gaussian white noise process

$$E\{n\} = 0, \quad E\{n(t)n^T(\tau)\} = R\delta(t - \tau). \tag{7.11}$$

Let θ denote the N x 1 vector of unknown parameters in the above equations. It is required to select the input $\{u(t), \ 0 \leq t \leq T\}$ to maximize the weighted trace of the information matrix M subject to an energy type constraint. The information matrix M for the unknown parameter set θ can be written as

$$M = \int_0^T (\nabla_\theta x)^T H^T R^{-1} H(\nabla_\theta x)\, dt \tag{7.12}$$

The energy constraint on u(t) is

$$\int_0^T u^T u \, dt = E \tag{7.13}$$

The performance index J can be written as,

$$J = \mathrm{tr}\{WM\} = \mathrm{tr}\{W^{1/2} M W^{1/2}\} \tag{7.14}$$

$$= \mathrm{tr}\left\{\int_0^T (\nabla_\theta x \cdot W^{1/2})^T H^T R^{-1} H(\nabla_\theta x \cdot W^{1/2})\, dt\right\}$$

In order to use this criterion, the quantity $(\nabla_\theta x)$ must be calculated as a function of time. A differential equation for $(\nabla_\theta x)$ can be easily found from the system equation. Since the multiplication by $W^{1/2}$ represents a column operation, this is equivalent to calculating successive equations for

$$w_i^{1/2} \frac{\partial x}{\partial \theta_i} \quad , \quad i = 1,\ldots,N \quad .$$

This can be accomplished by the following:

$$\left(w_i^{1/2} \frac{\partial x}{\partial \theta_i}\right)^{\bullet} = \left(w_i^{1/2} \frac{\partial F}{\partial \theta_i}\right)x + F\left(w_i^{1/2} \frac{\partial x}{\partial \theta_i}\right) + \left(w_i^{1/2} \frac{\partial G}{\partial \theta_i}\right)u \tag{7.15}$$

with $\dot{x} = Fx + Gu$.

An equivalent way of formulating the problem which makes use of the simultaneous computation of $w_i^{1/2}\, \partial x/\partial \theta_i$ involves the specification of augmented F_A, G_A, and H_A matrices and x_A vector. These are defined as follows (with m the number of inputs and p the number of outputs):

$$F_A = \begin{bmatrix} F & 0 & \cdots & 0 \\ w_1 \frac{\partial F}{\partial \theta_1} & F & & 0 \\ \vdots & \vdots & & \vdots \\ w_N \frac{\partial F}{\partial \theta_N} & 0 & & F \end{bmatrix} \qquad G_A = \begin{bmatrix} G & & \\ w_1 \frac{\partial G}{\partial \theta_1} & & \\ \vdots & & \\ w_N \frac{\partial G}{\partial \theta_N} & & \end{bmatrix} \tag{7.16}$$

$(N+1)n \times (N+1)n$ $\qquad$ $(N+1)n \times m$

$$H_A = \begin{bmatrix} 0 & H & 0 & \cdots & 0 \\ 0 & 0 & H & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & H \end{bmatrix}, \quad R_A = \begin{bmatrix} R & 0 & \cdots & 0 \\ 0 & R & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R \end{bmatrix} \tag{7.17}$$

H_A: Np x (N+1)n; R_A: Np x Np

and

$$x_A = \begin{bmatrix} x \\ w_1 \frac{\partial x}{\partial \theta_1} \\ w_2 \frac{\partial x}{\partial \theta_2} \\ \vdots \\ w_N \frac{\partial x}{\partial \theta_N} \end{bmatrix} \tag{7.18}$$

(N+1)n x 1

With these definitions and using Eq. (92), it is possible to write:

$$\dot{x}_A = F_A x_A + G_A u \tag{7.19}$$

The performance criterion is now redefined as

$$J = \int_o^T x_A^T H_A^T R_A^{-1} H_A x_A \, dt \tag{7.20}$$

In the next section, we obtain a set of necessary conditions for the optimal input using the Pontryagin's Maximum Principle (Ref. 52).

7.2 Optimal Engergy-Constrained Input Using Maximum Principle

Let $\lambda(t)$ denote the costate vector for $x_A(t)$ and μ be the constant multiplier associated with the constraint (7.13). The Hamiltonian of the augmented system is,

$$\mathcal{H} = \frac{1}{2}\left[-(x_A)^T H_A{}^T R_A^{-1} H_A(x_A) + \mu(u^T u - \frac{E}{T})\right] + \lambda^T[F_A x_A + G_A u] \qquad (7.21)$$

the necessary conditions of optimality are

$$\dot{\lambda} = -\left(\frac{\partial \mathcal{H}}{\partial x_A}\right)^T$$

or
$$\dot{\lambda} = -F_A^T\lambda + H_A^T R_A^{-1} H_A x_A \qquad (7.22)$$

and $\mathcal{H}_u = 0$

or
$$u^* = -\frac{1}{\mu}(G_A)^T\lambda \qquad (7.23)$$

The boundary conditions are homogeneous.

$$x_A(0) = 0 \ , \ \lambda(T) = 0$$

Substituting for u^* in Eq. 7.19, we obtain the two point boundary value problem,

$$\frac{d}{dt}\begin{bmatrix} x_A \\ \lambda \end{bmatrix} = \begin{bmatrix} F_A & -\frac{1}{\mu} G_A G_A{}^T \\ H_A^T R_A^{-1} H_A & -F_A^T \end{bmatrix}\begin{bmatrix} x_A \\ \lambda \end{bmatrix} \qquad (7.24)$$

Since the boundary conditions are homogeneous, the solution is trival viz. $x_A \equiv 0$, $\lambda \equiv 0$, $u \equiv 0$ except for certain values of μ which are the eigenvalues of the two point boundary value problem. In other words, the problem is of the Sturm-Liouville type.(Ref. 52) The eigenvalues and the optimal input can be determined in a number of ways. Two possible methods are (i) the transition matrix method and (ii) the Riccati equation method.

7.2.1 Transition Matrix Method

Let $\Phi(t,0;\mu)$ denote the transition matrix of 7.24 for a particular μ.

$$\Phi(t,0;\mu) = \exp\begin{pmatrix} F_A & -\frac{1}{\mu} G_A^T G_A \\ H_A^T R_A^{-1} H_A & -F_A^T \end{pmatrix} t \tag{7.25}$$

Partition $\Phi(t,0;\mu)$ into X_A and λ parts as follows:

$$\Phi = \begin{bmatrix} \Phi_{xx} & \Phi_{x\lambda} \\ \Phi_{\lambda x} & \Phi_{\lambda\lambda} \end{bmatrix} \tag{7.26}$$

Then

$$\begin{bmatrix} x_A(t) \\ \lambda(T) \end{bmatrix} = \begin{bmatrix} \Phi_{xx}(T,0;\mu) & \Phi_{x\lambda}(T,0;\mu) \\ \Phi_{\lambda x}(T,0;\mu) & \Phi_{\lambda\lambda}(T,0;\mu) \end{bmatrix} \begin{bmatrix} x_{A(o)} \\ \lambda(0) \end{bmatrix} \tag{7.27}$$

The second equation in (7.27)along with the boundary conditions gives

$$\lambda(T) = \Phi_{\lambda\lambda}(T,0:\mu)\ \lambda(0) = 0 \tag{7.28}$$

For a nontrivial solution

$$|\Phi_{\lambda\lambda}(T,0;\mu)| = 0 \tag{7.29}$$

Eq.(7.29) is the eigenvalue equation for the Hamiltonian system (7.24). It is a nonlinear algebraic equation in μ and can be solved by a Newton-Raphson iteration. In general, there is an infinite set of eigenvalues, but we will be only interested in the largest value of μ which will be shown to maximize J (Section 7.3).

7.2.2 Riccati Equation Method

The eigenvalues μ are functions of the interval length T. Therefore, one can fix μ and determine T for which $\Phi_{\lambda\lambda}(T,0;\mu)$ becomes singular. Another way is to use the Riccati matrix P(t) defined by the relationship

$$x_A(t) = P(t)\lambda(t) \tag{7.30}$$

An equation for P(t) is obtained by differentiating both sides of Eq. (7.30) and substituting from Eq. (7.24).

$$\dot{x}_A = \dot{P}\lambda + P\dot{\lambda}$$

or

$$[F_A P - \frac{1}{\mu} G_A G_A^T]\lambda = [\dot{P} + PH_A^T R_A^{-1} H_A P - PF_A^T]\lambda$$

or

$$\dot{P} = F_A\dot{P} + PF_A^T - PH_A^T R_A^{-1} H_A P - \frac{1}{\mu} G_A G_A^T \tag{7.31}$$

$$P(0) = 0. \tag{7.32}$$

The Riccati Eq. (7.31) differs from the usual Riccati-equation of the Linear-Quadratic problem in that the forcing term (last term) in Eq. (7.31) enters negatively. Eq. (7.30) can also be written as

$$\lambda(t) = P^{-1}(t)\, x_A(t)$$

whenever P^{-1} exists. At final time $t = T$, since $\lambda(T) = 0$,

$$P^{-1}(T) = 0. \tag{7.33}$$

which means that a conjugate point exists at $t = T$.

Eq. (7.33) provides us with a method to obtain the critical interval length T corresponding to an eigenvalue μ. The Riccati Eq. (7.31) is integrated forward in time for a particular μ using initial conditions (7.32). When the elements of $P(t)$ become very large, the critical length T corresponding to an eigenvalue is being reached. Now $P^{-1}(t)$ is integrated using the equation

$$\frac{d}{dt}(P^{-1}) = - P^{-1} \dot{P} P^{-1}$$

or

$$\frac{d}{dt}(P^{-1}) = - P^{-1}F_A - F_A^T P^{-1} + H_A^T R_A^{-1} H_A + \frac{1}{\mu} P^{-1} G_A G_A^T P^{-1} \tag{7.34}$$

At the critical interval length T, all the elements of P^{-1} go to zero. It follows from the Sturmian property (Ref. 52) that the smallest T corresponds to the largest eigenvalue μ.

After the critical length T corresponding to the largest value of μ has been determined, Eq. (7.24) is solved forward in time using $\lambda(0)$ obtained from Eq. (7.28) and (7.29) as an eigenvector of $\Phi_{\lambda\lambda}(T,0;\mu)$ corresponding to

the zero eigenvalue. Thereby, the boundary condition $\lambda(T) = 0$ is automatically satisfied. A unique value of $\lambda(0)$ is found by using the normalization condition of Eq. (7.13).

7.3 Application of Functional Analysis

In the last section, the optimal input u was characterized in terms of the solution to a two point boundary value problem. In this section, we show that the optimal u is an eigenfunction of a positive self-adjoint operator corresponding to the largest eigenvalue μ.

Let A denote the operator corresponding to Eq. (7.19) viz.

$$A[u] = \int_0^t e^{F_A(t-\tau)} G_A \, u(\tau) d\tau \tag{7.35}$$

Let $A^*[\cdot]$ denote the adjoint operator to $A[\cdot]$

$$A^*[w] = G_A^T \int_t^T e^{F_A^T(s-t)} w(s) ds \tag{7.36}$$

Let $\langle u,w \rangle$ denote the inner product

$$\langle u,w \rangle = \int_0^T u^T(t) w(t) dt \tag{7.37}$$

The performance index J can be written as

$$\begin{aligned} J &= \langle x_A \; H_A^T R_A^{-1} H_A \; x_A \rangle \\ &= \langle Au, \; H_A^T R_A^{-1} H_A Au \rangle \\ &= \langle u, \; A^* H_A^T R_A^{-1} H_A Au \rangle \end{aligned} \tag{7.38}$$

The energy constraint of Eq. (7.13) is written as

$$\langle u,u \rangle = E$$

It is well known that J is maximized subject to the above constraint by u^* which is an eigenfunction corresponding to the largest eigenvalue of the operator $A^* H_A^T R_A^{-1} H_A A$. Furthermore, since $A^* H^T R^{-1} H_A A$ is a positive self-adjoint operator, all its eigenvalues are real and positive (Ref. 52). For finite T, the operator is also compact and has a finite maximum eigenvalue. The optimal u^* is the eigenfunction of $A^* H_A^T R_A^{-1} H_A A$ corresponding to this eigenvalue and normalized according to $\langle u,u \rangle = E$.

$$A^* H_A^T R_A^{-1} H_A A\, u = \mu u \tag{7.39}$$

Also,

$$\max_u J = \mu E \tag{7.40}$$

To show the relationship of the above eigenvalue problem with the two point boundary value of Eq. (7.24), define

$$z = Au \tag{7.41}$$

and

$$\eta = \int_t^T e^{F_A^T (s-t)}\, w(s)\, ds \tag{7.42}$$

Then, using the definition of A and A^*,

$$\dot{z} = F_A z + G_A u\ , \quad z(0) = 0 \tag{7.43}$$

$$\dot{\eta} = -F_A^T \eta + H_A^T R_A^{-1} H_A z, \quad \eta(T) = 0 \tag{7.44}$$

From Eq. (7.39) and (7.42)

$$G_A{}^T\eta = A^* H_A^T R_A^{-1} H_A A u = \mu u \tag{7.45}$$

or

$$u = -\frac{1}{\mu} \quad G_A{}^T\eta \tag{7.46}$$

Therefore

$$\left.\begin{aligned} \dot{z} &= F_A z - \frac{1}{\mu} G_A G_A{}^T \eta \ , \quad z(0) = 0 \\ \dot{\eta} &= -F_A^T \eta + H_A^T R_A^{-1} H_A z \ , \quad \eta(T) = 0 \end{aligned}\right\} \tag{7.47}$$

A comparison of Eq. (7.46) and (7.47) with Eq. (7.24) shows that

$$z = x_A$$

$$\eta = \lambda.$$

7.4 Examples

We now apply the above results to two first order and a second order example.

7.4.1 First Order System with Unknown Gain

Consider the system

$$\dot{x} = -x + \theta u \tag{7.48}$$

where x and u are scalars and θ is the unknown gain.

$$y = x + v \ , \tag{7.49}$$

where $E\{v\} = 0$, $E\{v(t)v(\tau)\} = r\,\delta(t - \tau)$.

From Eq. (7.23) and (7.24)

$$u = -\frac{1}{\mu}\lambda \tag{7.50}$$

$$\frac{d}{dt}\begin{bmatrix} \nabla_\theta x \\ \lambda \end{bmatrix} = \begin{bmatrix} -1 & -\frac{1}{\mu} \\ \frac{1}{r} & 1 \end{bmatrix}\begin{bmatrix} \nabla_\theta x \\ \lambda \end{bmatrix} \tag{7.51}$$

Eigenvalues

Equation 7.51 gives

$$\dot{\lambda} = \frac{1}{r}\nabla_\theta x + \lambda$$

$$\ddot{\lambda} = \frac{1}{r}\left[-\nabla_\theta x - \frac{1}{\mu}\lambda\right] + \left[\frac{1}{r}\nabla_\theta x + \lambda\right]$$

$$= \left(1 - \frac{1}{\mu r}\right)\lambda \quad . \tag{7.52}$$

Let $\alpha^2 = 1 - 1/\mu r$. Three cases arise: $\alpha^2 > 0$, $\alpha^2 = 0$ and $\alpha^2 < 0$. It can be easily shown that only $\alpha^2 < 0$ leads to a nontrivial solution.

Let

$$w = \left(\frac{1}{\mu r} - 1\right)^{\frac{1}{2}} \tag{7.53}$$

Then

$$\lambda(t) = C_1 \sin wt + C_2 \cos wt$$

$$\nabla_\theta x(t) = r[C_1 w \cos wt - C_2 w \sin wt - C_1 \sin wt - C_2 \cos wt]$$

$$\nabla_\theta x(0) = 0 \Rightarrow C_1 w - C_2 = 0$$

$$\lambda(T) = 0 \Rightarrow C_1[\sin wT + w \cos wT] = 0$$

For nontrivial solution,

$$\tan wT = -w \quad . \tag{7.54}$$

Eq.(7.54) is the eigenvalue equation. The smallest root w_o of Eq.(7.54) corresponds to the largest value of μ. The optimal input u is obtained from Eq. (7.50) as

$$u^* = -\frac{C_1}{\mu_o}[\sin w_o t + w_o \cos w_o t] \tag{7.55}$$

where $\mu_o = 1/r(1 + w_o^2)$. Notice that u satisfies the same second order differential equation as λ viz. Eq.(7.52). C_1 is determined from the condition $\int_o^T u^2 dt = E$.

This gives

$$C_1 = \sqrt{\frac{E}{T/2 - 1/4 \sin 2T}} \tag{7.56}$$

7.4.2 Levadi's Example

Levadi considers the following example in his paper (Ref. 47).

$$\dot{x} = -x + bu \tag{7.57}$$

$$y = x + v \tag{7.58}$$

where v is a correlated noise process with autocorrelation function.

$$E\{v(t)v(\tau)\} = c\, e^{-a|t-\tau|}$$

It is required to estimate b only.

Levadi's (Ref. 47) results can be easily derived as follows:

v(t) can be represented as

$$\dot{v} = -av + \varepsilon \tag{7.59}$$

where ε is a white noise process and

$$E\{\varepsilon\} = 0 \quad , \quad E\{\varepsilon(t)\varepsilon(\tau)\} = 2ac\ \delta(t - \tau)$$

A new measurement can be generated by differentiating Eq.(7.58). (This procedure is similar to that of Bryson and Johansen, Ref. 53).

$$\dot{y} = \dot{x} + \dot{v}$$

$$= (a - 1)x + bu - ay + \varepsilon$$

Now $\dot{y}$ can be regarded as a new measurement which has white noise ε in it. The new information matrix is:

$$M = \int_0^T \frac{1}{2ac} [(a - 1)\nabla_b x + u]^2 dt \tag{7.60}$$

where

$$\dot{\nabla}_b x = -\nabla_b x + u \tag{7.61}$$

Now the problem is in the same form as example 7.4.1 except that the performance index is slightly different.

The following equations of optimality are easily derived.

$$u = -\frac{1}{2(\mu - \frac{1}{2ac})}\left[\lambda - \frac{a-1}{ac}\nabla_b x\right] \tag{7.62}$$

$$\dot{\lambda} = \frac{(a-1)^2}{ac}\nabla_b x + \frac{a-1}{ac}u + \lambda \quad . \tag{7.63}$$

An equation only in terms of u can also be obtained from Eq. (7.61) - (7.63)

$$\ddot{u} - \left[1 - \frac{a^2-1}{2\mu ac - 1}\right] u = 0 \quad . \tag{7.64}$$

The eigenvalue equation is

$$\tan(wT + \emptyset) = \frac{w}{a} \tag{7.65}$$

$$\text{where } w = \left[-1 + \frac{a^2-1}{2\ ac - 1}\right]^{1/2} \tag{7.66}$$

$$\emptyset = \tan^{-1} w \quad .$$

$$\text{The optimal input } u^* = A \sin(wt + \emptyset) \quad . \tag{7.67}$$

Notice that the results for example 7.4.1 can be obtained by letting $a \to \infty$ and $2c/a \to r$, where $2c/a$ represents the area under the autocorrelation function of v.

The optimum value of w is chosen to maximize μ. From Eq. (7.66),

$$\mu = \frac{1}{2ac}\left[1 + \frac{a^2-1}{1+w^2}\right] \quad . \tag{7.68}$$

It is seen from Eq.(7.68) that when $a^2 > 1$, the maximum of μ is attained for the smallest value of w. This corresponds to the case when the noise is wide-band. For the narrow band noise case viz. $a^2 < 1$, the second term in Eq. (7.68) is negative and the maximum of μ viz. $1/2ac$ is reached at $w = \infty$. The

practical implication of this result is that the input should be of as high frequency as possible. Since the noise is narrow band, this increases the high frequency signal to noise ratio at the output.

7.4.3 Second Order Example

The following system represents the short-period longitudinal dynamics of an aircraft.

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}a_1 & 1\\ a_2 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} + \begin{bmatrix}b_1\\ b_2\end{bmatrix}\delta_e \tag{7.69}$$

$$y = x_1 + v$$

x_1 represents pitch rate, x_2 is a linear combination of pitch rate and angle-of-attack and δ_e is the elevator deflection. The transfer function of the system from elevator to pitch rate is

$$\frac{x_1(s)}{\delta_e(s)} = \frac{b_1 s + b_2}{s^2 - a_1 s - a_2} \quad . \tag{7.70}$$

Optimal Input for Identifying b_1:

Sensitivity equations are

$$\begin{bmatrix}\nabla_{b_1}\dot{x}_1\\ \nabla_{b_1}\dot{x}_2\end{bmatrix} = \begin{bmatrix}a_1 & 1\\ a_2 & 0\end{bmatrix}\begin{bmatrix}\nabla_{b_1}x_1\\ \nabla_{b_1}x_2\end{bmatrix} + \begin{bmatrix}1\\ 0\end{bmatrix}\delta_e \tag{7.71}$$

Optimal input, $\delta_e = -\lambda_1/2\mu$

$$\dot{\lambda}_1 = \frac{2}{r}\nabla_{b_1} x_1 - \lambda_1 a_1 - \lambda_2 a_2 \tag{7.72}$$

$$\dot{\lambda}_2 = -\lambda_1 \quad . \tag{7.73}$$

Both the optimal input δ_e and λ_2 satisfy fourth order linear differential equations of the form

$$[D^4 + (\frac{1}{\mu r} - 2a_2 - a_1^2)D^2 + a_2^2]\lambda_2 = 0 \qquad (7.74)$$

where D denotes the differential operator d/dt.

A solution for λ_2 is easily written as

$$\lambda_2(t) = C_1 \sin w_1 t + C_2 \cos w_1 t + C_3 \sin w_2 t + C_4 \cos w_2 t \qquad (7.75)$$

where

$$w_1 = \frac{1}{\sqrt{2}} \sqrt{\sigma - \sqrt{\sigma^2 - 4a_2^2}}$$

$$w_2 = \frac{1}{\sqrt{2}} \sqrt{\sigma + \sqrt{\sigma^2 - 4a_2^2}}$$

$$\sigma = (\frac{1}{\mu r} - 2a_2 - a_1^2) = w_1^2 + w_2^2 \quad . \qquad (7.76)$$

Also $w_1 w_2 = a_2$.

It is assumed here that $\sigma > 2a_2$ or $1/\mu r > (4a_2 + a_1^2)$ since other cases lead to hyperbolic functions which become unbounded for large T. They are rejected as possible solutions using the same reasoning as used in examples 7.4.1 and 7.4.2.

The expressions for $\lambda_1(t)$, $\nabla_{b_1} x_1$, u(t) and $\nabla_{b_2} x_2$ are easily obtained from Eq.(7.75) using Eq.(7.71)-(7.73). The eigenvalue equation, assuming $w_1 \neq w_2$, is obtained as

$$\begin{vmatrix} \sin w_1 T & \cos w_1 T & \sin w_2 T & \cos w_2 T \\ -w_1 \cos w_1 T & w_1 \sin w_1 T & -w_2 \cos w_2 T & w_2 \sin w_2 T \\ -w_1 a_1 & a_2 + w_1^2 & -a_1 w_2 & a_2 + w_2^2 \\ w_1^3 - w_1 a_2 - \frac{w_1}{\mu r} & w_1^2 a_1 & w_2^3 - w_2 a_2 - \frac{w_2}{\mu r} & w_2^2 a_1 \end{vmatrix} = 0 \qquad (7.77)$$

where $w_2 = -a_2/w_1$. Equation (7.76) is used along with (7.77) to select w_1 and w_2 which maximize μ. From Eq. (7.76)

$$\frac{1}{\mu r} = (w_1 + \frac{a_2}{w_1})^2 + a_1^2 = (w_1 - w_2)^2 + a_1^2 \quad .$$

The minimum of $1/\mu r$ is attained at $w_1^2 = -a_2$ or $w_1 = w_2$, i.e., at the undamped natural frequency of the system. However since $w_1 = w_2$ is ruled out by the solution considered here, the root of Eq. (7.77) closest to the undamped natural frequency should be chosen. Since $w_1 w_2 = -a_2$, the two frequencies will bracket the natural frequency of the system.

The optimal input δ_e^* is

$$\delta_e^* = \frac{1}{2\mu} [C_1 w_1 \cos w_1 t - C_2 w_1 \sin w_1 t + C_3 w_2 \cos w_2 t - C_4 w_2 \sin w_2 t] \qquad (7.78)$$

C_1, C_2, C_3, and C_4 are determined as the eigenvector of Eq. (7.77) corresponding to the root w_1. They are normalized using the condition

$$\int_o^T \delta_e^2 \, dt = E \quad .$$

7.5 Extension to Systems with Process Noise

Consider the linearized aircraft equations of motion

$$\dot{x} = Fx + Gu + \Gamma n \qquad (7.79)$$

$$Z = H'x + v$$

where $\eta(t)$ is a gaussian white noise forcing function representing random gusts and modeling errors,

$$E\{\eta(t)\} = 0 \quad , \quad E\{\eta(t)\eta^T(\tau)\} = Q\delta(t - \tau) \quad .$$

The information matrix M in this case is given in terms of the Kalman filter for the above system.

$$\dot{\hat{x}} = F\hat{x} + Gu + K(y - H\hat{x}) \tag{7.80}$$

$$K = \Sigma H^T R^{-1} \tag{7.81}$$

$$\dot{\Sigma} = F\Sigma + \Sigma F^T + \Gamma Q \Gamma - \Sigma H^T R^{-1} H \Sigma \tag{7.82}$$

where $\hat{x}$ denotes the best filtered estimate of x and Σ denotes the covariance of $\hat{x}$. The Kalman filter provides a linear causally-invertible whitening transformation for the process y since the innovation sequence $(y - H\hat{x})$ is a gaussian white noise sequence. The likelihood function is easily written in terms of the innovation sequence (Refs. 24, 27). The information matrix M is given as

$$M = \int_0^T E\{(\nabla_\theta \hat{x})^T H^T R^{-1} H (\nabla_\theta \hat{x})\} dt \tag{7.83}$$

where $\nabla_\theta \hat{x}$ denotes the sensitivity function of the filtered estimate $\hat{x}$ with respect to the unknown parameter vector θ. Note that both K and Σ are functions of θ so that the sensitivity equations are much more complicated than for the no process noise case. Moreover M, in general, depends on the random quantities η and v so that its expected value needs to be maximized.

A special case arises when θ contains parameters from G only and the initial state is known exactly. Since K and P do not depend upon G, the sensitivity equation has a simple form

$$\dot{\nabla}_\theta \hat{x} = (F - KH)\nabla_\theta \hat{x} + \nabla_\theta G u \tag{7.84}$$

K is, in general, time-varying, but if the system is completely controllable and observable, K reaches a constant steady-state value (Ref. 54). Then Eq.(7.84) is essentially similar to Eq. (7.19) and most of the theory developed in Sections 7.2 and 7.3 carries over to this case.

7.5.1 Example: Let us consider example 7.4.1 with additive process noise.

$$\dot{x} = -x + \theta u + \eta$$
$$y = x + v \tag{7.85}$$

where

$$E\{\eta\} = 0 \quad , \quad E\{\eta(t)\eta(\tau)\} = q\delta(t - \tau) \quad , \quad x(0) = 0 \quad .$$

The filter sensitivity equation for θ under steady-state filter gain, $k > 0$ is

$$\nabla_\theta \dot{\hat{x}} = -(1 + k)\nabla_\theta \hat{x} + u \quad , \quad \nabla_\theta \hat{x}(0) = 0 \quad . \tag{7.86}$$

Proceeding as in example 7.4.1, and defining

$$\omega = \left[\frac{1}{\mu r} - (1 + k)^2\right]^{1/2} \quad \text{or} \quad \mu = \frac{1}{r[\omega^2 + (1 + k)^2]} \tag{7.87}$$

it is seen that the optimal input u^* obeys Eq. (7.54)-(7.56). Notice that by increasing process noise q, the gain increases and μ decreases. Thus the information $M = \mu E$ for the same input energy E decreases. The frequencies ω, however, remains unchanged.

7.6 State-Variable Constraints

Linear state variable constraints can be handled either directly by adding a quadratic penalty function to the performance index or indirectly by adjusting the total input energy E. The examples 7.4.1-7.4.3 show that the amplitude of the input u is determined by E. Thus by adjusting E, the amplitude of the input u and the state x can be bounded. Of course, the inputs obtained in this fashion are not strictly optimal.

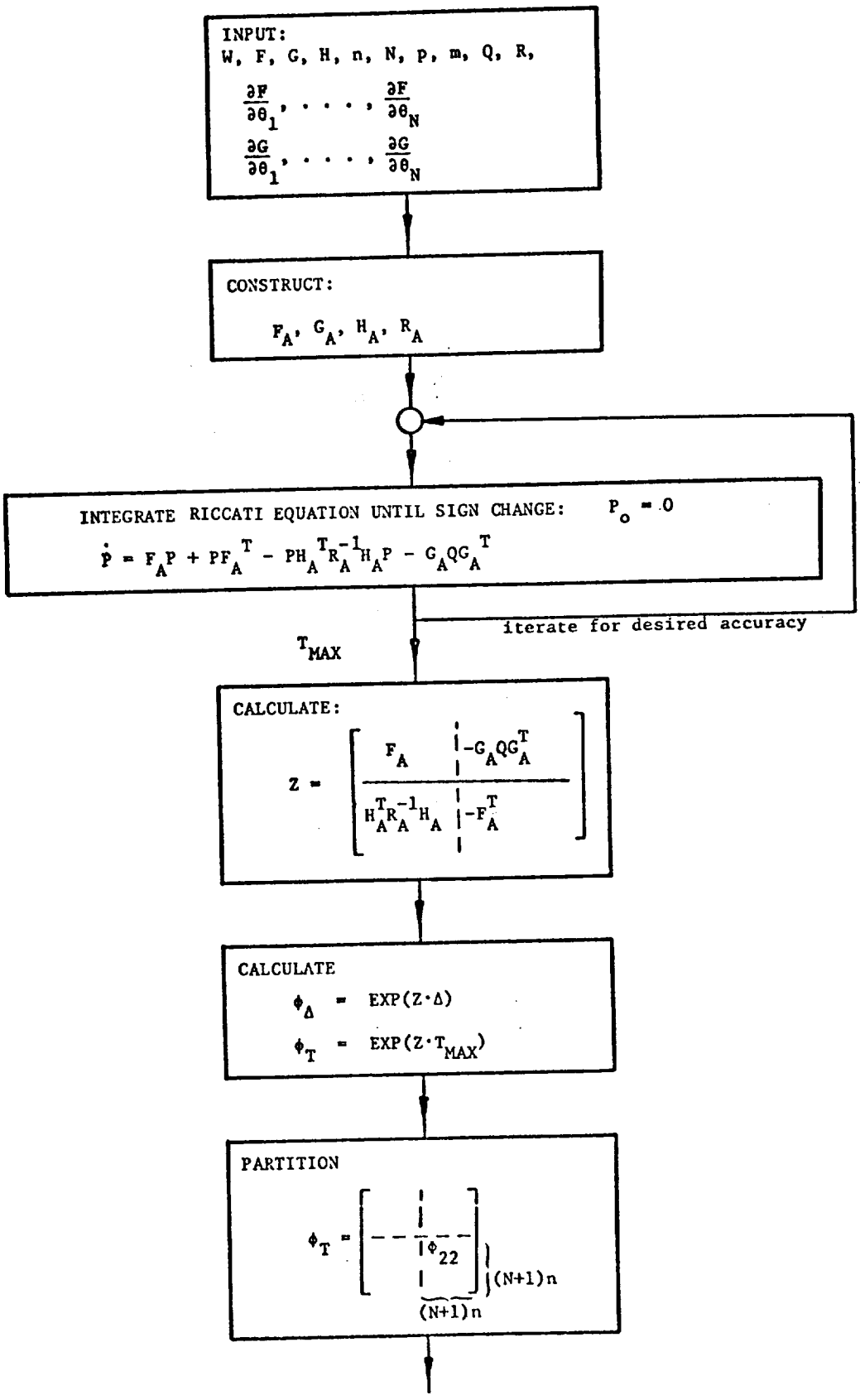
7.7 Steps in Optimal Input Program

As outlined in Sections 7.2 and 7.3, there are several possible computational techniques which can be used to solve for a consistent pair of interval length T and largest eigenvalue μ_{max}. The method that has been implemented in this program is to numerically find the first time instant at which the solution of the ricatti equation, relating $x_A(t)$ and $\lambda(t)$, becomes infinite. The instant at which this singularity occurs can be found with arbitrary accuracy by continually using a finer and finer step size and noting the instant at which the diagonal elements of the Riccati solution change sign.* This change in sign of the diagonal elements is one key indication that the ricatti solution has blown up through one direction (e.g., $-\infty$) and returned from the other (e.g. $+\infty$).

The complete flowchart of the computer program designed to calculate optimal inputs is given in Fig. 7.1. Many of the detailed steps have been combined into a single descriptive step since their description is beyond the scope of this report. For example, the actual technique used to integrate the ricatti equation is explained elsewhere (Ref. 55).

The only instant in the computational algorithm where the theoretical development was not followed exactly was in the calculation of the eigen-

*The alternative technique of calculating the largest eigenvalue of the Hamiltonian system is much more difficult since the determinant of the transition matrix for the Hamiltonian system exhibits a very sharp zero.


INPUT:
W, F, G, H, n, N, p, m, Q, R,
$\frac{\partial F}{\partial \theta_1}, \ldots, \frac{\partial F}{\partial \theta_N}$
$\frac{\partial G}{\partial \theta_1}, \ldots, \frac{\partial G}{\partial \theta_N}$
CONSTRUCT:
F_A, G_A, H_A, R_A
INTEGRATE RICCATI EQUATION UNTIL SIGN CHANGE: $P_o = 0$
$\dot{P} = F_A P + P F_A^T - P H_A^T R_A^{-1} H_A P - G_A Q G_A^T$
iterate for desired accuracy
T_{MAX}
CALCULATE:
$Z = \left[\begin{array}{c|c} F_A & -G_A Q G_A^T \\ \hline H_A^T R_A^{-1} H_A & -F_A^T \end{array}\right]$
CALCULATE
$\phi_\Delta = EXP(Z \cdot \Delta)$
$\phi_T = EXP(Z \cdot T_{MAX})$
PARTITION
$\phi_T = \left[\begin{array}{c|c} & \\ & \phi_{22} \end{array}\right]$ $\Big\}(N+1)n$
$\underbrace{\qquad}_{(N+1)n}$

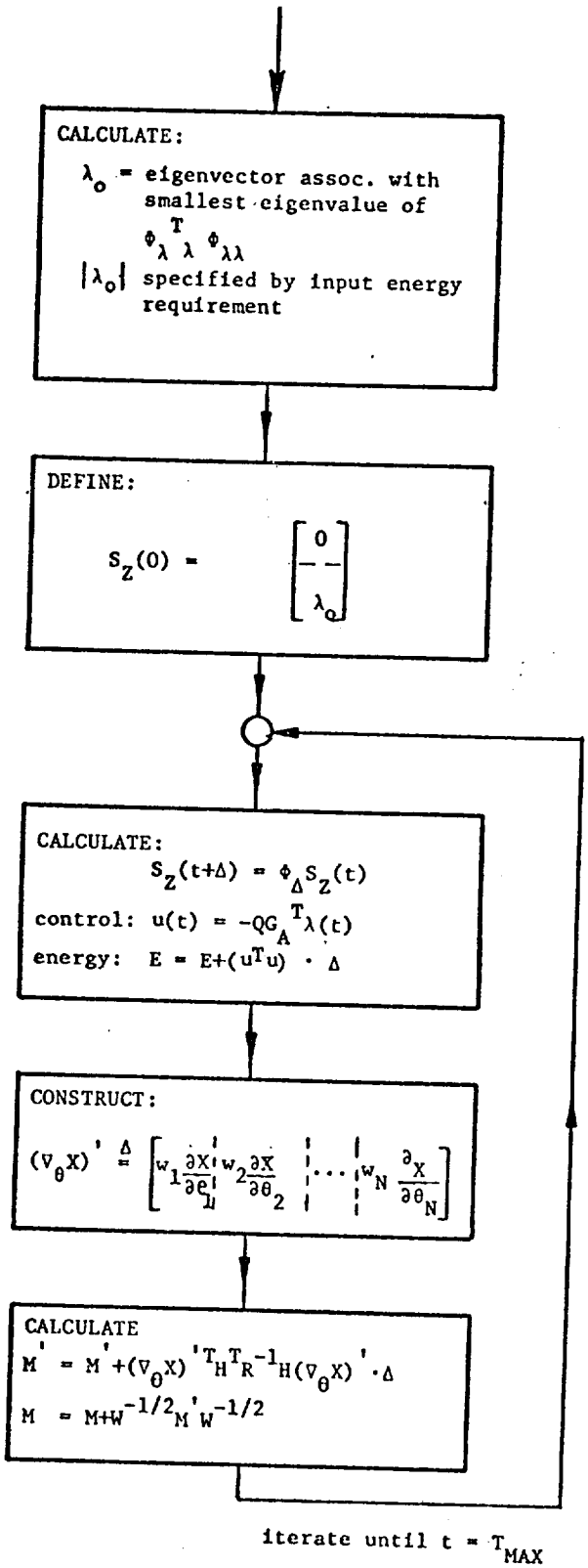


Fig. 7.1 FLOW-CHART OF OPTIMAL INPUT COMPUTER PROGRAM

vector of $\phi_{\lambda\lambda}(T,0;\mu)$ associated with the zero eigenvalue. The problem encountered was basically numerical. Very seldom, if at all, would the matrix $\phi_{\lambda\lambda}(T,0;\mu)$ exhibit an exactly zero eigenvalue. This was caused, to a large extent, by the fact that T is never the exact instant of singularity of the ricatti solution. This difficulty was resolved by choosing $\lambda(0)$ such that the product $\phi_{\lambda\lambda}(T,0;\mu)\ \lambda(0)$ had the smallest norm. If $\phi_{\lambda\lambda}(T,0;\mu)$ did, indeed have a zero eigenvalue, this would be the associated eigenvalue and $\phi_{\lambda\lambda}\ \lambda(0) \equiv 0$. The solution of this minimum norm problem is to choose $\lambda(0)$ to be the (normalized) eigenvector associated with the smallest eigenvalue of $\phi_{\lambda\lambda}{}^{T}\phi_{\lambda\lambda}$. This was the technique incorporated into the optimal input program.

Two additional items should be mentioned concerning the steps in the computational algorithm. The first involves specifying some additional fact about the eigenvector $\lambda(0)$ so that it can be uniquely specified. It was shown in Section 7.2 that the control at any time point, t, is a linear functional of $\lambda(t)$. Therefore, if a particular input energy is required, the norm of $\lambda(0)$ is scaled to the proper amount.

The second item concerns the reconstruction of the M matrix as the last step in the computer algorithm. To just calculate the performance index, tr{WM}, it would be sufficient to use

$$\mathrm{tr}\{\int_0^T x_A^T H_A^T R_A^{-1} H_A x_A dt\} \quad .$$

However, the information matrix M, the Cramer-Rao lower bound M^{-1}, and the det(M) give important information about the identification of the parameters which is not reflected in tr{WM}. For this reason, the augmented state vector, x_A, is rearranged to construct the matrix $(\nabla_\theta x)W^{1/2}$, which is used to compute the information matrix M. In addition, the eigenvalues of M and a figure-of-merit $\mathrm{tr}\{M\} \cdot \mathrm{tr}\{M^{-1}\}$ is also computed.

7.8 Specialized Algorithms

This section describes the two specialized algorithms which have been incorporated into the main program. These are: (i) the algorithm for computing the optimal set of weights when using the product of diagonal elements of M criterion, and (ii) the algorithm for computing μ_{max} when the data length, T, is specified.

An Algorithm for Choosing Weights:

As outlined in Appendix B, one purpose of the weighted trace is to obtain a closer approximation to the determinant of M as the performance criterion. This is done by bringing the ratio of the largest to the smallest diagonal element of M as close to 1 as possible. The computer algorithm takes the form of an iterative sequence of choosing weights, calculating the resulting M and then updating the weights and repeating.

The formula for the updating of the weights, once an optimal input has been computed along with the accompanying information matrix, is as follows:

$$w_{new_i} = w_{old_i} + \alpha\left[\frac{1}{m_{ii}} - w_{old_i}\right] \tag{7.88}$$

where m_{ii} is the i^{th} diagonal element of M. Additional logic was subsequently added to the program to reduce the factor α by successive factors of 2 if the new set of weights failed to reduce the ratio of the largest to smallest diagonal element of M. This was made necessary since the equation (7.88) for updating the weights is by no means optimal.

An Algorithm for Determining μ_{max} for a Given T:

The second special algorithm built into the program enables a user to determine the μ_{max} for a specific length of data, T. The most direct, but costly, way of determining a value of μ_{max} would be to pick several values of μ_{max} and run through the program, finding the associated values of T. These pairs (μ_{max},T) could then be used to construct a μ_{max}

from this curve, the μ_{max} for a desired T could be determined. However, this procedure would require a great many (μ_{max},T) pairs. A more direct method is to employ the optimal input program with associated zero crossing logic which would converge onto the correct μ_{max}. For the simple case of (μ_1,T_1) and (μ_2,T_2) as known associated pairs and $T_1 > T_d > T_2$ where T_d is the desired data length, the iterative equation for successive choices of μ_d is $(\mu_{d_2} \equiv \mu_2)$ (Ref. 56).

$$\frac{1}{\mu_{d_{i+1}}} = \frac{(T_i - T_d)}{T_i - T_1}\frac{1}{\mu_1} - \frac{(T_2 - T_d)}{T_i - T_1}\frac{1}{\mu_{d_i}} \quad . \tag{7.89}$$

With a new, updated choice of μ_{d_i}, the program is run and an associated T_i is found. This new pair is used in Eq. (7.89) to find an updated μ_{d_i}, and so on. Once the change in values for μ_d becomes smaller than some ε, the procedure is stopped. For ε of 1%, this procedure usually requires only four or five repetitions. Of course, if the value of T_d is outside the range of the given initial pair (μ_1,T_1) and (μ_2,T_2), the logic of Eq. (7.89) is altered appropriately.

Examples of both these specialized algorithms are given in Section VIII.

VIII

APPLICATION OF OPTIMAL INPUT DESIGN TO C-8 AIRCRAFT

This section discusses the results of applying the optimal input design technique described in Section VII to a two and four state model of the longitudinal dynamics of a C-8 aircraft. In both cases the input was designed to optimize the identification of the five parameters associated with the short period mode. Additional constraints were put on the input in the form of limiting the maximum excursion of the angle of attack and pitch rate by adjusting the total energy of the optimal input. In the first part of this section the results of the two-state example are described, including a frequency domain analysis of the optimal input, a comparison of the input performance versus a suboptimal doublet elevator input and use of the weighted trace criterion. The second part of this section describes the results of a Monte Carlo simulation of the identification process for the five short period parameters from simulated flight test data for both optimal input and a suboptimal doublet input.

8.1 Short Period (Two-State) Longitudinal Dynarics of C-8 Aircraft

8.1.1 Optimal Input for two-state model

This investigation involved finding the optimal elevator deflections δ_e to identify the parameters in the two dimensional short period longitudinal equations of motion for a C-8 aircraft. The state variables are the pitch rate q and angle-of-attack α. The equations for the short period dynamics of the C-8 aircraft were

$$\begin{bmatrix} \dot{q} \\ \dot{\alpha} \end{bmatrix} = \begin{bmatrix} -1.588 & -.562 \\ 1 & -0.737 \end{bmatrix} \begin{bmatrix} q \\ \alpha \end{bmatrix} + \begin{bmatrix} -1.66 \\ 0.005 \end{bmatrix} \delta_e$$

and the measurement equations were

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} q \\ \alpha \end{bmatrix} + \begin{bmatrix} n_q \\ n_\alpha \end{bmatrix}$$

In determining the power spectral densities of n_q and n_α, the values given in Ref. 13 (1 deg./sec. error in q, and 2^o in α) were multiplied by two times the correlation time of the noise sources, which is assumed to be 0.01 secs. The measurement noise spectral density matrix is therefore given as

$$R = \begin{bmatrix} 0.02 & 0 \\ 0 & 0.04 \end{bmatrix}$$

For this investigation the data length T was fixed at 4 secs. The appropriate μ_{max} was found from a μ_{max}-T curve shown in Fig. 8.1. The value of μ_{max} associated with a T of 4 secs is 0.015. The shape of the μ_{max}-T curve in Fig. 8.1 is characteristic of the general relationship between these two variables.

For the μ_{max} and R values indicated above the optimal input with the respect to the three parameters in the F matrix and the two parameters in G is given in Fig. 8.2. The energy of the input was 311 and tr{M} = 20,460. The check value of μ_{max} E was approximately 20,200, indicating a numerical error of 0.1%. The determinant of M was computed to be 1×10^{15}, with the ratio of the largest to smallest eigenvalue of M being almost three orders of magnitude. The eigenvalues themselves indicated that much of the relative uncertainty in the parameter estimates was concentrated in two of the five dimensions. The standard deviations of the parameter estimates were

$$\text{Standard Deviation for F} = \begin{bmatrix} 0.167 & 0.0639 \\ & 0.035 \end{bmatrix}, \quad \text{Standard Deviation for G} = \begin{bmatrix} 0.095 \\ 0.025 \end{bmatrix}$$

The time histories of the states α and q, resulting from the optimal

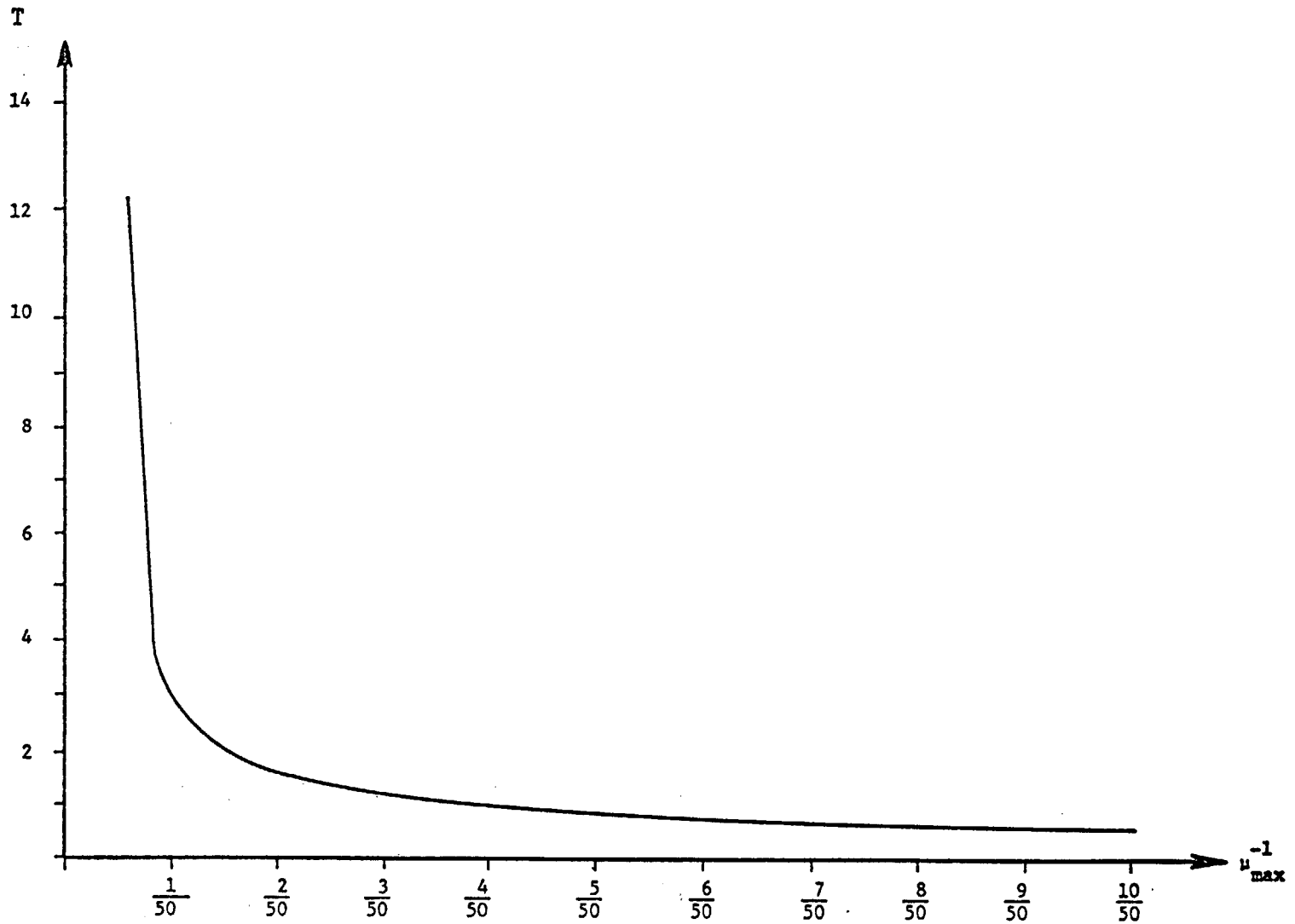


Fig. 8.1 - μ_{max}^{-1} VS. T CURVE FOR A 2 STATE/5 PARAMETER MODEL

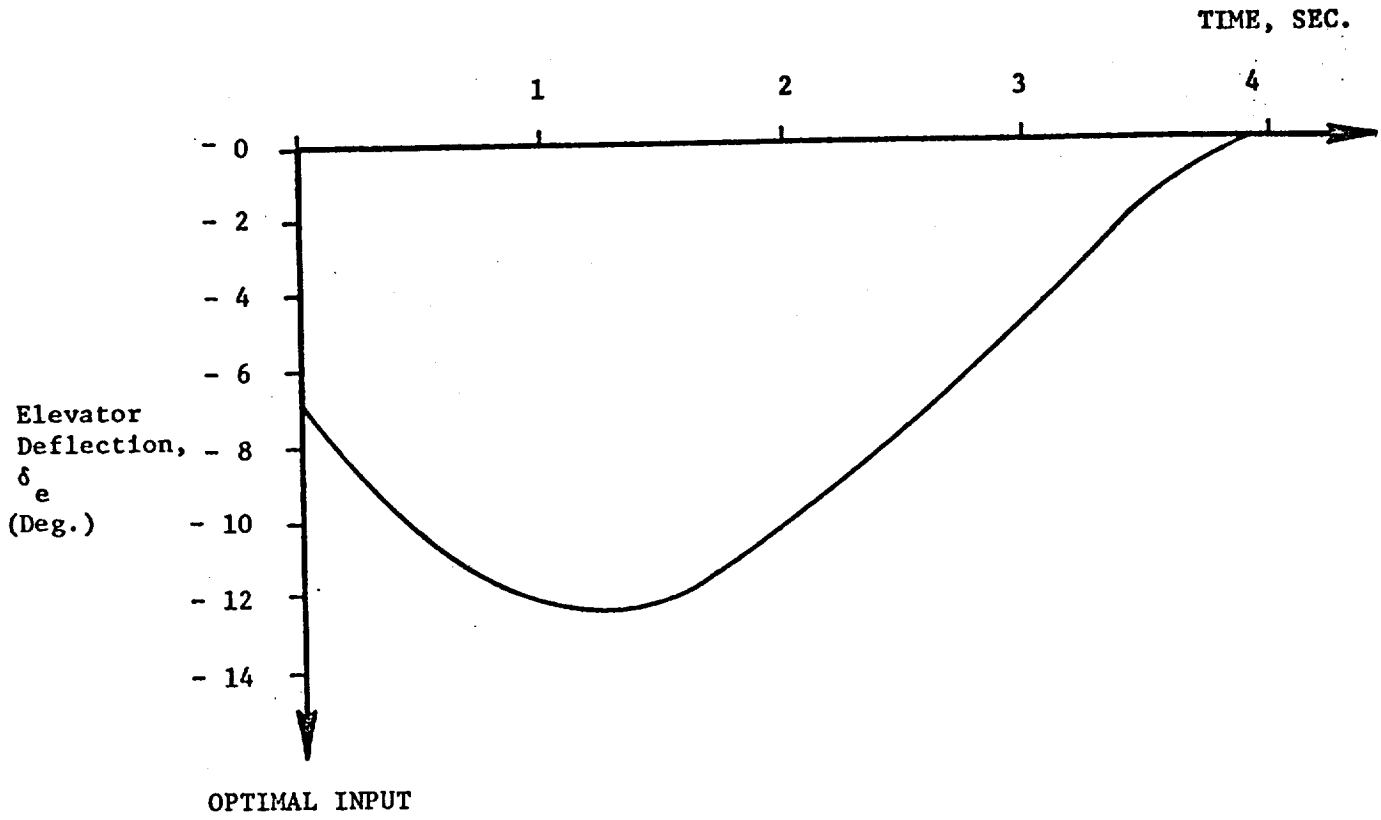


Fig. 8.2 - OPTIMAL INPUT FOR SHORT PERIOD LONGITUDINAL DYNAMICS

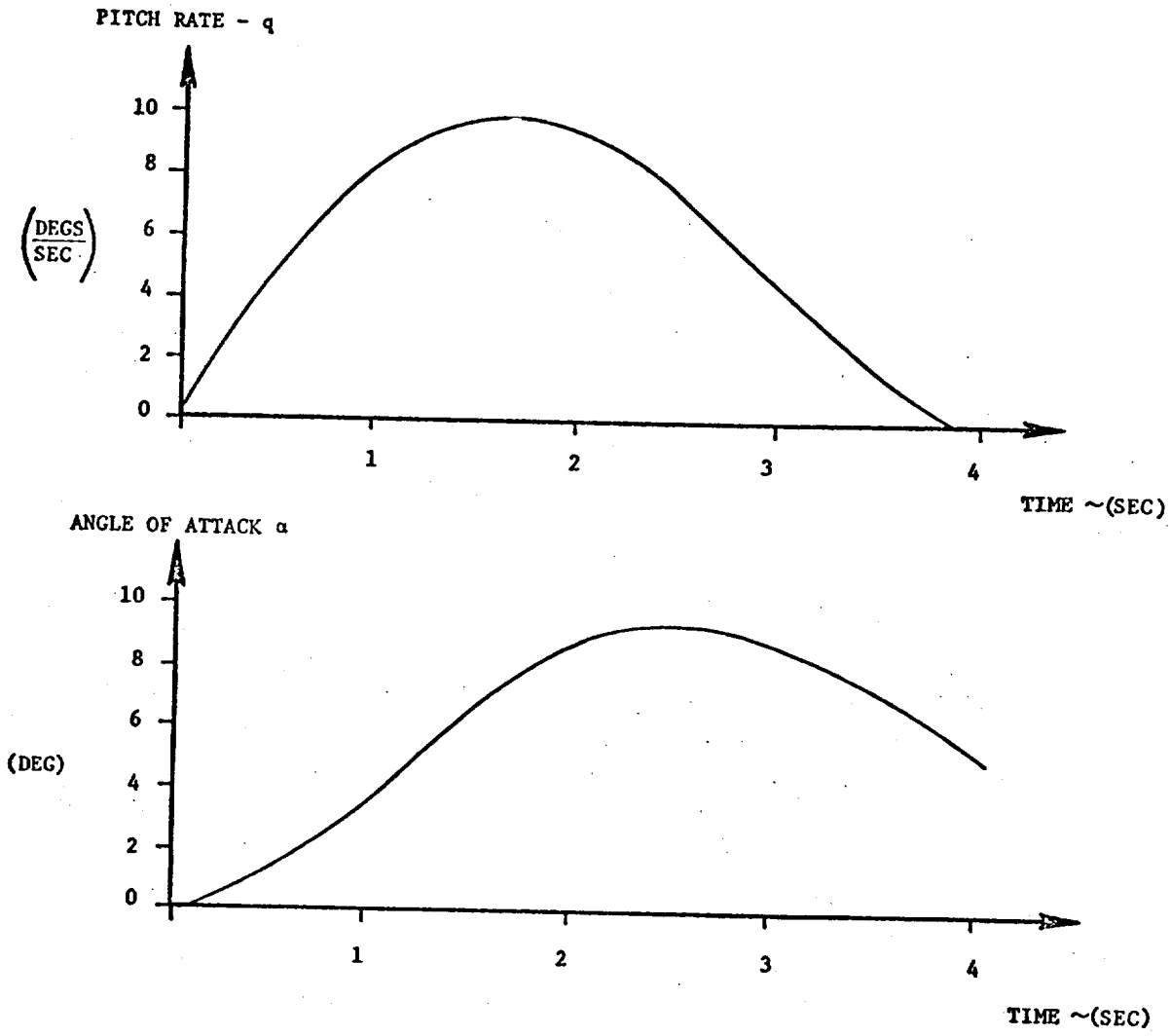


Fig. 8.3 - PITCH RATE AND ANGLE-OF-ATTACK TIME HISTORIES WITH OPTIMAL INPUT

input are shown in Fig. 8.3. The energy of the input was constrained so that α does not exceed 10°. This method of energy limitation is the most direct way of applying state constraints; although, as mentioned in Section VII, the penalty function approach can also be used.

8.1.2 Fourier Transform of the Optimal Input

Since it was specified that the input be designed with respect to the parameters both in F and G, it would be reasonable to infer that the input would have a low frequency component for identifying the parameters in G and a higher frequency component for identifying the parameters in F. The actual Fourier transform of the input signal is given in Fig. 8.4. The vertical scale has been reduced to 1/10 its actual height in order to illustrate the smaller variations. The DC component is 0.98. The small peak in the transform occurs at a frequency of 0.375 cyc/sec, which is close to the short period frequency of the aircraft.

The most important point demonstrated by the frequency domain analysis is that the sinusoidal like component of the input signal occurred at a very specific frequency. It is well known that the maximum signal power can be obtained from a second order system if it is excited at its natural frequency. Therefore, in order to maximize the sensitivity of the output signal to the parameters in the F matrix (which is given by the M matrix), or in other words, to maximize the component of the output signal due to the parameters in F, the input signal had a specific component set at the natural system frequency.

8.1.3 Comparison with a Doublet Input of Equal Duration and Energy

The third part of the investigation was to compare the performance of the optimal input to that of a doublet input of equal duration and energy. The doublet input used for this comparison is shown in Fig. 8.5.

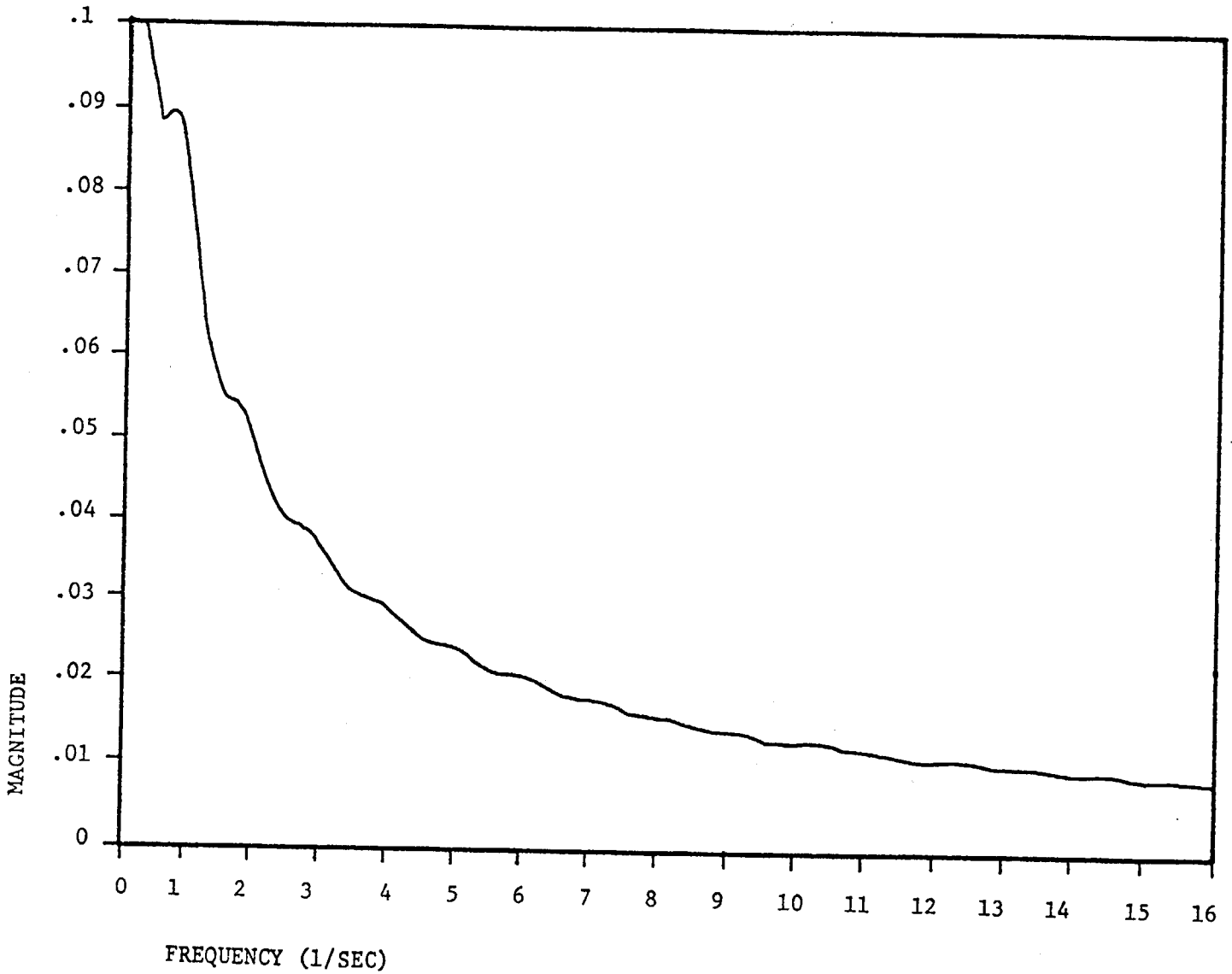


Fig. 8.4 - FOURIER TRANSFORM OF OPTIMAL INPUT

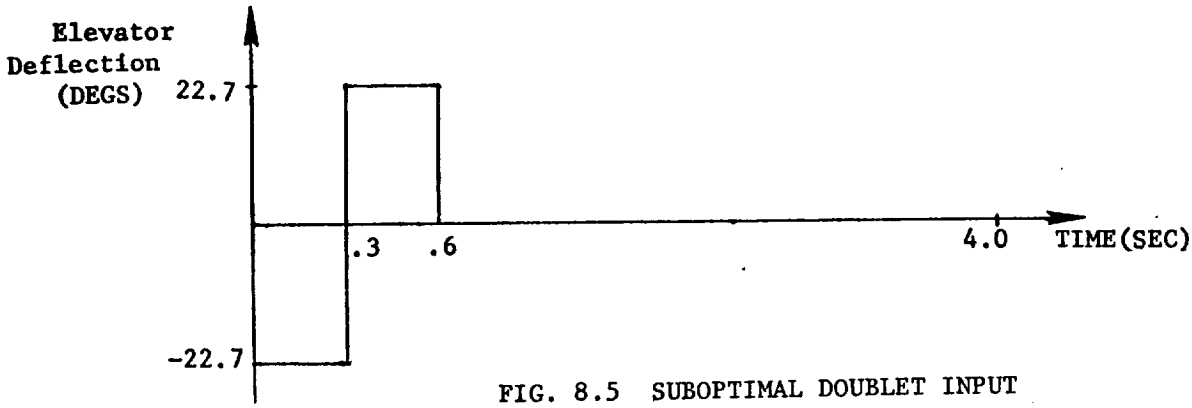


FIG. 8.5 SUBOPTIMAL DOUBLET INPUT

The determinant of M was 7×10^8, seven orders of magnitude less than for the optimal, while the trace of M was only 1600. The standard deviations of the parameter estimates were

$$\text{Standard Deviation for F} = \begin{bmatrix} .147 & .247 \\ & .164 \end{bmatrix}; \quad \text{Standard Deviation for G} = \begin{bmatrix} .0631 \\ .0346 \end{bmatrix}$$

Two of these values are smaller than those obtained for the optimal input, however the standard deviations in the parameter estimates are not explicitly in the optimization criterion.

8.1.4 Effect of Small Parameter Value Changes on Optimal Input

Since, in an operational application, the actual parameter values are not known, it was important to investigate the effect that changes in the parameter values had on the optimal input shape. These changes might reflect, for example, the difference between wind tunnel or theoretical estimates of an aircraft's stability and control derivative and the actual stability and control derivative values. It would be these estimates, however, that would be used to compute the optimal input for the identification of those same derivatives. The modified F matrix which was used in this investigation was

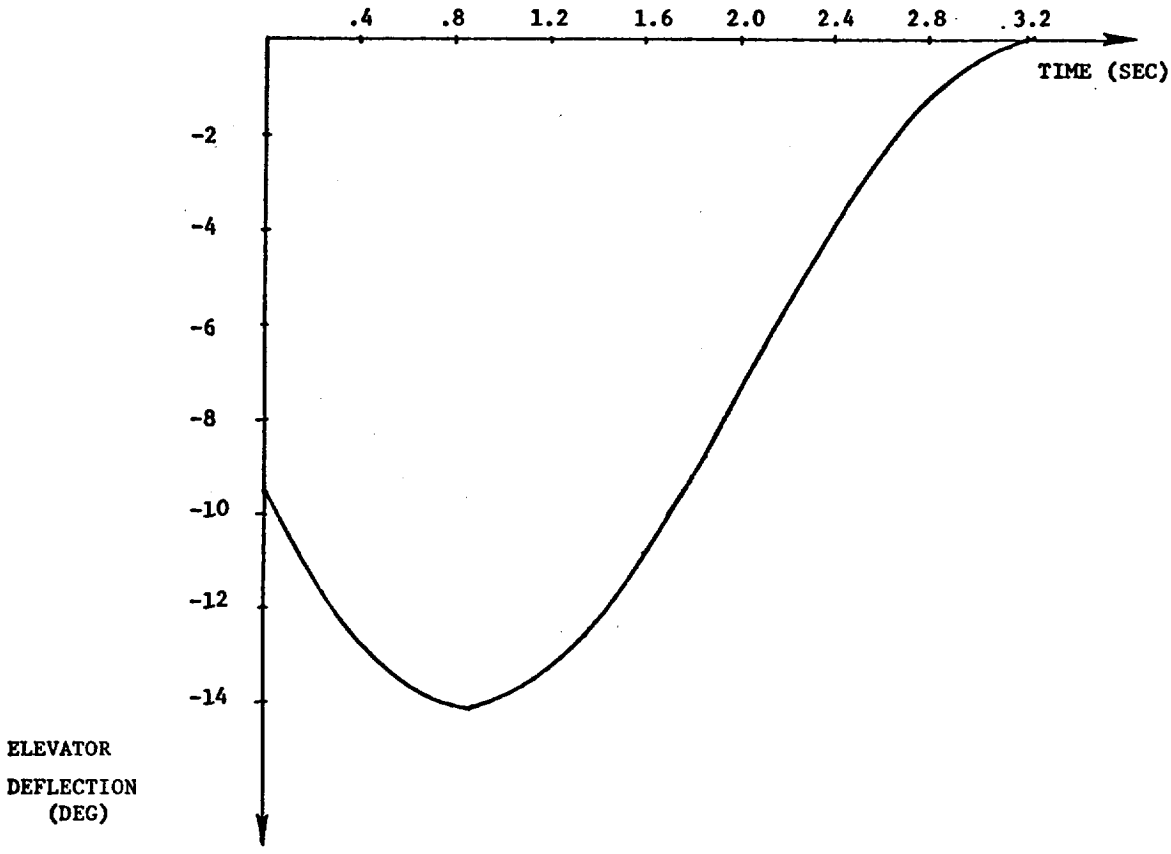


Fig. 8.6 - OPTIMAL INPUT FOR SYSTEM WITH 10% PARAMETER VARIATION

$$F = \begin{bmatrix} -1.429 & -.613 \\ 1 & -.663 \end{bmatrix}$$

The optimal input resulting from this approximately 10% change in the parameter values is shown in Fig. 8.6. The most obvious difference is that the input length is now 3.21 seconds instead of 4.0 secs. Otherwise, the input has the same qualitative shape as the optimal input. The expected standard deviations in the parameter estimates resulting from this input when it is applied to the original system are as follows:

$$\text{Standard Deviation for } F = \begin{bmatrix} .255 & .0886 \\ & .036 \end{bmatrix} \quad ; \quad \text{Standard Deviation for } G = \begin{bmatrix} .148 \\ .026 \end{bmatrix}$$

Three of these values are very close to the standard deviations obtained for the optimal input, while the other two represent increases of approximately 50%. The improvement over the suboptimal input, however, still exists.

8.1.5 Weighted Trace Criterion:

This part of the investigation involved using the weighted trace criterion to derive the optimal input and choosing the weights to make the diagonal elements of the information matrix equal. As detailed in Section VII the performance criterion is $tr\{WM\}$ where W is a diagonal matrix of weights chosen to set $\omega_1 m_{11} = \ldots = \omega_p m_{pp}$. When all the diagonal elements of WM are equal, maximizing the trace of WM is equivalent to maximizing the product of the diagonal elements which is a better approximation to det(M).

Since the input which maximizes the performance criterion depends on the values of the weights, which in turn affect the input, an iterative scheme was used to update the weights until convergence was achieved.

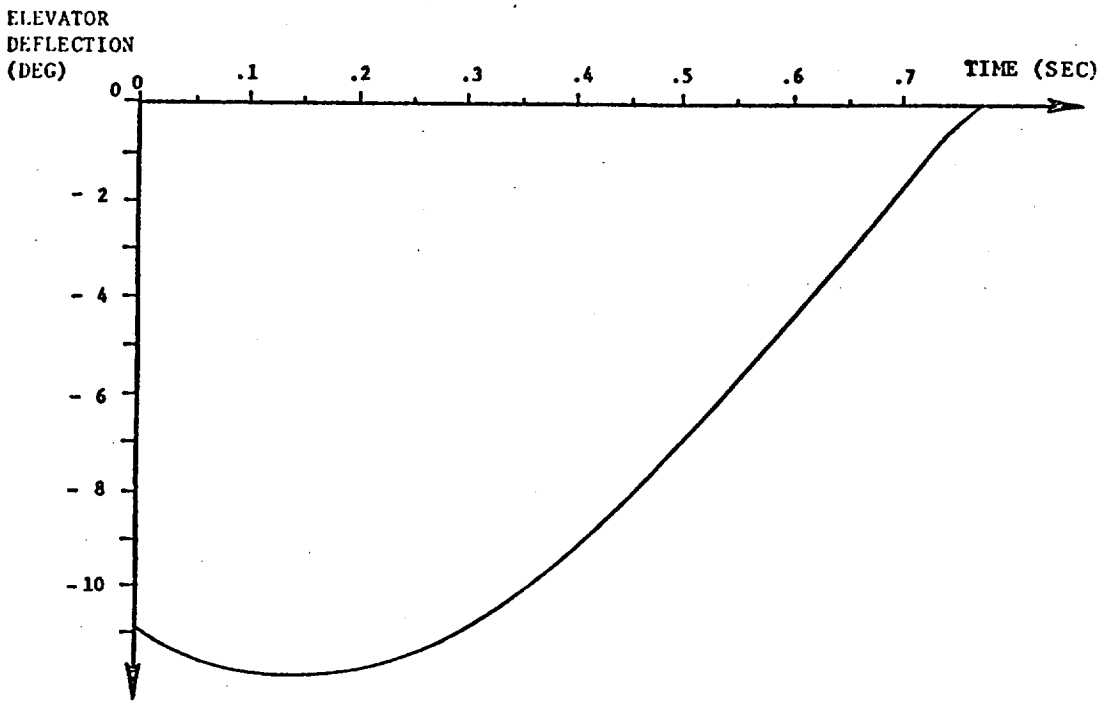


Fig. 8.7 - OPTIMAL ELEVATOR DEFLECTION WITH UNITY WEIGHTS

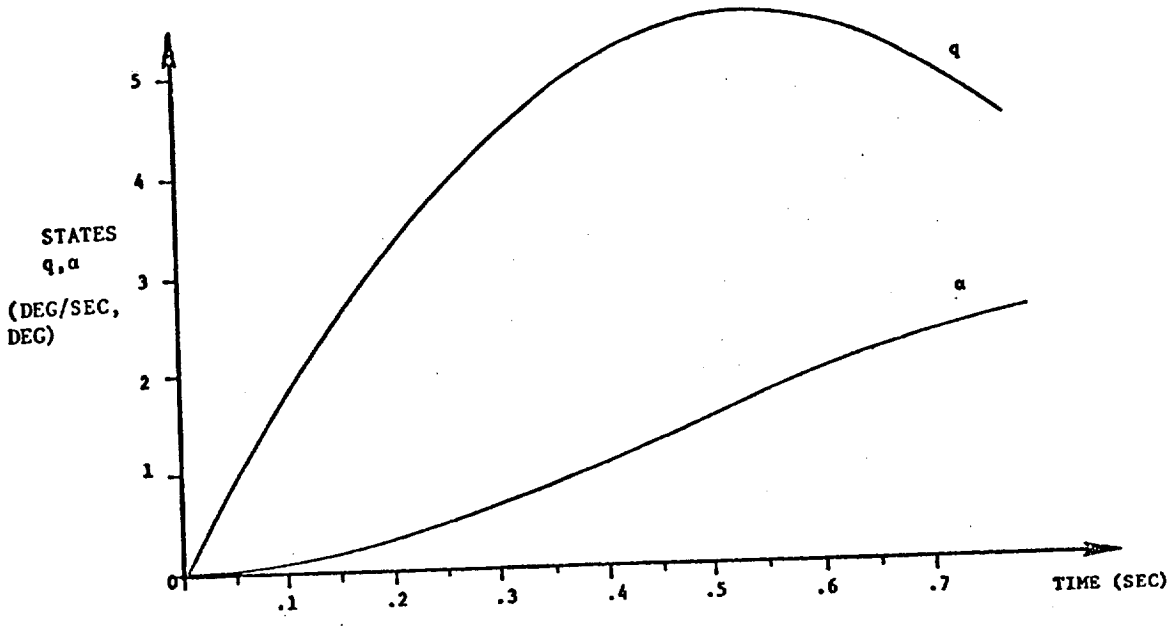


Fig. 8.8 OPTIMAL STATE TIME HISTORIES FOR UNITY WEIGHTS

The optimal input and the state time histories for a T of 0.77 sec and an energy of 62.61, and unity weights are given in Figs. 8.7 and 8.8. The trace of M for this input was 611 and the determinant of M was 8.11×10^3. The ratio of the largest to smallest eigenvalue of M was 1270. The eigenvalues themselves indicated that the parameter uncertainty was quite disproportionate along two of the eigenvector directions. The standard deviations in the parameter estimates are given below.

$$\text{Standard Deviation for } F = \begin{bmatrix} 2.03 & 5.57 \\ & 1.43 \end{bmatrix}, \quad \text{Standard Deviation for } G = \begin{bmatrix} 0.381 \\ 0.129 \end{bmatrix}$$

Notice the increase in standard deviations due to a shorter data length (0.77 sec. vs 4 sec.) and smaller input energy (62.6 vs 311).

Using the weighted trace criterion, and 12 iterations to bring the ratio of the largest to the smallest element of WM down to 1.14, the optimal input and state time histories given in Fig. 8.9 were obtained. The determinant of M was calculated to be 2.27×10^4 which is five times greater than the unity weights determined. The volume of the uncertainty ellipsoid decreased by the same factor. Another indication of this was the fact that the ratio of the largest to smallest eigenvalue of M was reduced by a factor of 2 to 640, with the largest eigenvalue itself being reduced by a factor of 2. The standard deviations for the parameter estimates are as follows:

$$\text{Standard Deviation for } F = \begin{bmatrix} 1.56 & 4.04 \\ & 1.23 \end{bmatrix}, \quad \text{Standard Deviation for } G = \begin{bmatrix} 0.304 \\ 0.122 \end{bmatrix}$$

Notice on Fig. 8.8 that the parameters which are poorly estimated with unity weights were assigned higher weights. As a result, the lengths of the error ellipsoid along each of the axes has become more uniform and the total volume of the uncertainty ellipsoid has been decreased.

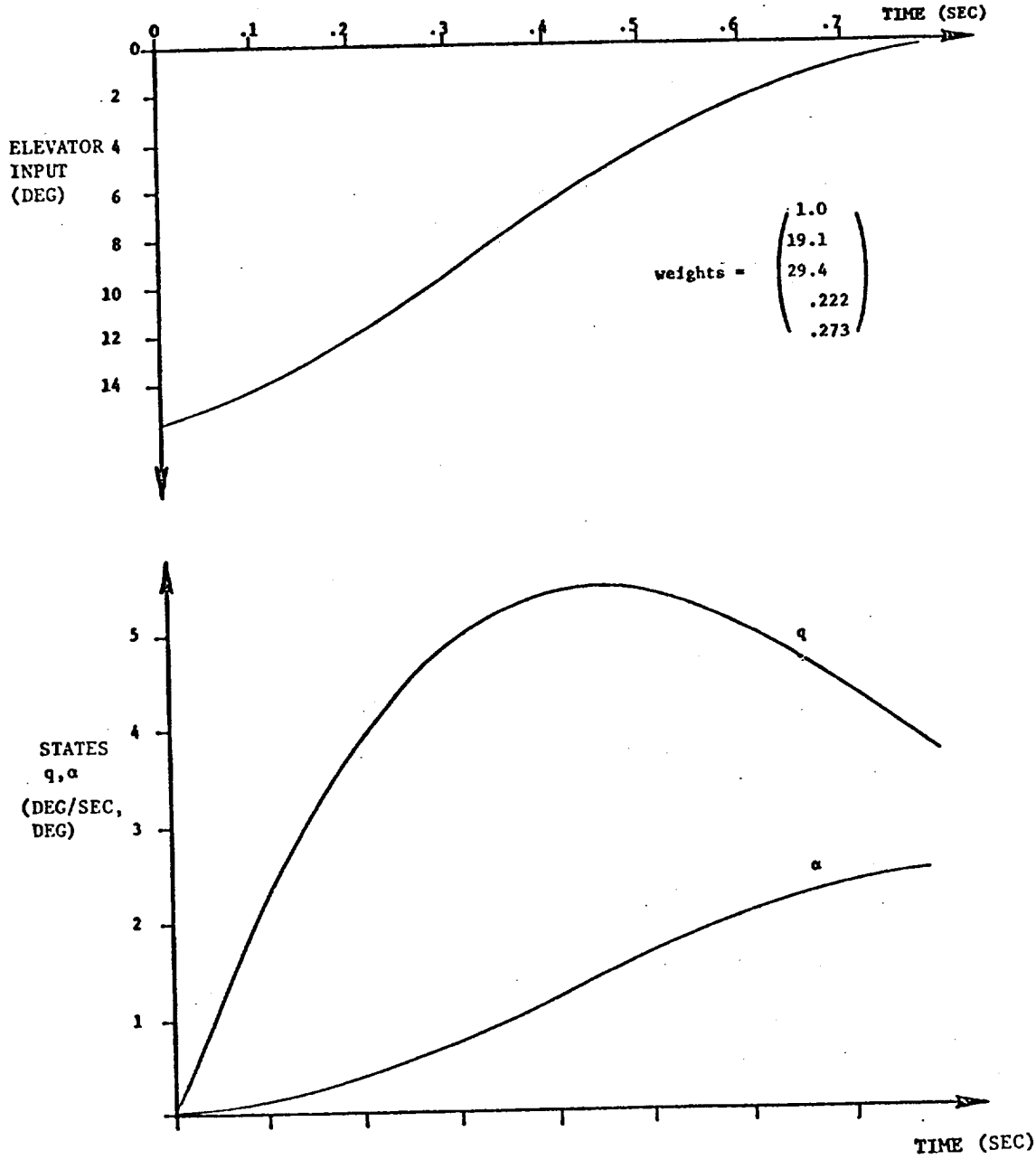


Fig. 8.9 OPTIMAL INPUT AND STATE TIME HISTORIES - WITH WEIGHTED TRACE

It is clear from this example that using a weighted trace criterion does result in an input which can, in an overall sense, identify the unknown parameters with improved accuracy. This improvement is measured by the increase in the value of the determinant of M.

8.2 C-8 Monte Carlo Simulation

A more realistic test for verifying the advantages of the optimal input over a sub-optimal doublet input of equal energy was to perform a monte carlo simulation of the identification process using simulated flight test data. To make the simulation representative of an actual flight test the following were included. First, both phugoid and the short period modes in the longitudinal equations of motion were used to generate the data, although only the short period derivatives were to be identified. The optimal input had the characteristic of suppressing, as much as possible, the phugoid mode in order to maximize the sensitivity of the output to the values of the short period parameters. Second, short period parameters of the four state model that were used to compute the optimal input were changed by approximately 50% from the equivalent parameter in the model that generated the data. In this way the situation of designing the input based on the wind tunnel values of the stability and control derivative was simulated.

There were several criteria for comparing the performance of the optimal input with that of the suboptimal doublet input. Since the optimal input was computed on the basis of maximizing the trace of the expected information matrix, this is an obvious candidate. Maximizing the determinant of the information matrix, or equivalently minimizing the volume of the uncertainty ellipsoid, in parameter space, is another. However, by the nature of a monte carlo simulation, the parameter estimates themselves can be used to calculate the sample covariance, and its trace, determinant or the eigenvalues can be used as performance criteria. Finally, histograms of the parameter estimates themselves and the associated probability distributions can also be used in determining input performance.

8.2.1 Optimal and Suboptimal Inputs for Monte Carlo Simulation

The four state longitudinal equations of motion for the C-8 aircraft which were used for computing the optimal input are

$$\begin{bmatrix} \dot{u} \\ \dot{\theta} \\ \dot{q} \\ \dot{\alpha} \end{bmatrix} = \begin{bmatrix} -.02 & -32.2 & 0 & 33.74 \\ 0 & 0 & 1 & 0 \\ .003 & 0 & -1.588 & -.562 \\ -.004 & 0 & 1 & -.737 \end{bmatrix} \begin{bmatrix} u \\ \theta \\ q \\ \alpha \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ -1.658 \\ .005 \end{bmatrix} \delta_e$$

$$\begin{bmatrix} z_{1_k} \\ z_{2_k} \\ z_{3_k} \\ z_{4_k} \end{bmatrix} = \begin{bmatrix} u_k \\ \theta_k \\ q_k \\ \alpha_k \end{bmatrix} + \begin{bmatrix} n_u \\ \eta_\theta \\ n_q \\ n_\alpha \end{bmatrix}$$

where $R = E\{\underline{n}_i \underline{n}_j\} = \begin{bmatrix} .125 & & & \\ & .125 & & \\ & & .125 & \\ & & & .250 \end{bmatrix} \delta_{ij}$

is the discrete measurement noise covariance. The optimal input itself was designed with respect to enhancing the ability to identify the five short period parameters, given in the locations marked by an x

in F : $\begin{bmatrix} - & - & - & - \\ & & x & x \\ & & & x \end{bmatrix}$; in G: $\begin{bmatrix} - \\ x \\ x \end{bmatrix}$

These parameters are M_q, M_α, Z_α, M_{δ_e} and Z_{δ_e}. Considering the amount of computer time needed to generate the optimal input for five parameters and four state variables, a 2 second data length was decided upon. The optimal input, for an energy of 58.0 is shown in Fig. 8.10, along with the suboptimal doublet input, of equal energy, which was used for comparison.

8.2.2 Generation of Simulated Flight Data

The different stability and control derivatives which were used in the generation of the simulated flight data are illustrated below.

$$F = \begin{bmatrix} \text{Same} & \text{Same} \\ & -2.238 \quad -.28 \\ \text{Same} & 1. \quad -.368 \end{bmatrix} \qquad G = \begin{bmatrix} \text{Same} \\ -.829 \\ -.0075 \end{bmatrix}$$

The random noise added to each of the four measurements were derived from a Gaussian random number generator. These noise sequences presented a slight problem since, with a 2 second data length and a .02 sampling period, the 100 samples were sometimes insufficient for the noise statistics to have the desired mean and covariance. However, since the performance of the optimal and suboptimal inputs were always compared for the same measurement noise sequence, the problem of incorrect noise statistics should be of minor importance for comparison purposes.

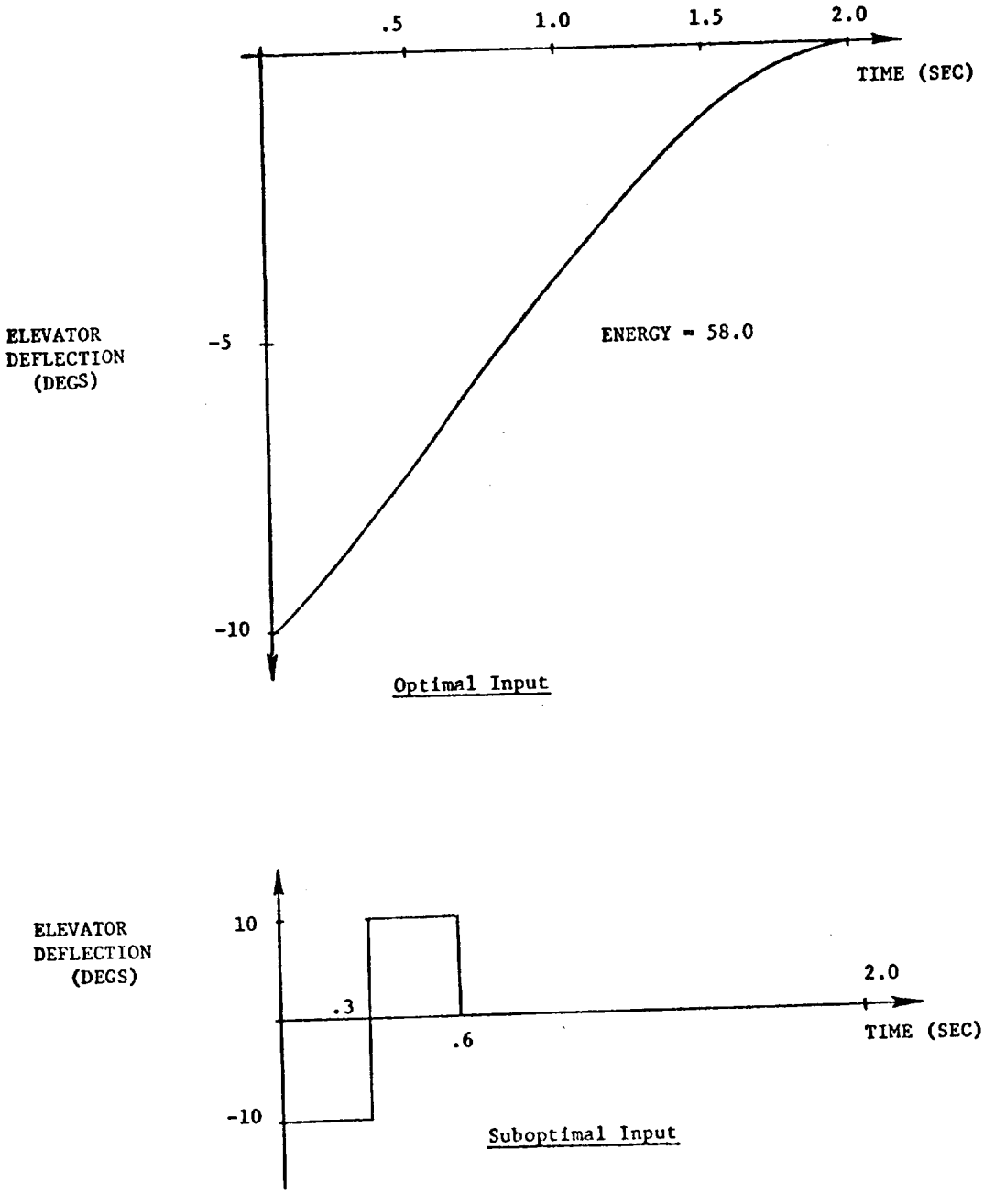


Fig. 8.10 - OPTIMAL AND SUBOPTIMAL INPUT FOR MONTE CARLO SIMULATION

8.2.3 Description of Monte Carlo Identification Simulation

A flow diagram of the Monte Carlo identification simulation is given in Fig. 8.11 . Fifty different sets of random numbers are added to the computed time histories in order to simulate a statistically significant set of runs at the same flight condition. Each of the 50 sets of simulated flight test data was processed in exactly the same manner using the SCI maximum likelihood identification program. The parameters to be identified consisted of the five unknown stability and control derivatives and the measurement noise covariance matrix. The initial stability and control derivative estimates were set equal to the values used in computing the optimal input, while the initial state estimates were set equal to the value of the first data point, since all four variables were being measured. The estimate of the measurement noise covariance matrix is obtained, at each iteration, from the sample error covariance (fit error covariance). The four diagonal elements of the sample error covariance are taken to be the estimates of the four independent measurement noise sequences.

For each set of data the identification procedure is carried out until either the change in cost, between two iterations, or the norm of parameter step size, or the norm of the gradient becomes smaller than some present tolerances. For both the optimal and suboptimal inputs, convergence, indicated by one of these three conditions being satisfied, was usually obtained after 4 iterations through the data. At the end of processing a set of data, the final parameter estimates and the information matrix would be stored on tape for use in the latter compilation of the results.

Since there wasn't any process noise being considered in the simulation, the maximum likelihood identification technique reduces to an output error method, with the weighting matrix being the inverse of the sample error covariance. The required calculations consist of, for a given set of parameter estimates, integrating the system equations and the sensitivity equations. Since there are 4 states and 5 parameters, this makes a total of 24 simultaneous differential equations. The information matrix is computed from the time history of the sensitivity functions.

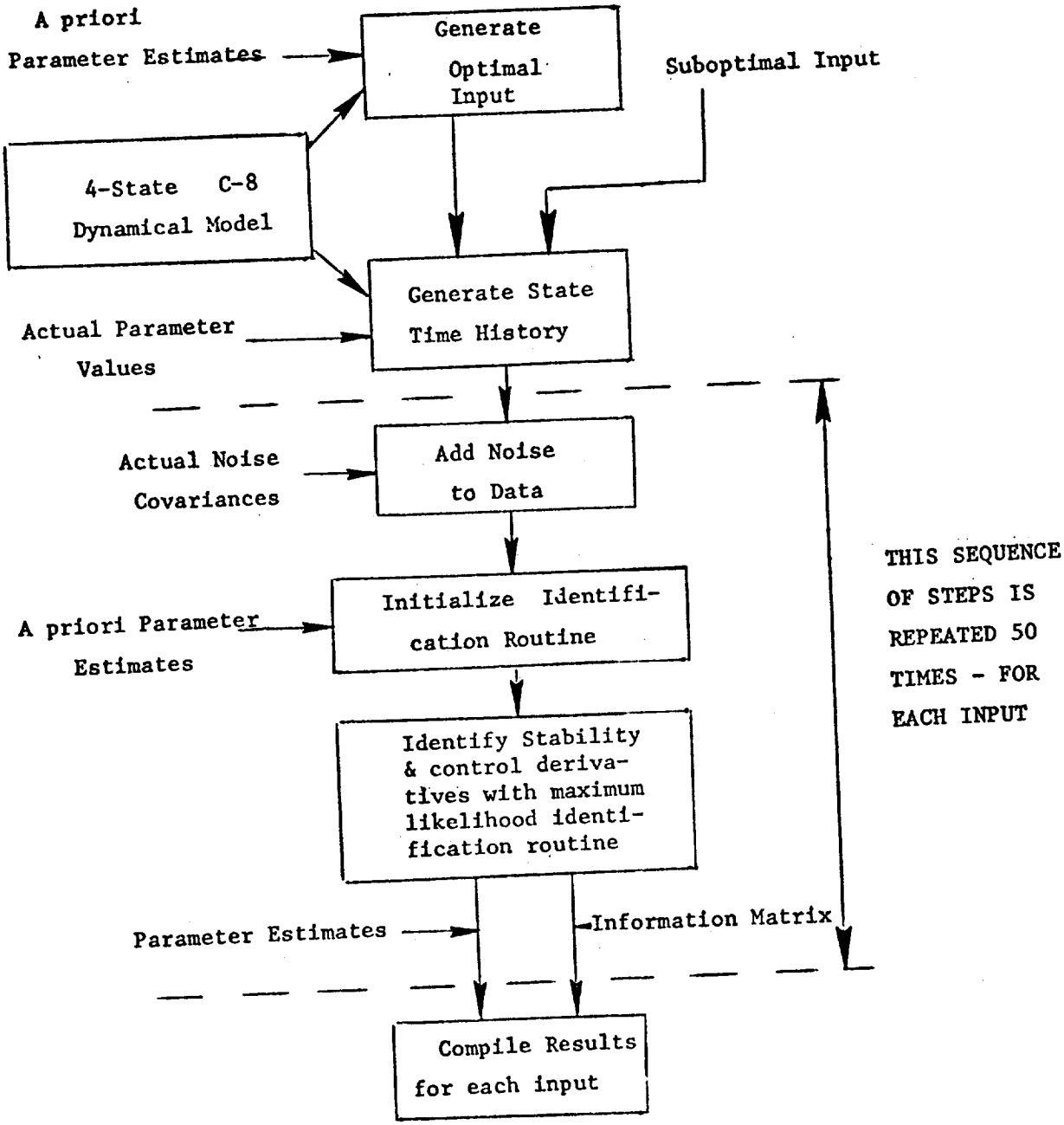


Fig. 8.11 BLOCK DIAGRAM OF MONTE CARLO SIMULATION

It is interesting to note that since the weighting matrix is the sample error covariance, the value of the likelihood function will be determined completely by the log of the determinant of the sample error covariance. There are also two different methods for choosing the weighting matrix. One is that it should be considered a diagonal matrix since the measurement noise sources are all independent. The other is that, non zero off diagonal terms be allowed in the weighting matrix. For this Monte Carlo simulation, the former method gave better results. In the latter case, the error covariance matrix had a tendency to become singular at times.

For this application, none of the options discussed in Section 4.6 , including parameter fixing, rank-deficient solution for M^{-1} or bounding the parameter estimates were used. Neither was it necessary during any of the iterations to cut the step size in order to improve the convergence properties.

8.2.4 Results of Monte Carlo Simulation

The monte carlo simulation consisted of 50 runs of the identification procedure, with the parameter estimates, information matrix and its eigenvalues, and the parameter covariance matrix and its eigenvalues being saved at the end of each run. The ensemble results are tabulated in Table 8.1.

The theoretical values of the trace of the information matrix, using the actual values for the stability and control derivatives, were computed to be 2.12×10^7 and 4.74×10^5 for the optimal and suboptimal inputs respectively. The average values, from Table 8.1, for the 50 runs were 2.15×10^7 and 4.79×10^5, indicating that 50 runs were sufficient for obtaining accurate parameter estimate and information matrix statistics. In addition, the trace of the sample covariance, computed from

$$\text{Sample Covariance} = \frac{1}{50} \sum_{j=i}^{50} (\Delta p_j - \overline{\Delta p})\ (\Delta p_j - \overline{\Delta p})^T$$

TABLE 8.1 MONTE CARLO RESULTS BASED ON IDENTIFICATION FOR 50 SETS OF SIMULATED DATA

	Optimal Input	Suboptimal Input
Trace of Sample Covariance	.242	.315
Determinant of Sample Covariance	1.62×10^{-19}	1.501×10^{-15}
Eigenvalues of Sample Covariance	.234	.262
	$.725 \times 10^{-2}$	$.514 \times 10^{-1}$
	$.252 \times 10^{-3}$	$.115 \times 10^{-2}$
	$.188 \times 10^{-4}$	$.766 \times 10^{-4}$
	$.202 \times 10^{-7}$	$.126 \times 10^{-5}$
Parameter Standard Deviations	.407	.307
	.295	.491
	.00925	.0235
	.0771	.0537
	.000570	.00168
Average Trace of Information Matrix	2.15×10^{7}	4.79×10^{5}
Eigenvalues of Average Information Matrix	2.14×10^{7}	4.79×10^{5}
	2.95×10^{4}	8.46×10^{3}
	6.56×10^{3}	4.77×10^{2}
	1.39×10^{2}	2.18×10^{1}
	1.12	4.14
Average Determinant of Information Matrix	4.70×10^{18}	1.95×10^{14}
Average Trace of the Covariance Matrix (Cramer-Rao Lower Bound)	.182	.312
Lower bound on parameter standard deviations (from Cramer-Rao Lower Bound)	.351	.303
	.234	.441
	.00876	.0311
	.0665	.0568
	.000247	.00262

where Δp_j is the error in the parameter estimates for the j^{th} run and $\overline{\Delta p}$ is its mean. As stated, it was found that the sample covariance matrix was fairly close to the Cramer-Rao lower bound.

By almost all measures of performance the optimal input performed better than the suboptimal input. The determinant of the sample covariance, giving a measure of the overall parameter uncertainty based on the actual derived parameter estimates, was four orders of magnitude smaller for the optimal input. The eigenvalues of the sample covariance were smaller, on a one-to-one basis, for the optimal input, indicating a smaller dimension for each axis of the uncertainty ellipsoid.

The histograms of parameter estimates are shown in Figs. 8.12 - 8.16 For the M_α, Z_α and Z_{δ_e} parameters, the optimal input definitely produced a better ensemble of parameter estimates. The mean estimate value was much closer to the actual parameter value and the standard deviations and mean square errors were smaller. The performance for the two inputs was about the same for the M_q, while the suboptimal input did outperform the optimal input on the fourth parameter. (M_{δ_e}) However, it should be kept in mind that the accuracy of the parameter estimates themselves was not a direct performance objective. Rather, the overall input performance, as measured by the determinant or trace of the covariance matrix was the criterion of interest.

It is also interesting to note that the standard deviations for the parameter estimates obtained from the inverse of the information matrix (Cramer-Rao Lower Bound) did in fact bound the parameters standard deviations obtained from compiling the individual parameter estimates. For the optimal input, a comparison of these standard deviations is given below:

Lower Bound From Information Matrix Inverse (Cramer-Rao Lower Bound)	Actual Standard Deviation of Parameter Estimates
.351	.407
.234	.295
.00876	.00925
.0665	.0771
.000247	.000570

The histograms of the error in estimating two of the four components of the observation noise covariance are shown in Fig. 8.17 and 8.18. The errors in estimating the covariance are plotted rather than the covariance estimates themselves because the actual (sample) covariance of the noise varied from run to run due to the finite data length. As these histograms indicate, the error in estimating the 2nd component was consistently less than 5% for both the optimal and sub-optimal inputs and most often within 1% for the optimal input. For the third component, the error with the optimal input is consistently less than 7% and with the sub-optimal input, 9%. Overall, it is accurate to state that the performance of the optimal input in identifying the measurement noise covariances was only slightly improved over that of the sub-optimal input.

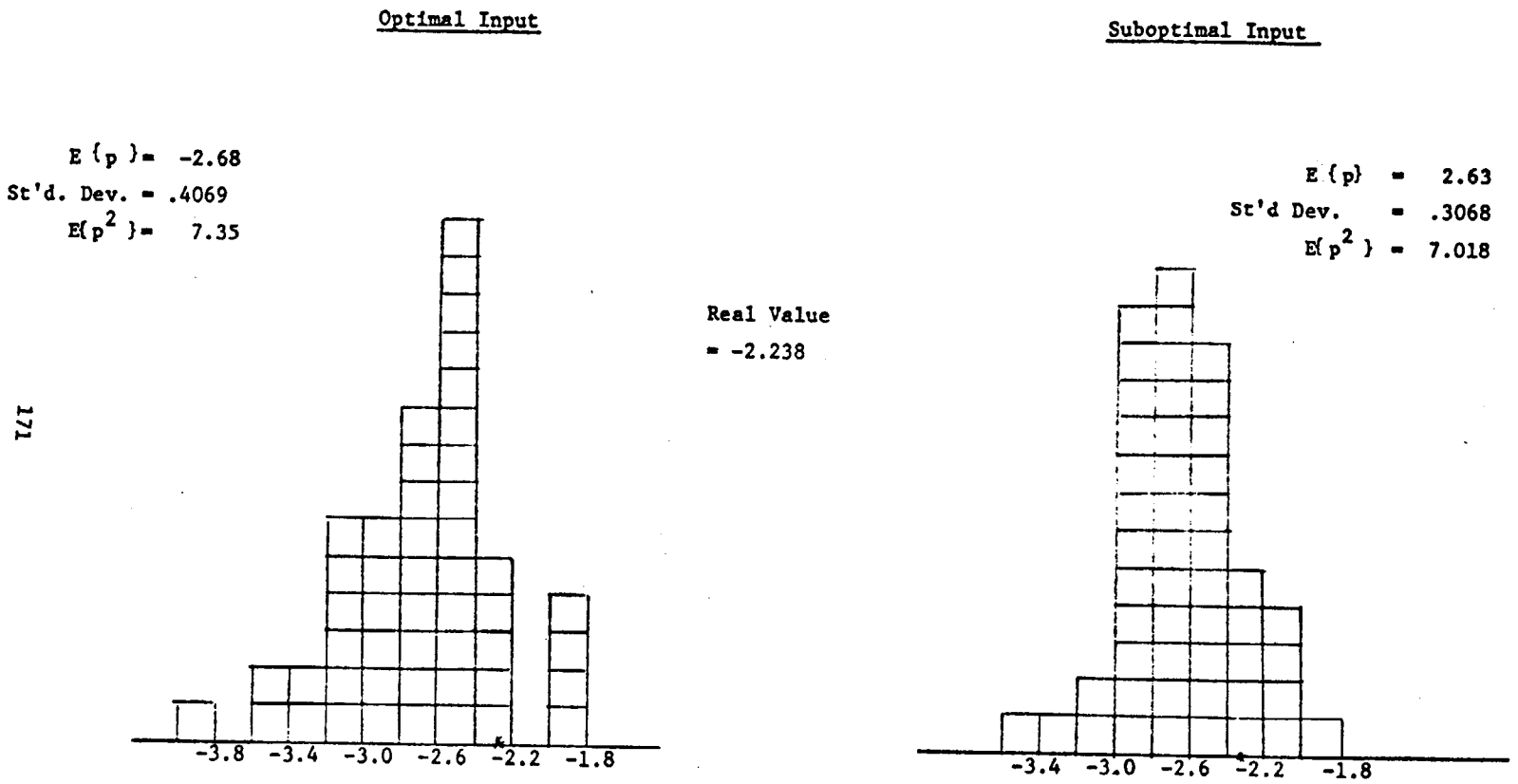


Fig. 8.12 PARAMETER ESTIMATE HISTOGRAMS FOR M_q

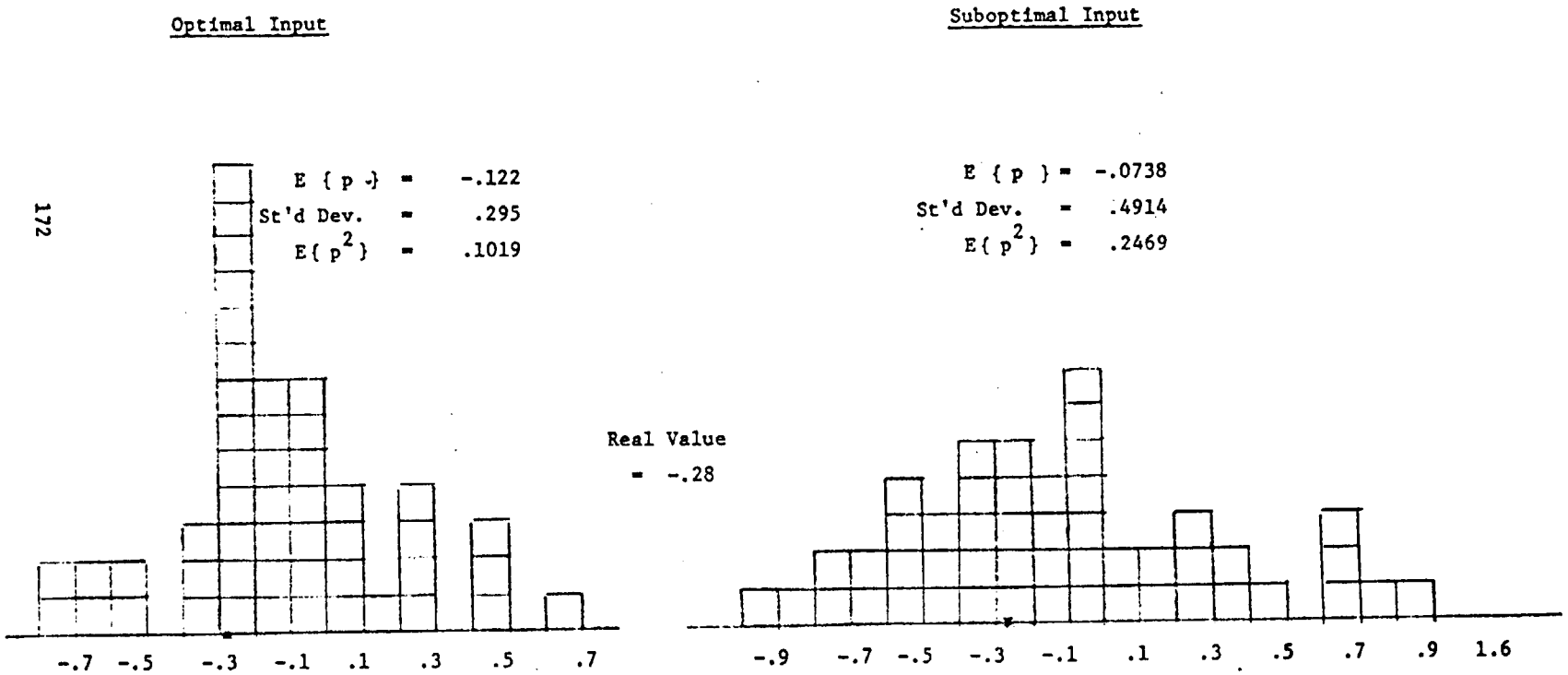


Fig. 8.13 PARAMETER ESTIMATE HISTOGRAMS FOR M_α

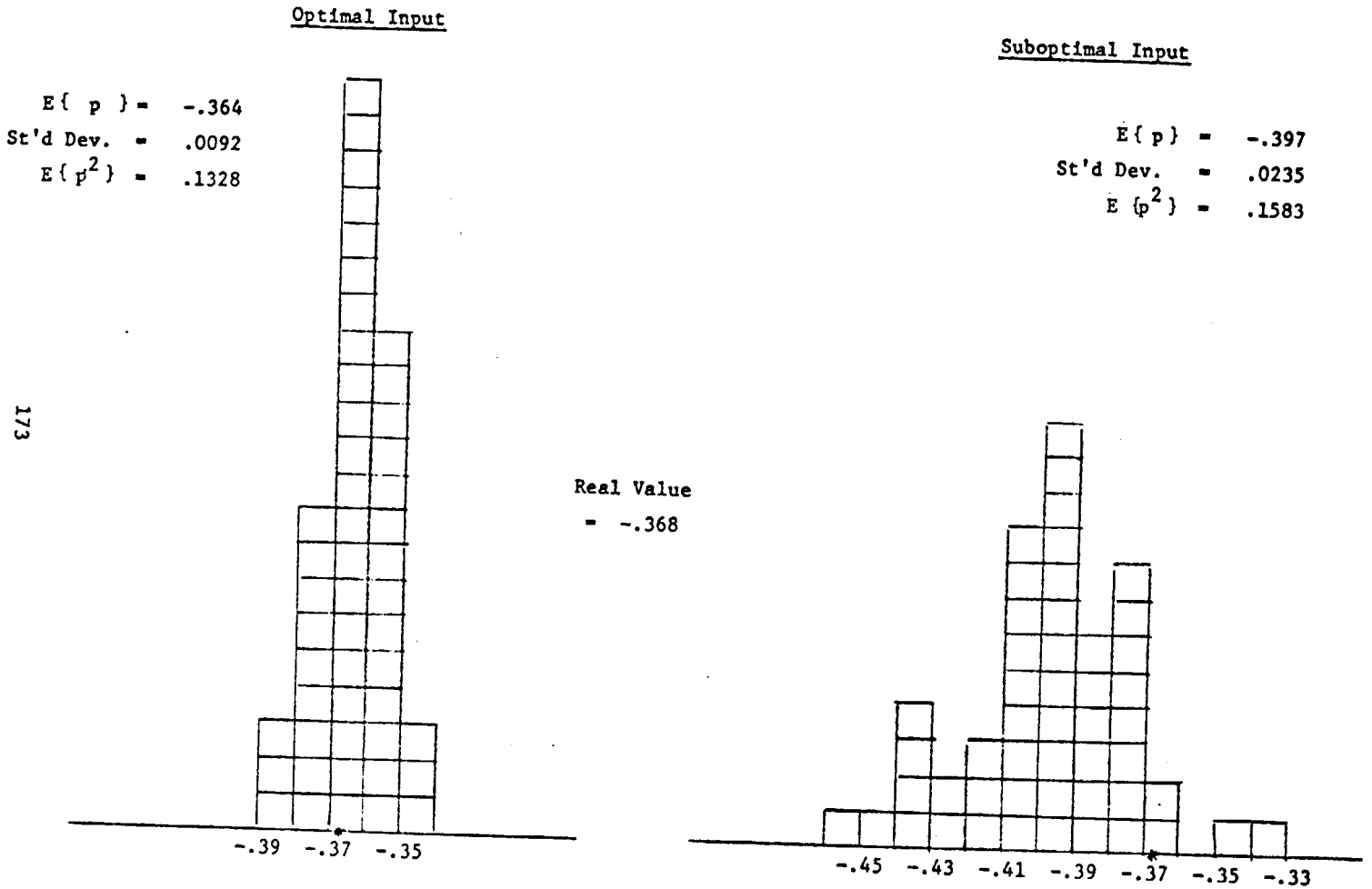


Fig. 8.14 PARAMETER ESTIMATE HISTOGRAMS FOR Z_α

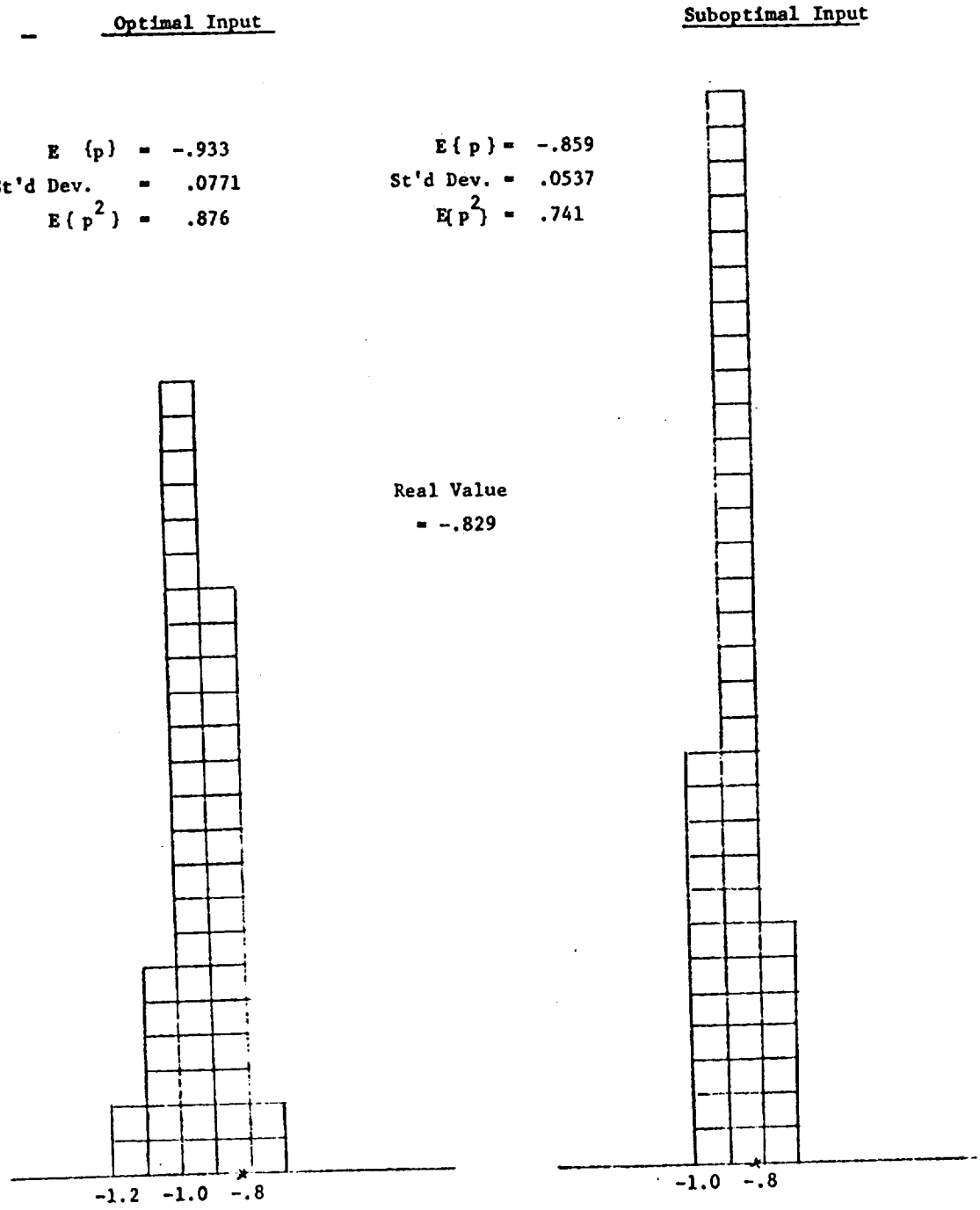


Fig. 8.15 PARAMETER ESTIMATE HISTOGRAMS FOR M_{δ_e}

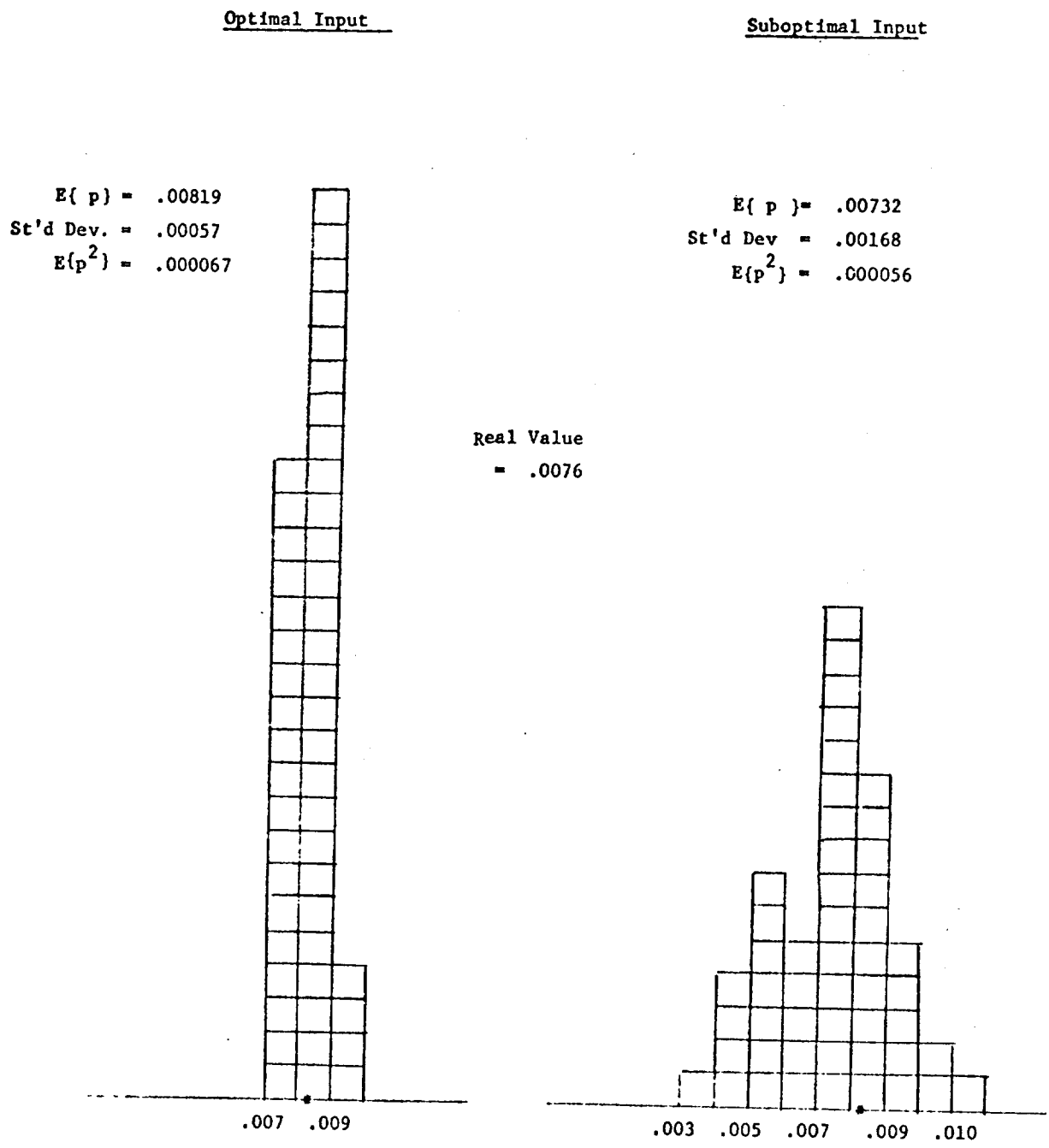


Fig. 8.16 PARAMETER ESTIMATE HISTOGRAMS FOR z_{δ_e}

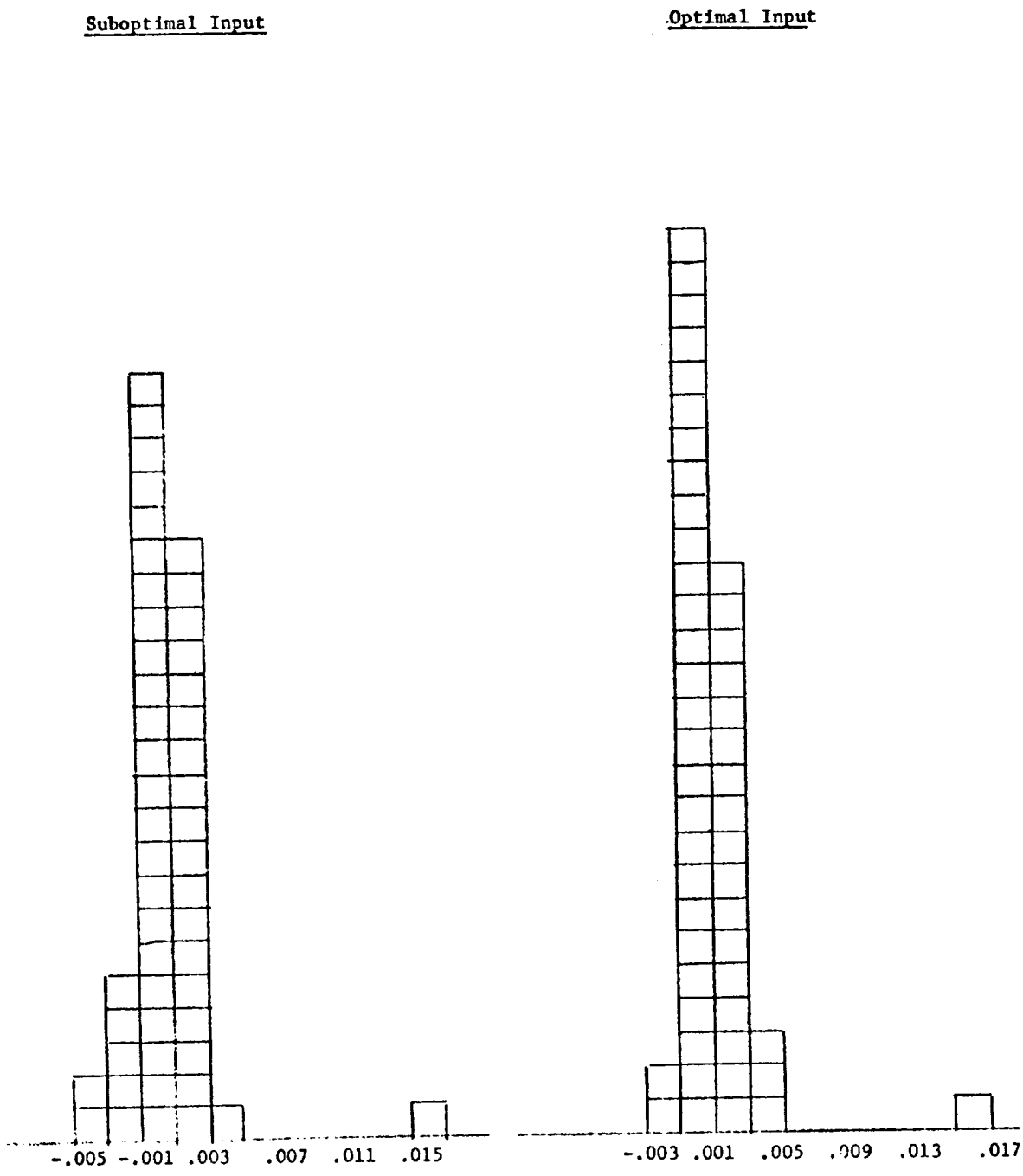


Fig. 8.17 HISTOGRAM OF ESTIMATION ERRORS FOR n_θ

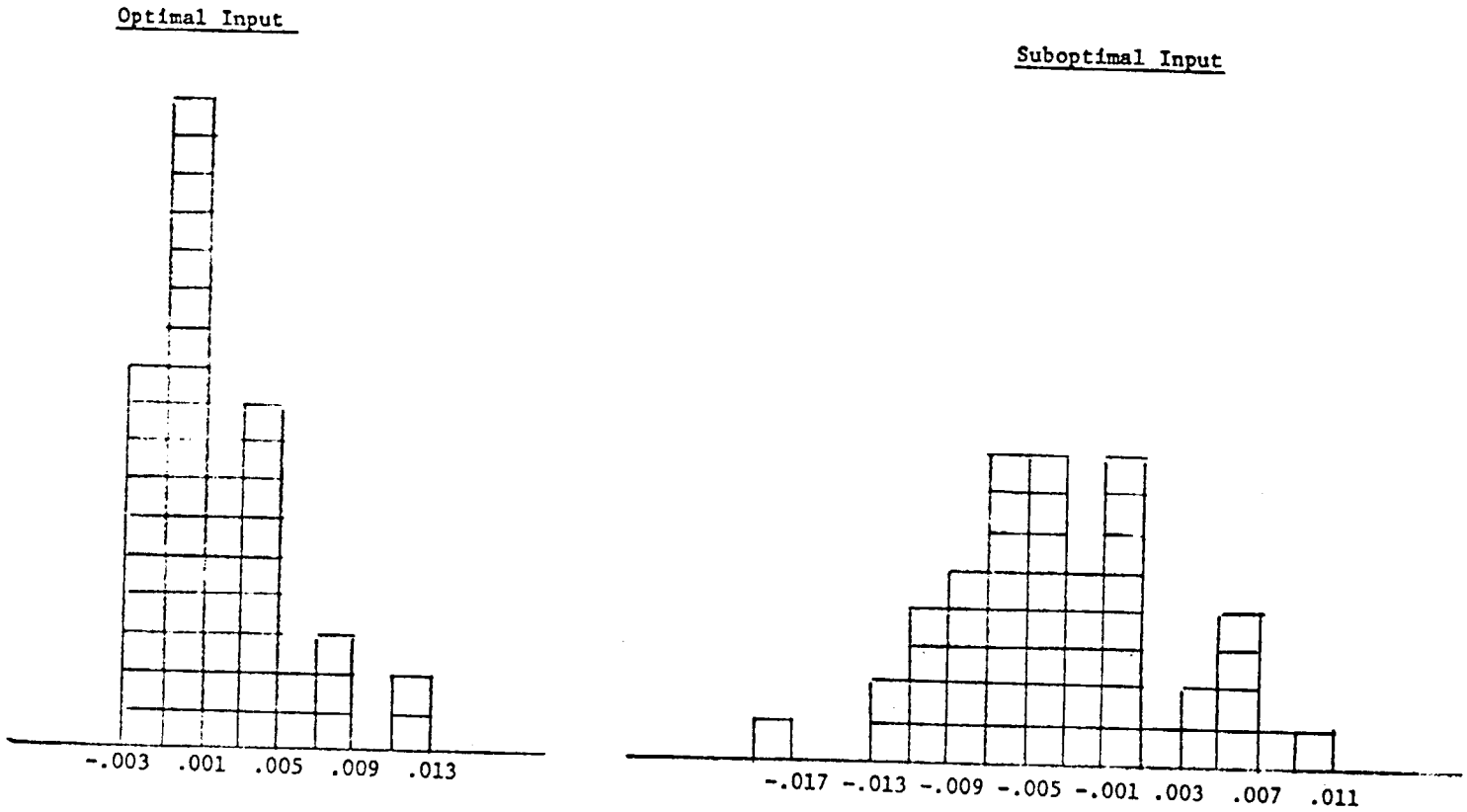


Fig. 8.18 HISTOGRAM OF ESTIMATION ERRORS FOR n_q

8.3 Optimal Input Through First-Order Filter

One possible objection to the optimal inputs which are derived from the optimization technique is that they start from nonzero value which cannot be realized in practice. For example, in the case of the C-8 Monte Carlo simulation, the initial elevator deflection for the optimal input was 10°. Although the suboptimal doublet input had an initial discontinuity of the same magnitude, a more realistic model of the situation would include the lag action of the control servos. This would prevent an instantaneous change in elevator deflection for both the optimal and suboptimal inputs. In addition, the optimal input, for identifying the short period parameters, has the disadvantage that the longitudinal velocity does not return to the nominal level after the input has been applied. For a nominal steady state value, the input must be two sided, i.e. both positive and negative values.

This section compares the performance of a two sided optimal input, with a total energy of 101.25 and length of 4 seconds to the performance of a suboptimal doublet input, of equal energy.
The 2 second optimal input which was used to generate the 4 - second input was derived for the C-8 four state longitudinal equations of motion given in section 8.2.1. A first order dynamical system with a time constant of .2 seconds and unity gain was used to simulate the control surface servo mechanism.

The 4 second optimal input after passed through the model of the servo dynamics, appears as in Fig. 8.19 . Note that the actual control surface deflection has no discontinuities, starts at 0° and has no violent maneuvers. The suboptimal input, on the other hand, which is shown in Fig. 8.20 , while also beginning at 0°, entails some much more drastic deflections.

Although a Monte Carlo indentification simulation was not performed for these two inputs, their performance, based on the charateristics of the information matrix, can still be compared. The determinant of the

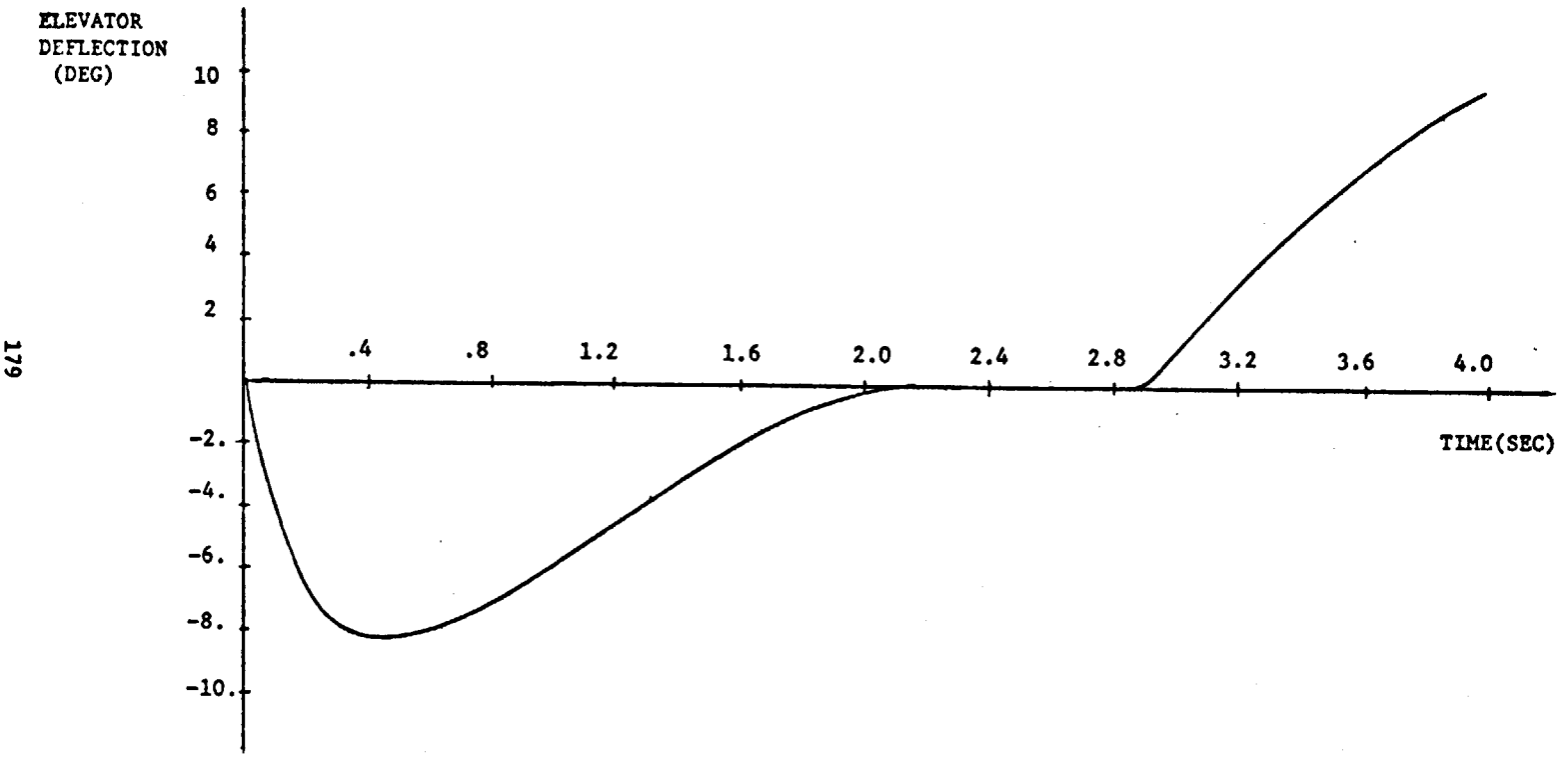


Fig. 8.19 TWO-SIDED OPTIMAL INPUT FROM OUTPUT OF FIRST ORDER SERVO

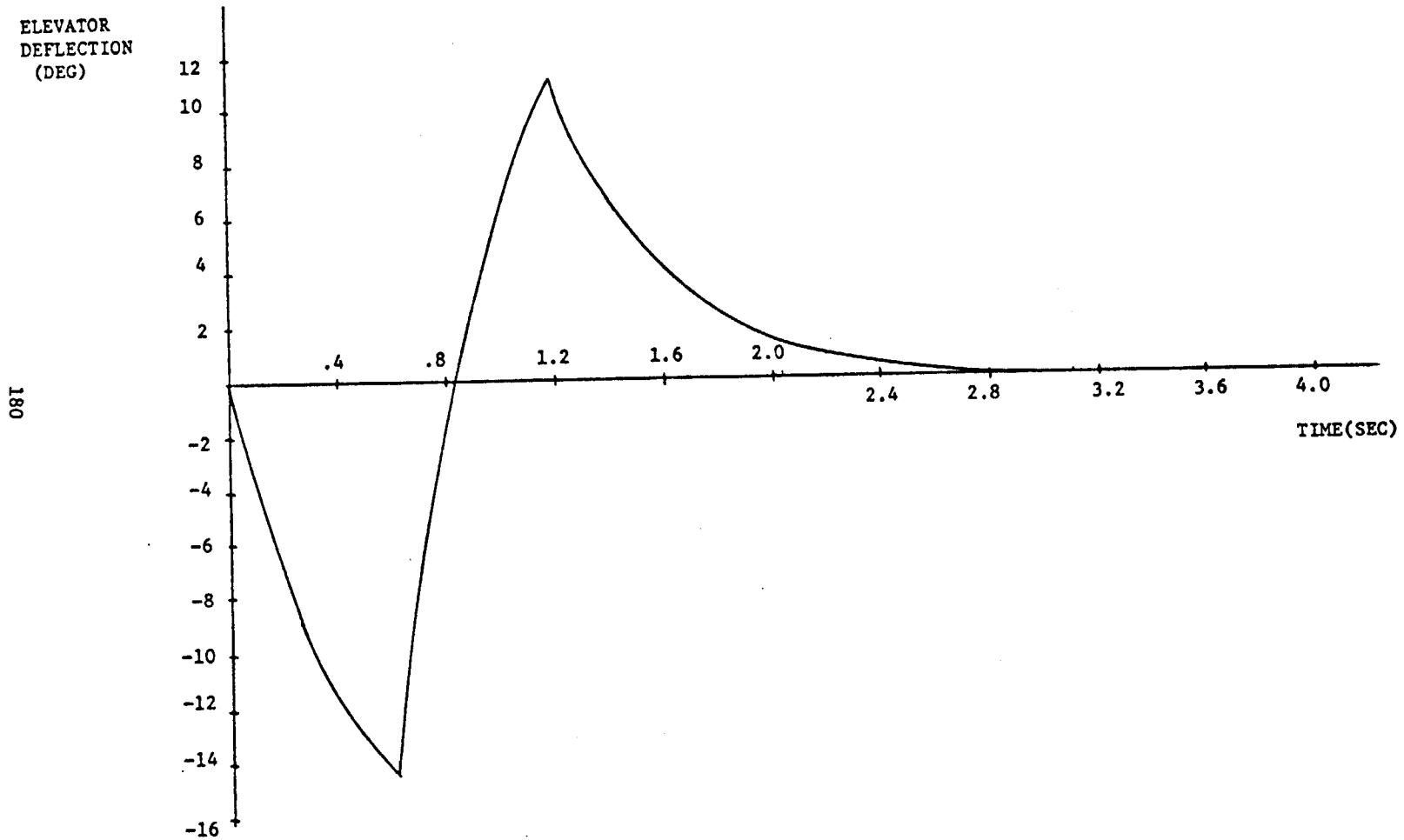


Fig. 8.20 TWO-SIDED SUB-OPTIMAL INPUT FROM OUTPUT OF FIRST ORDER SERVO

information matrix, representing the reciprocal of the volume of the uncertainty ellipsoid, was 1.6×10^{26} for the optimal input and 7.3×10^{24} for the suboptimal doublet input. Three out of five eigenvalues of the information for the optimal input were larger, on a one-to-one basis, than for the suboptimal input, and the trace of M was 1.4×10^{8} versus 9.7×10^{6}, with the one for the optimal input again being larger.

In summary, the performance of the 4 second optimal input versus the suboptimal input including the effects of the making the input two-sided and the control surface servo delay, changed only slightly from that observed during the extensive Monte Carlo simulation. Further work along these lines is needed to make optimal inputs realizable in practice.

IX

CONCLUSIONS

In this report, the problem of identifying aircraft stability and control derivatives from flight test data was discussed. It was shown that the three important elements of the identification process are (i) the identification method, (ii) the design of control inputs, and (iii) the instrumentation. The first part of the report described various methods that have been used in the past for identification. The special characteristics and limitations of these methods were discussed. A new technique, based on the Maximum Likelihood Criterion, was then derived and discussed in detail. The technique is applicable to nonlinear models containing both process and measurement noise. As such, the Maximum Likelihood Method presented here is the most general technique that has been developed for identifying stability and control derivatives from flight test data. It consists of a combination of a Kalman filter (linear) or an Extended Kalman filter (nonlinear) for estimating the state and a Modified Newton-Raphson iterative procedure for estimating the parameters.

The Maximum Likelihood Identification Method was applied successfully to three different problems :

(i) Identification of parameters from simulated data for a nonlinear model of X-22 VTOL aircraft containing both process and measurement noise.

(ii) HL-10 Lifting Body flight test data without gusts.

(iii) M2/F3 Lifting Body flight test data containing gusts.

The Output Error and Equation Error methods which were also tried gave either poor results or failed to converge on problems (i) and (iii).

Additional problems of identifiability were encountered in the processing of the HL-10 and M2/F3 flight test data. Those problems were manifested

as a) opposite signs on the parameter estimates compared to wind-tunnel and theoretical values, b) large parameter covariances, and c) high correlation between the parameter estimates. Three different methods were tried for alleviating these problems:

(i) A priori weighting

(ii) Fixing certain parameters

and (iii) Rank-deficient solutions.

The last method seems to hold the greatest promise for further developments.

The second part of the report discusses the theory and the computation of optimal inputs for linear system identification. The criterion used for input design is the maximization of the output sensitivity to parameter variations. This criterion is related to the Fisher Information matrix and the Cramer-Rao lower bound for the covariance of the parameter estimates. The objective of this effort was to determine the control surface deflections that enhance the effectiveness of identifying stability and control derivatives from flight test data. The specific application considered was the longitudinal dynamics of a C-8 aircraft. The optimal elevator surface deflection time histories were computed and compared with a doublet input of the same energy and time duration. A Monte Carlo simulation was done using the optimal and the doublet inputs and identifying short-period parameters from noisy output data. The optimal inputs were shown to provide more accurate estimates of aircraft parameters compared to a doublet input.

X

AREAS FOR FURTHER INVESTIGATION

The following areas are proposed for further research on and practical application of the maximum likelihood identification method and the design of optimal inputs:

(i) Complete preflight simulation of the Integrated Aircraft Identification process for a future flight test program.

(ii) Identification of stability and control derivatives from flight test data for specific important (e.g., high angle-of-attack , transonic) flight regimes. This would require the development of methods for model structure determination since the aerodynamic models in such flight regimes are not well known.

(iii) Correlation of the flight test results with the wind-tunnel results for high performance aircraft.

(iv) Determination of the effect of SAS on the identifiability of the parameters.

(v) Further developments of the Rank-Deficient Method for solving identifiability problems.

(vi) Extensions of the Input Design procedure to nonlinear models and process noise.

(vii) Modifications of the optimal inputs for pilot acceptability and ease of implementation.

(viii) Flight Test validation of optimal inputs.

APPENDIX A

EQUATION OF MOTION FOR X-22 VTOL

The model for the equations of motion for the X-22 was designed to represent the longitudinal-vertical three degree of freedom motions of the vehicle in the body axis coordinate system. Defining

u = change in velocity along the x axis from trim condition

w = change in velocity along the z axis from turn condition

θ = change in pitch angle (fuselage attitude relative to the horizontal) from the trim value

q = change in pitch rate from zero trim value; since vehicle is restricted to longitudinal-vertical, wings level motion only, $q = \dot{\theta}$

δ_{es} = change in elevator stick deflection (positive δ_{es} gives positive pitching moment) from trim condition

n_x = accelerometer signal along the x axis

n_z = accelerometer signal along the z axis

the nonlinear model, derived from an examination of wind tunnel data, is given by

$$\begin{bmatrix} \dot{u} \\ \dot{w} \\ \dot{\theta} \\ \dot{q} \end{bmatrix} = \begin{bmatrix} X_w(u)w - qw - g \sin\theta \\ Z_w(u)w + qu + g\cos\theta \\ \dot{\theta} \\ M_w(u)w + M_q(u)q \end{bmatrix} + \begin{bmatrix} X_o(u) & X_{\delta es}(u) \\ Z_o(u) & Z_{\delta es}(u) \\ o & o \\ M_o(u) & M_{\delta es}(u) \end{bmatrix} \begin{bmatrix} 1 \\ \delta_{es} \end{bmatrix}$$

$$+ \begin{bmatrix} X_u(u) & X_w(u) & 0 \\ Z_u(u) & Z_w(u) & 0 \\ 0 & 0 & 0 \\ M_u(u) & M_w(u) & M_q(u) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

The stability and control derivatives are all expressed as first order functions of u (e.g. $M_w \triangleq M_{w_o} + M_{w_u} \cdot u$) except for $M_o(u)$, $X_o(u)$ and $Z_o(u)$ which are second order functions. The derivatives of $X_u(u)$, $Z_u(u)$ and $M_u(u)$ are given as $\frac{d}{du}(X_o(u) + X_w(u) \cdot w + X_{\delta es}(u) \cdot \delta_{es})$; $\frac{d}{du}(Z_o(u) + Z_w(u)w + Z_{\delta es}(u) \cdot \delta_{es})$ and $\frac{d}{du}(M_o(u) + M_w(u) \cdot w + M_q(u) \cdot q + M_{\delta es}(u) \cdot \delta_{es})$ respectively and v_1, v_2, and v_3 are independent, zero mean white noise sequences with (diagonal matrix) covariance Q. There are therefore a total of 23 parameters to identify, enumerated as

$$M_o \begin{pmatrix} 1 \\ u \\ u^2 \end{pmatrix} ; \quad M_w \begin{pmatrix} 1 \\ u \end{pmatrix} ; \quad M_q \begin{pmatrix} 1 \\ u \end{pmatrix} ; \quad M_{\delta es} \begin{pmatrix} 1 \\ u \end{pmatrix}$$

$$X_o \begin{pmatrix} 1 \\ u \\ u^2 \end{pmatrix} ; \quad X_w \begin{pmatrix} 1 \\ u \end{pmatrix} ; \quad X_{\delta es} \begin{pmatrix} 1 \\ u \end{pmatrix}$$

and $$Z_o \begin{pmatrix} 1 \\ u \\ u^2 \end{pmatrix} ; \quad Z_w \begin{pmatrix} 1 \\ u \end{pmatrix} ; \quad Z_{\delta es} \begin{pmatrix} 1 \\ u \end{pmatrix}$$

There are seven measurements being made of the aircraft state. These include the four state variables u,w,θ and q, the pitch acceleration, $\dot{q}$, and the two accelerometer outputs n_x and n_z. The equations for

the accelerometer outputs are

$$n_x = \frac{1}{g}(\dot{u} + qw) + \sin\theta$$

$$n_z = \frac{1}{g}(\dot{w} - qu) - \cos\theta$$

Substituting in $\dot{u}$ and $\dot{w}$, these expressions can be written as

$$\begin{bmatrix} g \cdot n_x \\ g \cdot n_z \end{bmatrix} = \begin{bmatrix} 0 & X_w(u) & 0 & 0 \\ 0 & Z_w(u) & 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ w \\ \theta \\ q \end{bmatrix} + \begin{bmatrix} X_o(u) & X_{\delta es}(u) \\ Z_o(u) & Z_{\delta es}(u) \end{bmatrix} \begin{bmatrix} 1 \\ \delta_{es} \end{bmatrix}$$

$$+ \begin{bmatrix} X_u(u) & X_w(u) & 0 \\ Z_u(u) & Z_w(u) & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

APPENDIX B

GRADIENT AND INFORMATION MATRIX CALCULATION WITH ADDED PARTIAL DERIVATIVE TERMS

Since the presence of the process noise requires the use of a Kalman filter in computing the sensitivity functions, the exact equations for the gradient and information matrix will include the additional partials of the Kalman gain and state estimate error covariance with respect to the parameters. In those cases where the Kalman gain reaches a steady state, this steady state value can often be included as unknown parameters (with the only error being during the transient period). However, whenever the system equations are non-linear, as with the X-22 model, the Kalman filter will not reach a steady state and added partials will appear.

Using the notation of Section 5.1, the gradient and information computations as well as the sensitivity function computations are given below. The additional terms included in these calculations, due to the time varying Kalman gain, are indicated by a $\boxed{}$. The term p_j is used to denote the jth entry in the vector of unknown parameters.

Gradient:

$$\frac{\partial J}{\partial p_j} = -\frac{1}{N} \sum_{i=1}^{N} \begin{pmatrix} z_i - \hat{x}_{i/i-1} \\ \\ y_i - \hat{x}_{i/i-1} \end{pmatrix}^T \sum_{i/i-1}^{-1} \left(\begin{array}{c} \dfrac{\partial \hat{x}_{i/i-1}}{\partial p_j} \\ \\ \dfrac{\partial f}{\partial \hat{x}} \dfrac{\partial \hat{x}_{i/i-1}}{\partial p_j} + \dfrac{\partial f'}{\partial p_j} \end{array} \right) -$$

$$- \frac{1}{2N} \sum_{i=1}^{N} \begin{pmatrix} z_i - \hat{x}_{i/i-1} \\ y_i - \hat{\dot{x}}_{i/i-1} \end{pmatrix}^T \sum_{i/i-1}^{-1} \frac{\partial \sum_{i/i-1}}{\partial p_j} \sum_{i/i-1}^{-1}$$

$$\begin{pmatrix} z_i - \hat{x}_{i/i-1} \\ y_i - \hat{\dot{x}}_{i/i-1} \end{pmatrix}$$

where $$H_1 = \frac{\partial}{\partial x} \left\{ \begin{bmatrix} \hat{x}_{i/i-1} \\ \hline f' \end{bmatrix} \right\}$$

$$\sum_{i/i-1} = H_i \; P_{i/i-1} \; H_i^T + R$$

$$\frac{\partial \sum_{i/i-1}}{\partial p_j} = \frac{\partial H_i}{\partial p_j} \; P_{i/i-1} \; H_i^T + H_i \frac{\partial P_{i/i-1}}{\partial p_j} H_i^T + H_i \; P_{i/i-1} \frac{\partial H_i^T}{\partial p_j}$$

$$+ \quad \frac{\partial}{\partial P_j} \quad \left\{ R + \begin{bmatrix} 0 & 0 \\ \hline 0 & gQg^T \end{bmatrix} \right\}$$

Information Matrix:

$$\frac{\partial^2 J}{\partial P_K \partial P_j} = \frac{1}{N}\sum_{i=1}^{N}\left(\begin{array}{c}\frac{\partial \dot{\hat{x}}_{i/i-1}}{\partial P_K} \\ B_i \quad \frac{\partial \hat{x}_{i/i-1}}{\partial P_K} + \frac{\partial f'}{\partial P_K}\end{array}\right)^T \sum_{i/i-1}^{-1}\left(\begin{array}{c}\frac{\partial \hat{x}_{i/i-1}}{\partial P_j} \\ B_i \frac{\partial \hat{x}_{i/i-1}}{\partial P_j} + \frac{\partial f'}{\partial P_j}\end{array}\right)$$

$$+\frac{1}{N}\sum_{i=1}^{N}\left(\begin{array}{c}z_i - \hat{x}_{i/i-1} \\ y_i - \dot{\hat{x}}_{i/i-1}\end{array}\right)^T \sum_{i/i-1}^{-1} \frac{\partial \sum_{i/i-1}}{\partial P_K} \sum_{i/i-1}^{-1} \cdot \left(\begin{array}{c}\frac{\partial \hat{x}_{i/i-1}}{\partial P_j} \\ B_i \frac{\partial \hat{x}_{i/i-1}}{\partial P_j} + \frac{\partial f'}{\partial P_j}\end{array}\right)$$

$$+\frac{1}{N}\sum_{i=1}^{N}\left(\begin{array}{c}z_i - \hat{x}_{i/i-1} \\ y_i - \dot{\hat{x}}_{i/i-1}\end{array}\right)^T \sum_{i/i-1}^{-1} \frac{\partial \sum_{i/i-1}}{\partial P_j} \sum_{i/i-1}^{-1} \cdot$$

$$\cdot \left(\begin{array}{c}\frac{\partial \hat{x}_{i/i-1}}{\partial P_K} \\ B_i \frac{\partial \hat{x}_{i/i-1}}{\partial P_K} + \frac{\partial f'}{\partial P_K}\end{array}\right)$$

Sensitivity Function:

Between measurements:

$$\frac{d}{dt}\left(\frac{\partial \hat{x}}{\partial p}\right) = \frac{\partial f}{\partial \hat{x}} \cdot \frac{\partial \hat{x}}{\partial p} + \frac{\partial f}{\partial p}$$

At a measurement:

$$\frac{\partial \hat{X}}{\partial P_j}\left(t_i^{+}\right) = \frac{\partial \hat{X}}{\partial P_j}\left(t_i^{-}\right) - W_i \begin{pmatrix} \dfrac{\partial \hat{X}_{i/i-1}}{\partial P_j} \\ \dfrac{\partial f'}{\partial \hat{x}} \dfrac{\partial \hat{X}_{i/i-1}}{\partial P_j} + \dfrac{\partial f'}{\partial P_j} \end{pmatrix} + \frac{\partial W_i}{\partial P_j} \begin{bmatrix} z_i - \hat{x}_{i/i-1} \\ y_i - \dot{\hat{x}}_{i/i-1} \end{bmatrix}$$

where

$$\frac{\partial W_i}{\partial P_j} = \frac{\partial P_{i/i-1}}{\partial P_j} H_i^T \sum\nolimits_{i/i-1}^{-1} + P_{i/i-1} \left(\frac{\partial H_i}{\partial P_j}\right)^T \sum\nolimits_{i/i-1}^{-1}$$

$$- P_{i/i-1} H_i^T \sum\nolimits_{i/i-1}^{-1} \frac{\partial \sum_{i/i-1}}{\partial P_j} \sum\nolimits_{i/i-1}^{-1}$$

$$\frac{\partial P_{i/i-1}}{\partial P_j} = (I + F_i \Delta t) \frac{\partial P_{i/i}}{\partial P_j} (I + F_i \Delta t)^T + \frac{\partial}{\partial P_j}(GQG^T) \Delta t^2$$

$$+ \frac{\partial F_i}{\partial P_j} \Delta t \; P_{i/i} (I + F_i \Delta t)^T + (I + F_i \Delta t) P_{i/i} \left(\frac{\partial F_i}{\partial P_j}\right)^T \Delta t$$

$$\frac{\partial P_{i/i}}{\partial P_j} = (I - W_i H_i) \frac{\partial P_{i/i-1}}{\partial P_j} - \left(\frac{\partial W_i}{\partial P_j} H_i + W_i \frac{\partial H_i}{\partial P_j}\right) P_{i/i-1}$$

and $F_i = \left.\dfrac{\partial f}{\partial \hat{x}}\right|_{t=t_i}$

LIST OF REFERENCES

1. C. H. Wolcowicz, "Considerations in the Determination of Stability and Control Derivatives and Dynamic Characteristics from Flight Data", AGARD Report 549, Part I

2. C. H. Wolcowicz, K. W. Iliff, G. B. Gilyard, "Flight Test Experience in Aircraft Parameter Identification", presented at AGARD Symposium on Stability and Control, Braunschweig, Germany, April 1972

3. R. Turley, Private Communication, U.S. Air Force Flight Test Center, April 1971.

4. T. M. Kastner, J. A. Eney and J. J. McCue, "Flight Evaluation of Various Short Period Dynamics in the Variable Stability F-80", NATC TR FT-13R-TO.

5. H. Rediess, Private Communication, NASA Flight Test Center, October 1971.

6. Dante DiFranco, "In-Flight Parameter Identification by the Equation-of-Motion Technique -- Application to the Variability Stability T-33 Airplane", Cornell Aeronautical Laboratory Report No. TC-1921-F-3, 15 December 1965.

7. O. H. Gerlach, "Determination of Performance and Stability Parameters from Non-Steady Flight Test Maneuvers", SAE National Business Aircraft Meeting, Wichita, Kansas, March 1970.

8. G.C. Goodwin, "Application of Curvature Methods to Parameter and State Estimation", Proc. IEEE, Vol. 16, No. 6, June 1969.

9. L. Taylor,et. al., "A Comparison of Newton-Raphson and Other Methods for Determining Stability Derivatives from Flight Data", Third Technical Workshop on Dynamic Stability Problems, Ames Research Center, 1968. Also presented at AIAA Third Flight Test, Simulation and Support Conference, Houston, Texas, March 1969.

10. R. Bellman, et. al., "Quasilinearization, System Identification, and Prediction, RAND Corporation RM-3812, August 1963.

11. K.S.P. Kumar and R. Shridhar, "On the Identification of Control Systems by the Quasilinearization Method", IEEE Trans., Vol. AC-10, pp 151-154, April 1964.

12. D. Larson, "Identification of Parameters by Method of Quasilinearization", CAL Report 164, May 1968.

13. D. G. Denery, "An Identification Algorithm Which is Insensitive to Initial Parameter Estimates", AIAA Eighth Aerospace Science Conference, January 1970.

14. P.C. Young, "Process Parameter Estimation and Adaptive Control", In Theory of Self-Adaptive Control Systems, P. Hammond, ed., Plenum Press, New York, 1966.

15. R. D. Schalow, "Quasilinearization and Parameter Estimation Accuracy", Ph.D. Thesis, Syracuse University, 1967.

16. B. Dolbin, "A Differential Correction Method for the Identification of Airplane Parameters from Flight Test Data", University of Buffalo, Masters Thesis, December 1968.

17. L. S. Lason, et. al., "The Conjugate Gradient Method for Optimal Control Problems", IEEE Trans., G-AL, Vol. 12, No. 2, April 1967.

18. R. Wingrove, Private Communication, NASA Ames Research Center, October 1971.

19. W.T. Suit, "Aerodynamic Parameters of the Navion Airplane Extracted from Flight Data", NASA TN D-6643, March 1972.

20. G.S. Steinmetz, R.V. Parrish, and R.L. Bowles, "Longitudinal Stability and Control Derivatives of a Jet Fighter Airplane Extracted from Flight Test Data by Utilizing Maximum Likelihood Estimation", NASA TN D-6532, March 1972.

21. J.S. Tyler, J.D. Powell, R. K. Mehra, "The use of Smoothing and Other Advanced Techniques for VTOL Aircraft Parameter Identification", Final Report to Cornell Aeronautical Laboratory under Naval Air Systems Command Contract No. N00019-69-C-0534, June 1970.

22. K.J. Astrom and S. Wenmark, "Numerical Identification of Stationary Time Series", Sixth International Instruments and Measurements Congress, Stockholm, Sept. 1964.

23. R.L. Kashyap, "Maximum Likelihood Identification of Stochastic Linear Dynamic Systems", IEEE Trans. AC, Feb. 1970.

24. R.K. Mehra, "Identification of Stochastic Linear Dynamic Systems Using Kalman Filter Representation", AIAA Journal, Vol. 9, No. 1, Jan. 1971

25. A. E. Bryson and Y.C. Ho, Applied Optimal Control, Blaisdell Publishing Co., Waltham, Mass., 1969.

26. R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Trans. ASME, J. Basic Eng. Vol. 82, March 1960.

27. T. Kailath, "A General Likelihood - Ratio Formula for Random Signals in Gaussian Noise", IEEE Trans. Info. Theory, Vol. IT-15, May 1969.

28. A. Papoulis, Probability, Random Variables and Stochastic Processes, McGraw-Hill, New York, 1965.

29. T. Kailath, "An Innovations Approach to Least-Squares Estimation, Part I", IEEE Trans. on AC, Vol AC-13, No. 6, Dec. 1968.

30. R. E. Larson, R. M. Dressler, and R.S. Ratner, "Application of the Extended Kalman Filter to Ballistic Trajectory Estimation", Final Report, SRI, Proj. 5188-103, Jan. 1967.

31. R. K. Mehra, "A Comparison of Several Non-linear Filters for Radar Target Tracking", IEEE T-AC, Vol. AC-16, No. 4, Aug. 1971.

32. R.T. Chen, B.J. Eulrich and J.V. Lebacgz, "Development of Advanced Techniques for the Identification of V/STOL Aircraft Stability and Control Parameters", CAL Report No. BM-2820-F-1, Aug. 1971.

33. R.K. Mehra, "Maximum Likelihood Identification of Aircraft Parameters", 1970 JACC, Atlanta, Georgia

34. R. Bellman, Introduction to Matrix Analysis, McGraw-Hill, New York, 1970

35. R. J. Hanson and C. L. Lawson, "Extensions and Applications of the Householder Algorithm for Solving Linear Least Squares Problems," Math. of Comp., October 1969.

36. D.G. Denery, "Simplification in the Computation of the Sensitivity Functions for Constant Coefficient Linear System", IEEE Trans. on AC, Aug. 71.

37. R.K. Mehra and R.E. Davis, "A Generalized Gradient Method for Optimal Control Problems with Inequality Constraints and Singular Arcs", SCI Research Report No. 2, Sept. 1970.

38. H. Akaike, "Statistical Predictor Identification", Ann. Inst. Statist. Math., Vol. 22, 1970.

39. J. Sorensen, "Analysis of Instrumentation Error Effects on the Identification Accuracy of Aircraft Parameters", Contract NAS 1-10791, May 1972. NASA CR-112121

40. C. L. Muzzey and E.A. Kidd, "Measurement and Interpretation of Flight Test Data for Dynamic Stability and Control", Chapter 11, Vol II, AGARD Flight Test Manual, Pergammon Press, 1959.

41. H. Greenberg, "A Survey of Methods for Determining Stability Parameters of an Airplane from Dynamic Flight Measurements", NASA TN 2340, April 1951.

42. R.K. Mehra, "Optimal Inputs for Linear System Identification, Part I - Theory", JACC, Stanford, California 1972.

43. P.A.N. Briggs, K.R. Godfrey and P.H. Hammond, "Estimation of Process Dynamic Characteristics by Correlation Methods Using Pseudo-Random Signals", I.F.A.C. Symposium on Identification and Process Parameter Estimation, Prague, June 1967.

44. K.R. Godfrey, "The Appplication of Pseudo-Random Sequences to Industrial Processes and Nuclear Power Plants", I.F.A.C. Symposium on Identification and Process Parameter Estimation, Prague, 1970.

45. I.G. Cumming, "Frequency of Input Signal in Identification", I.F.A.C. Symposium on Identification and Process Parameter Estimation, Prague 1970.

46. M. Aoki and R.M. Staley, "On Input Signal Synthesis in Parameter Identification", Automatica, Vol. 6, 1970.

47. V.C. Levadi, "Design of Input Signal for Parameter Estimation," IEEE G-AC, Vol. AC-11, No. 2, April 1966.

48. N.E. Nahi and D.E. Wallis, Jr., "Optimal Inputs for Parameter Estimation in Dynamic Systems with White Noise Observation Noise," Preprints, JACC Boulder, Colo. Aug. 1969.

49. M.J. Levine, "Estimation of a System Pulse Transfer Function in the Presence of Noise," IEEE AC, 1964, pp 229-235.

50. R.J. McAulay, "Optimal Control Techniques Applied to PPM Signal Design", Information and Control 12, 1968, pp 221-235.

51. R. Esposito and M.A. Schumer, "Probing Linear Filters -- Signal Design for the Detection Problem", IEEE Trans Information Theory, Vol T-16, No 2, March 1970.

52. L.S. Pontryagin, V. Boltyanskii and E. Mishehenko, The Mathematical Theory of Optimal Processes, New York, Interscience, 1962.

53. A.E. Bryson and D.E. Johansen, "Linear Filtering for Time-Varying Systems Using Measurements Containing Colored Noise," IEEE Trans. AC Vol. AC-10 pp 4-10, January 1965.

54. R.E. Kalman, "New Methods and Results in Linear Prediction and Filtering Theory," Proc. Symp. on Engineering Applications of Random Function Theory and Probability," New York, Wiley, 1961.

55. R.E. Kalman and T.S. Englar, "A User's Manual for the Automatic Synthesis Program," NASA CR-475, June 1966.

56. A. Ralston, A First Course in Numerical Analysis, McGraw-Hill, New York, 1965.

57. R. K. Mehra, "On-Line Identification of Linear Dynamic Systems with Applications to Kalman Filtering", IEEE Trans. Auto. Control, April 1970

U.S. GOVERNMENT PRINTING OFFICE: 1971 - 716-412, REGION NO. 8-11