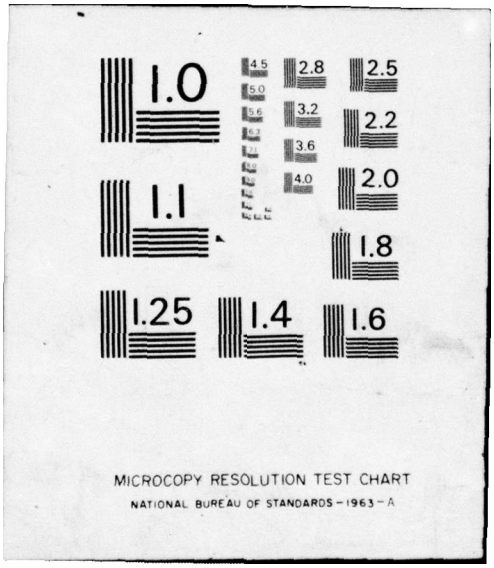
AD-A032 145 SYSTEMS CONTROL INC PALO ALTO CALIF F/G 12/1
METHODS FOR THE REAL TIME IDENTIFICATION OF VEHICLE PARAMETERS.(U)
FEB 75 N K GUPTA, W E HALL N00014-72-C-0328
UNCLASSIFIED NL

1 OF 2
AD A032145


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS - 1963 - A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER (9)	
4. TITLE (and Subtitle) (6)	5. TYPE OF REPORT & PERIOD COVERED (12)
METHODS FOR THE REAL TIME IDENTIFICATION OF VEHICLE PARAMETERS	Engineering Technical Report, 1 Jan 1974 - 31 Dec 1974
	6. PERFORMING ORG. REPORT NUMBER 1 Jan-31 Dec 74
7. AUTHOR(s) (10)	8. CONTRACT OR GRANT NUMBER(s) (15)
N.K. Gupta and W.E. Hall	N00014-72-C-0328 NEW
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Systems Control, Inc. (Vtae) 1801 Page Mill Road Palo Alto, CA 94304 389333	5984
11. CONTROLLING OFFICE NAME AND ADDRESS (11)	12. REPORT DATE
Office of Naval Research 800 North Quincy Road Arlington, VA 22217	February 1975
	13. NUMBER OF PAGES 95
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 119 p.	15. SECURITY CLASS. (of this report)
Same	UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Parameter Identification
Real Time Parameter Identification
On-line Identification
SCIDNT

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report presents two methods for on-line and real time estimation of parameters for application to aircraft, missiles, surface and subsurface marine vehicles and various subsystems. The first method is based on the instrumental variables approach. A new technique for choosing the instrument matrices is developed. This technique ensures algorithm convergence even with bad initial conditions and when the system parameters change. The second method uses the sensitivity functions reductions technique in the maximum likelihood approach to → NEXT PAGE

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

389333

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. (Continued) give a viable on-line or real time method. A maximum likelihood approach is also suggested for nonlinear systems. The trade offs of the two methods are discussed.

These techniques are applied to simulation and flight test data. Excellent results are obtained in both cases.

ACKNOWLEDGMENT

This report is prepared for the Office of Naval Research, Arlington, Virginia under Contract N00014-72-C-0328. This work was performed between Januarly 1, 1974 and December 31, 1974. The technical monitor at ONR is Mr. David Seigel. This project was supervised by J.S. Tyler, Jr., and managed by W.E. Hall, Jr. N.K. Gupta and W.E. Hall, Jr. are the project engineers. Report preparation and artwork were supervised by D.A. Buenz.

The authors with to thank Prof. R.K. Mehra of Harvard University, Mr. David Siegel of ONR and their colleagues at Systems Control, Inc. (Vt) for many stimulating discussions on the subject of this report.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONCLUDED)

I. INTRODUCTION AND SUMMARY

1.1 INTRODUCTION

The requirement to reduce test and evaluation time has become a primary objective in Navy systems development. This requirement, however, must be applied to increasingly more complicated and sophisticated hardware, such as aircraft, surface and subsurface marine vehicles, and inertial navigation components. The result is increasing dependence on advanced data acqusition systems by research, test, and evaluation agencies. To maximize the effectiveness of such data acquisition capabilities, specialized software algorithms are needed to provide test system evaluation parameters in the minimum engineering and test time. These software programs, for example, are developed to quantify aircraft performance, calibrate instrumentation, and monitor critical subsystems (e.g., engines).

One important type of algorithm is parameter identification. These algorithms use test data to estimate the fundamental parameters which govern the response characteristics of the particular system being evaluated. Advanced parameter identification techniques, for example, have been implemented and successfully used to reduce evaluation time for stability and control parameters of advanced Naval aircraft [1]. These applications have demonstrated the potential benefits of even more advanced parameter identification algorithms which can be used for real time calculations (e.g., "as the data comes in").*

*For the purposes of this report, "real time" implies available calculation results within a test data length; "near real time" implies results which are available sufficiently soon after a particular data length that sequential data lengths can be continually monitored without interruption; "on-line" implies that sensor data is passed to processing computers as the data is generated; "off-line" implies storage of data to be processed at a later time.

Such algorithms cannot only be used for substantial reductions in testing and processing time, but also for system diagnostic monitoring and control. A summary of such applications is given in Table 1.

The objective of the analyses conducted for this effort has been to develop the fundamental theoretical and applications methods for providing real and near-real time parameter estimates from test data. The principal results are summarized in the following section.

1.2 PRINCIPAL RESULTS

There are basically two approaches to developing a real time parameter identification algorithm. These are as follows:

a. Simplification of an Off-Line Algorithm: In order to optimally process a data set, a minimal number of assumptions are evoked. The result is a complex computational algorithm requiring sufficient execution time and machine memory allocation that only post-test calculations are practical. By restricting the test system analytical complexity (e.g., limiting to linear systems), by making the program more dependent on a priori start-up data, and by a general "streamlining" of the program structure, significant reductions in computer resources are achievable.

b. Development of a Real Time Algorithm: An alternate approach is to formulate a parameter identification algorithm within the constraint of real-time capability. This approach therefore evokes simplifying assumptions at the beginning, instead of the end, of the algorithm development.

In summary, one may simplify an off-line algorithm, which was developed for accuracy, not speed, or one may specifically develop a fast algorithm, with possible compromises on accuracy. Both approaches have been developed and evaluated for this work. In particular, the following methods have been developed and applied to flight test data.

Table 1

Summary of Applications of On-Line Parameter Identification Systems

ALGORITHM	PURPOSE	APPLICATION
● Data Quality Analysis	Processing of data yields on-line assessment of information content of measurements. If data suitable for further processing, a "GO" condition is displayed, and test passes to next stage. If data not suitable, do test again, possibly discarding old data.	Reduces test time and amount of extraneous or redundant data. In either case, leads to cost savings.
● Fault Detection	Analyzes data for sudden changes in output statistical signature (i.e., mean, variance). Does not give fault isolation without further processing.	Provides on-line verification of system integrity before and during test.
● Parameter Identification	To determine, on-line, values of significant parameters of process. Not all parameters are necessarily identified since some may be known exactly.	Real time system identification is required in RPV, navigation systems, as well as high performance aircraft, where some parameters change unpredictably.
● Fault Isolation	An extension of parameter identification, this processing analyzes deviations in parameter estimates from various data channels to isolate failure mode (e.g., structure, control, or instrumentation).	Provides internal redundancy of instruments instead of equipment redundancy, thus reducing equipment cost. Once established as reliable technique, can be used as an on-line safety feature.
● Adaptive Control	A continuing objective of much research, adaptive control changes process characteristics to minimize response to disturbances. On-line parameter identification is basic to this task.	Performance envelope expansion of aircraft, missiles, ships.

a. Near Real Time Maximum Likelihood Algorithm: The implementation of a new algorithm into an advanced maximum likelihood parameter identification program has been found to reduce execution time by a factor of 5:1. Denoted as sensitivity function reduction, this algorithm provides the potential of even further minimization of execution time for linear systems.

b. Real Time Instrumental Variable Method: A new formulation of a sequential least square filter based on the statistical method of instrumental variables, has produced a real time capability for providing parameter estimates for linear systems.

In addition to these implemented algorithms, three other related algorithms have been developed for possible future implementation. These are as follows:

a. Real Time Maximum Likelihood Method: A unique reformulation of the maximum likelihood method has been completed. This algorithm promises estimates of high statistical efficiency.

b. On-Line Data Consistency Analysis: A recurring problem in system testing is the poor quality of certain key data from low accuracy or even partially failed instruments. Miscalibrations, data channel losses, or spurious noise inputs contribute to such problems. To alleviate attempts to process such data, whose results may be marginal, a method for assessing the relative accuracy of data channels has been developed.

c. On-Line Data Quality Analysis: One further step in data quality analysis is the determination of the information about a desired parameter set. This quality, denoted as identifiability, provides an assessment of the test input, instrumentation accuracy, and correctness of the assumed model for the system under test. A new technique for evaluating the identifiability of data has thus been formulated to improve the overall identification process.

1.3 SUMMARY

This report details the theoretical basis of real time parameter identification algorithms. Of necessity, the formulation of these algorithms is of a basic mathematical nature, and Chapter II provides the background concepts and definitions upon which the subsequent algorithmic developments rest. Chapter III details a new on-line instrumental variable method which overcomes many problems encountered in previous work. Also presented is a unique maximum likelihood algorithm for real time processing requirements. Chapter IV presents the implementation of the instrumental variable method and results. Some results on on-line evaluation of data quality are given in Chapter V.

II. REQUIREMENTS FOR ON-LINE PARAMETER IDENTIFICATION AND REVIEW OF PREVIOUS METHODS

This chapter reviews the identification problem with respect to typical computational requirements imposed by accurately estimating a large number of parameters (Section 2.1). Then, a background of mathematical terminology and system representations is provided to classify the constraints which must be considered in reducing these computational requirements (Sections 2.2 and 2.3). A review of previous approaches is then presented which introduces the developments of the following Chapter III.

2.1 COMPUTATIONAL REQUIREMENTS FOR ADVANCED PARAMETER IDENTIFICATION ALGORITHMS

In general, there is no universal specification for the length of time or computer capability needed to estimate the coefficients of a specific system from input-output data. Such computational requirements are functions of the type of system (e.g., linear or nonlinear, dimension of the system), the data length, the number of parameters to be identified, and the estimate accuracy desired. These are the basic factors governing the computer requirement. In general, as also discussed in Chapter I, there is a trade off between computer requirements and estimate accuracy. This trade off has not be quantified in general due to the difficulty (and expense) of trying to do so for a practical problem, with many state and parameters to be identified, over all possible types of computers, algorithms, and data lengths.

For reference, a typical, highly accurate off-line parameter identification program can be used to establish a basis for discussing computer requirement reduction. A general-purpose (e.g., linear or nonlinear system capability) maximum likelihood computer program, for example, has a computer time requirement as shown in Figure 1. This computer program will identify an arbitraily large number of parameters of the system dynamics, the measure-

ment instrumentation errors, and process noise characteristics. The system may have linear or nonlinear dynamics and the program is designed to be able to treat nonlinear aerodynamics with coding modification. Figure 1 shows, for two cases of data point number N, the approximate increase in execution time versus number of parameters identified. For example, about six minutes are required (on a UNIVAC 1108) to estimate ten parameters from poor initial estimates. This program has not been speed optimized, however, by programming some of the subroutines in machine language or specializing the coding to that required for a linear system.

The characteristic times shown in Figure 1 may not be satisfactory for operational on-line use of the program. The question is then as to whether the slope of this characteristic can be decreased, and what accuracy compromises are required, if any, in so doing.

It will be shown in subsequent chapters that the characteristic slope of Figure 1 can in fact be reduced, at no loss in accuracy if the program is applied only to linear systems. Before proceeding to these new results, however, it is necessary to establish more precise meanings of parameter estimate accuracy, and review some previous work.

2.2 PROPERTIES OF PARAMETER ESTIMATES AND IDENTIFICATION ALGORITHMS

To illustrate the various properties of parameter estimates and identification algorithms, consider a system parameterized on m parameters whose true values θ are not known. Suppose that, based on a certain set of data, x, an identification technique gives the estimate $\hat{\theta}$ for θ. In general, we are interested in knowing how close the estimates are to the true values of parameters. Various indicators of the closeness of $\hat{\theta}$ and θ are described by the statistical properties of the estimate $\hat{\theta}$, some of which are described as follows [2].

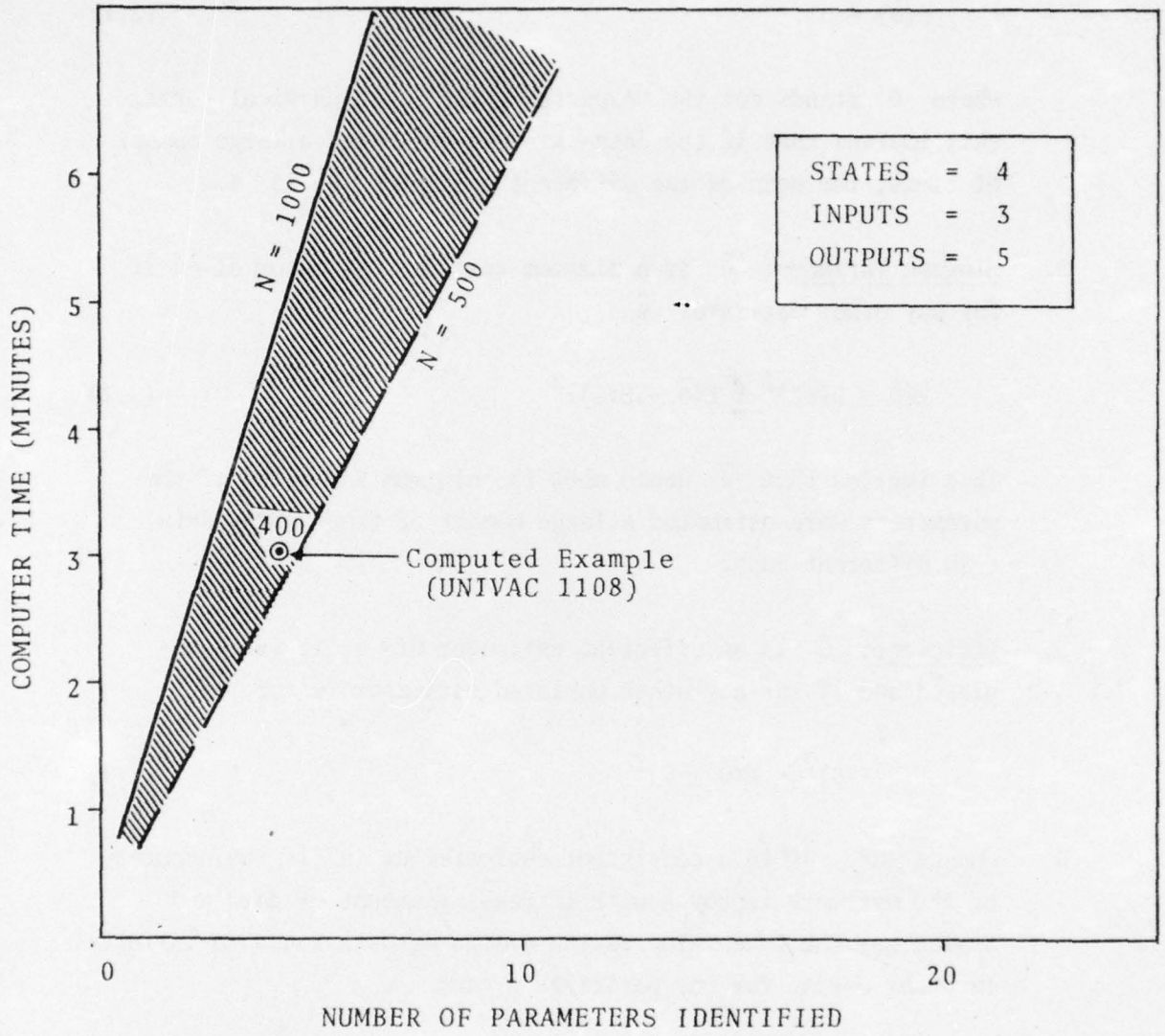


Figure 1 Computer Execution Time Versus Number of Parameters Identified for a General Purpose Maximum Likelihood Program

1. Bias: $\hat{\theta}$ is an unbiased estimator of θ if

$$E(\hat{\theta}) = \theta \tag{2.1}$$

where E stands for the "expected value". In physical terms, this implies that if the data x were collected a large number of times, the mean of the different estimates $\hat{\theta}$ is θ.

2. Minimum Variance: $\hat{\theta}$ is a minimum variance estimator of θ if for any other estimator $\tilde{\theta}$

$$E\{\hat{\theta} - E(\hat{\theta})\}^2 \leq E\{\tilde{\theta} - E(\tilde{\theta})\}^2 \tag{2.2}$$

This implies that $\hat{\theta}$ would show the minimum variation if the parameters were estimated a large number of times using data from different runs.

3. Efficient: $\hat{\theta}$ is an efficient estimator of θ if it is unbiased and if for any other unbiased estimator $\tilde{\theta}$ of θ

$$E(\hat{\theta} - \theta)^2 \leq E(\tilde{\theta} - \theta)^2 \tag{2.3}$$

4. Consistent: $\hat{\theta}$ is a consistent estimator of θ if the accuracy of the estimate improves with increasing amount of data and approaches the true value as the amount of data tends to infinity. In other words, for any positive η and ε

$$P(|\hat{\theta}_n - \theta| < \varepsilon) > 1 - \eta \qquad n > N \tag{2.4}$$

This definition is analogous to the definition of convergence in real analysis.

In addition to these statistical properties, the convergence of the algorithm should be assured even with poor starting values and when the parameters change either slowly or abruptly. This is extremely important, much more important than in an off-line algorithm, because the on-line parameters may be used in critical real time tasks where human intervension may be undesirable or infeasible. The convergence of the parameter estimation algorithm must, in other words, be robust with respect to starting parameter values.

At this point, we also define the concept of "probability in the limit". Let y_n be a random variable, which is a function of n. Then the probability in the limit of y_n is y, if the distribution of y_n collapses to a single point y as n becomes large. This is different than the convergence of the expected value of y_n to y.

2.3 TWO REPRESENTATIONS OF THE DYNAMIC SYSTEM

Two main representations have been used for linear time invariant systems, which can be described by a Markov process. These are: (a) state space form, and (b) autoregressive moving average (ARMA) form.

2.3.1 State Space Representation

The central concept in the state space representation of the system is the definition of state. At any point in time, it summarizes the past history of the system from the viewpoint of predicting the future response. Examples of state are the position and velocity of a particle acted upon by a force. If x is the state and y is the measurement, then the system representation in discrete form is

$$x(k+1) = \phi x(k) + Gu(k) + \Gamma w(k) \tag{2.5}$$

$$x(k) = Hx(k) + v(k) \tag{2.6}$$

The state equations (2.5) are driven by inputs u and random noise w (called process noise). Note that the determination of $x(k+1)$ requires the knowledge of $u(k)$, $w(k)$ and $x(k)$ but not $x(1)$, $x(2)$, ..., $x(k-1)$. ϕ, G and Γ describe the behavior of the system. The measurements y are also contaminated by random noise. The matrix H describes the instrumentation system. The number of measurements can be greater than, less than or equal to the number of states. This system can be written in the innovations representation

$$\hat{x}(k+1) = \phi\hat{x}(k) + Gu(k) + \phi K\nu(k) \tag{2.7}$$

$$y(k) = H\hat{x}(k) + \nu(k) \tag{2.8}$$

$\nu(k)$ are the innovations and are white, and K is the Kalman gain. In state estimation context, the innovations represent new information brought in by each measurement. The innovations representation is particularly useful in estimation and identification problems, since $\hat{x}(k)$ is the only component of $x(k)$ which can be estimated.

2.3.2 Autoregressive Moving Average (ARMA) Representation

For a multi-output, multi-input system, the autoregressive moving average representation is

$$\sum_{i=0}^{p-1} A_i\, y(k-i) = \sum_{i=1}^{q} \{B_i\, u(k-i) + C_i\, \nu(k-i)\}\ ,\quad A_o = C_1 = I \tag{2.9}$$

where $\nu(k)$ is white random noise. p is the order of the autoregressive part and q is the order of the moving average part. Notice that there is no concept of state in this representation. In effect, the descriptions of the system and the instrumentation have been combined in one equation. Equation (2.9) shows directly the effect of the input and the random noise on the output of the system.

2.3.3 Relationship Between the State Space Representation and ARMA Representation

A system which can be described by state space representation can also be described by an ARMA representation and vice versa. However, there are important differences, which often make the choice of one form more desirable over the other. It should be mentioned that certain canonical forms in state space representation are directly reducible to ARMA representation. On the other hand, by a suitable selection of the state vector the ARMA representation can be converted into a state space form.

The state space representation results from a knowledge of the physical laws which govern the system behavior. Therefore, the system state and instrument outputs are modeled separately. The ϕ, G, Γ and H have a specified structure. Examples of this are various physical processes. The ARMA representation is a natural consequence of systems in which the time series does not follow any well defined or known law. Examples of this are various economic series, biomedical time series and complicated physical processes. In each of these ARMA cases, a state variable model in canonical form could also be applied.

In short, the state space and the ARMA representations should be considered as two complimentary techniques to mathematically describe the response of linear dynamic systems.

2.4 REVIEW OF PREVIOUS ON-LINE IDENTIFICATION TECHNIQUES

Several on-line algorithms have been proposed in the past for both the state variable and the autoregressive moving average representations. These methods have tended to concentrate on ARMA representations and, to avoid the identifiability problem, assumed all parameters unknown and identifiable. The problems of statistical efficiency and bias and algorithm divergence have been addressed by various authors but are not completely resolved. There are a number of schemes to make the algorithms

recursive. These methods are covered in two recent survey papers by Saridis [3] and by Isermann, et al. [4]. Some of the recently proposed on-line schemes are in Reference 5.

All on-line methods can be classified as follows:

(1) Least Squares (LS) and Generalized Least Squares (GLS), Including Two Stage Least Squares (2SLS) and Three Stage Least Squares (3SLS).

(2) Stochastic Approximations (SAP).

(3) Kalman/Filter Smoother (KF/S).

(4) Instrumental Variables (IV).

(5) On-Line Maximum Likelihood (OLML).

The least squares methods have been in use for a long time, both in econometrics and in control applications. The basic appeal of these methods is their simplicity. The difference between the deterministic parts of the left hand side and the right hand side of the equation is minimized. For the ARMA representation, A_i and B_i are chosen to minimize

$$\sum_{k=p}^{N} \left\| \sum_{i=0}^{p-1} A_i\, y(k-i) - \sum_{i=1}^{q} B_i\, u(k-i) \right\|_Q^2 \qquad (2.10)$$

An explicit solution is straightforward. There are two problems with the estimate obtained by minimizing (2.10). First, since the measurements are correlated with noise, the estimates are biased. Secondly, since the measurement noise sequence is moving average, the estimates are inefficient. A method similar to Equation (2.10) can be used for state variable model.

Attempts at removing this bias and inefficiency in parameter estimates has given birth to the generalized least square and two and three stage least square methods. After the parameters A_i and B_i are estimated, the residuals are determined. The residuals are nonwhite and a model is fitted to the residuals. Consider a single-input, single-output ARMA representation and let $\hat{a}_i$ denote the estimated value of a_i. Let

$$e(k) = \sum_{i=0}^{p-1} \hat{a}_i \, y(k-i) - \sum_{i=1}^{q} \hat{b}_i \, u(k-i) \qquad (2.11)$$

and

$$e(k) = \eta(k) + \gamma_1 \, n(k-1) + \gamma_2 \, n(k-2) \ldots \gamma_{q'} \, n(k-q') \qquad (2.12)$$

where $\eta(k)$ is a white noise process. The order $q'+1$ is determined from significance tests and experience, while the parameters are determined from another least squares. The y and u time series are passed through the filter of Equation (2.12). The noise is then white and a new set of parameters $\hat{a}_i$ and $\hat{b}_i$ can be estimated. This process is repeated until the final estimates of a_i, b_i and γ_i are obtained. This is called the generalized least square method [6]. The two stage and three stage least square methods follow a similar approach. The least square methods can be recast into a recursive form [6, 7], making them very useful for on-line and real-time applications.

The stochastic approximation methods were developed by statisticians and introduced to the control literature by Saridis and others [8]. In essence, the stochastic approximation methods are extensions of gradient methods to stochastic problems. Their application in on-line and real time identification is motivated by the fact that they can be converted into recursive methods, in which the parameters are corrected by an appropriately weighted error correction term based on the observed and the expected output. The correction term can be chosen in a variety of ways. Its main appeal is simplicity and flexibility in implementation. The parameter estimates are consistent, but usually inefficient. To make the method more efficient, second order and accelerated stochastic approximation algorithms have been developed. The efficiency is still poor.

The Kalman filter/smoother approach for system state estimation has found some application in on-line and real time parameter identification. The parameters are appended to the state vector to give a larger augmented state. This converts an otherwise linear system into a nonlinear one. The augmented state is estimated using the Kalman filter approach for systems governed by nonlinear differential equations. The parameters and states can be simultaneously estimated on-line. This method has not found a wide application because of several problems. The estimates are biased even in simple systems. To start the algorithm, good estimates of parameters, their covariances, and process and measurement noise covariances are required. The algorithm tends to diverge and neglect measurements after a certain time because parameters constitute undisturbable states. To avoid this divergence, arbitrary white noise terms are added to the parameter equations. Thus, the equations governing the parameters become

$$\theta(k+1) = \theta(k) + \eta(k) \tag{2.13}$$

where $\eta(k)$ is a white noise process. The equations governing the augmented state are

$$\begin{bmatrix} x(k+1) \\ \hline \theta(k+1) \end{bmatrix} = \begin{bmatrix} \phi & A \\ \hline 0 & I \end{bmatrix} \begin{bmatrix} x(k) \\ \hline \theta(k) \end{bmatrix} + \begin{bmatrix} G \\ \hline 0 \end{bmatrix} u(k) + \begin{bmatrix} \Gamma & 0 \\ \hline 0 & I \end{bmatrix} \begin{bmatrix} w(k) \\ \hline \eta(k) \end{bmatrix} \tag{2.14}$$

$$y(k) = [H \,\vdots\, 0] \begin{bmatrix} x(k) \\ \hline \theta(k) \end{bmatrix} + v(k) \tag{2.15}$$

Matrix A depends upon the state and is obtained by linearizing the first two terms on the right hand side of Eq. (2.5) around the current state and control. The Kalman filter equations to estimate the augmented state are

$$\begin{bmatrix} \hat{x}(k+1) \\ \hline \hat{\theta}(k+1) \end{bmatrix} = \begin{bmatrix} \phi & A \\ \hline 0 & I \end{bmatrix} \begin{bmatrix} \hat{x}(k) \\ \hline \hat{\theta}(k) \end{bmatrix} + K(k)\,(y - H\hat{x}(k)) + \begin{bmatrix} G \\ \hline 0 \end{bmatrix} u(k) \tag{2.16}$$

The Kalman gains K and the estimation error covariance follow the usual equations. Note that the Kalman gain is a function of time. The Kalman filter method has been used by Chen [9].

The instrumental variable method is an extension of least squares techniques which are easy to implement but result in biased estimates whenever the measurements are noisy. The instrumental variables technique was developed for econometric applications to combine the ease of least squares type methods with a technique to give unbiased estimates. This is done in a very elegant way by the choice of the instrumental variables. The instrumental variables are chosen such that they are independent of the measurement noise. Consider a single output ARMA process. Choose $z(k)$ as the vector of instrumental variables. Then,

$$z(k)\, y(k) = z(k)\, \{y(k-1),\, y(k-2),\ldots,y(k-p+1),\, u(k-1),\ldots,u(k-q)\}\theta + z(k)\, \eta(k) \tag{2.17}$$

where θ is the vector of unknown parameters.

Summing this for the N data points

$$\sum_{k=1}^{N} z(k)\, y(k) = \sum_{k=1}^{N} z(k)\, y^*(k)\theta + \sum_{k=1}^{N} z(k)\, \eta(k) \tag{2.18}$$

$$y^*(k) \triangleq \{y(k-1),\, y(k-2),\ldots,y(k-p+1),\, u(k-1),\ldots,u(k-q)\}$$

and we can estimate θ using the equation

$$\hat{\theta} = \left\{ \sum_{k=1}^{N} z(k)\, y^*(k) \right\}^{-1} \sum_{k=1}^{N} z(k)\, y(k) \tag{2.19}$$

assuming that the first matrix is invertible.

It is straightforward to show that

$$E(\hat{\theta}) = \theta$$

$$\text{iff} \quad E \sum_{k=1}^{N} \eta(k)\, z(k) = 0 \tag{2.20}$$

In other words, the estimate of θ is unbiased if the instrumental variables are chosen to be independent of the noise process. This algorithm can be used with state space models also. For on-line applications, Eq. (2.19) can be converted into a recursive form. The main problem in this method is the choice of the instrumental variables. The technique has been described by Wong and Polak [10] and applied by Young [11] and Pandaya [12]. Both these techniques achieve efficiency in certain cases, but do not ensure convergence. Moreover, the results of Pandaya and Young are applicable to single output systems in ARMA representation.

The efficiency of the off-line maximum likelihood approach [13] has lead to several on-line schemes for parameter identification based on maximizing the likelihood function [14]. Basically, these techniques compute the first and the second gradients of the likelihood function and then take a single step of the Newton-Raphson gradient algorithm. This step can be updated as new measurements are received. There are two problems with past on-line maximum likelihood techniques. First, no attention is given to the efficient computation of the gradients of the likelihood function except in very simple autoregressive or autoregressive moving average models. Second, since the likelihood function is not exactly quadratic at the optimum parameter values, a single step of the Newton-Raphson does not provide sufficiently accurate parameter estimates. The advantage is that all parameters in the system model are not considered unknown.

2.5 SUMMARY

Considerable work has been done in the development of on-line and real time algorithms for identification. Nevertheless, it is clear from the brief discussion of this chapter that many important problems have been left unanswered. The need for developing algorithms which are statistically efficient and have guaranteed convergence cannot be overstressed. Further work is also required in algorithms for state variable models. For specific applications, the algorithms may be tailored to particular computers and systems.

In the work reported in the next chapter, convergent and efficient algorithms have been developed for state variable models. Two approaches are considered: (a) the instrumental variables approach, and (b) the maximum likelihood approach. Attention is also given to computation time and storage.

III. ON-LINE AND REAL TIME IDENTIFICATION OF PARAMETERS

3.1 INTRODUCTION

The last chapter presented a brief discussion of the several existing on-line techniques. Various trade offs in choosing specific on-line algorithms are also discussed. It is concluded that there is a need for statistically efficient and convergent algorithms for on-line and real time identification of parameters in state variable models. It is desired that such algorithms be implementable within small storage and computation time.

Two techniques have been developed to this end. They are: (a) the instrumental variable approach, and (b) the on-line maximum likelihood method. Attention is given to statistical properties like the unbiasedness, efficiency, and consistency. Convergence characteristics of these algorithms are also studied and improved, in some cases, with a slight degradation in the statistical properties. Computation time and storage requirements are minimized so that these techniques can be used in real time application.

Section 3.2 introduces the instrumental variable approach. Two cases are considered. In the first case, all the state variables are measured and in the second, only some of the states are measured. Identifiability problems are also discussed. This is followed by a discussion of the on-line maximum likelihood methods in Section 3.3. On-line methods are given for both linear and nonlinear systems. The last section discusses the relationship between the instrumental variables and on-line maximum likelihood methods and gives the advantages and disadvantages of each method.

3.2 THE INSTRUMENTAL VARIABLES APPROACH

As introduced in Chapter II, the instrumental variable approach is an equation error method in which the bias of the least square methods is removed by proper choice of the instrument matrix. The techniques of Wong and Polak [10], Young [11], and Pandaya [12] can be used with a single-input single-output system in canonical form or for a scalar autoregressive moving average (ARMA) process. Moreover, their methods are applicable only for systems in which the state equation or the ARMA equation is noise free but the outputs are contaminated by colored noise. All these authors select efficient instrumental variables with a possibility of divergence. Only Pandaya gives attention to the divergence problem. The technique, developed here, can be used with general multi-input multi-output systems when both process noise and measurement noise are present. The technique gives special attention to the convergence properties of the algorithm.

3.2.1 Problem Statement

Consider a discrete time linear system

$$x(k+1) = \phi x(k) + Gu(k) + w(k)$$

$$x(0) = x_o \tag{3.1}$$

and the measurements

$$z(k) = Hx(k) + Du(k) + v(k) \tag{3.2}$$

where $x(k)$ is an $n \times 1$ state vector, $u(k)$ is a $q \times 1$ input vector, and $y(k)$ is a $m \times 1$ output vector. $w(k)$ and $v(k)$ are white noise sources such that

$$E(w(k)) = 0$$

$$E(v(k)) = 0 \qquad k = 1, 2, \ldots \tag{3.3}$$

and

$$E(w(k)\ w^T(\ell)) = Q\ \delta_{k,\ell}$$

$$E(v(k)\ v^T(\ell)) = R\ \delta_{k,\ell}$$

$$E(w(k)\ v^T(\ell)) = 0 \qquad k,\ell = 1, 2, \ldots \tag{3.4}$$

ϕ, G are either constant or slowly varying matrices and are not known. The measurement distribution matrices H and D are assumed known. An alternate representation of the system is in terms of the innovations and predicted state estimate

$$\hat{x}(k+1) = \phi\hat{x}(k) + Gu(k) + \phi K\nu(k)$$

$$z(k) = H\hat{x}(k) + Du(k) + \nu(k) \tag{3.5}$$

where K is the Kalman gain which in "steady state" follows the equations (constant ϕ, Γ, Q, R and H)

$$P = \phi(I-KH)P\phi^T + \Gamma Q\Gamma^T$$

$$K = PH^T(HPH^T + R)^{-1} \tag{3.6}$$

where $\hat{x}(k)$ is the expected value of the state $x(k)$ given the measurements $z(1), z(2) \ldots z(k-1)$. It is well known that the innovation sequence $\nu(1), \nu(2) \ldots \nu(k)$ is a white noise process. Since both $z(k)$ and $u(k)$ are known, it is possible to rewrite the measurements by subtracting the contribution of linear combination of u . Thus it is general to assume that D is zero.

The on-line identification problem consists of tracking parameters in ϕ, G and possibly noise sources. For sake of simplicity, we consider two cases: (a) Case 1, where matrix H has rank n, and (b) Case 2, where matrix H has rank smaller than n. The first case leads to a simple instrumental variable algorithm.

3.2.2 Instrumental Variable Approach When the Observability Index is One (Rank (H) = n)

In aircraft application, usually, there are many measurements available. In most cases it is possible to reconstruct a noisy estimate of the state at any point by using noisy measurements at that point. This implies that the observability index is one and the rank of H is equal to the number of states. This is possible only if the number of measurements is not smaller than the dimension of the state vector.

When rank (H) is n, for any positive definite m x m matrix $\hat{R}$, $H^T\hat{R}^{-1}H$ is invertible. Premultiplying Equation (3.2) by

$$T = (H^T\hat{R}^{-1}H)^{-1} H^T\hat{R}^{-1} \tag{3.7}$$

we have

$$z_T(k) = x(k) + v_T(k) \tag{3.8}$$

where

$$z_T = Tz \quad \text{etc.} \tag{3.9}$$

and

$$E(v_T(k)\ v_T^T(\ell)) = TRT^T\delta_{k,\ell} \tag{3.10}$$

As a special case when H is square and of rank n

$$T = H^{-1} \tag{3.11}$$

It is clear from (3.7) and (3.10) that if $\hat{R} = R$

$$E(v_T(k)v_T^T(\ell) = H^T R^{-1} H\delta_{k,\ell} \tag{3.12}$$

Since the Fisher information matrix for a set of measurements is an integral of $H^T R^{-1} H$, it is clear that the above transformation on the measurements maintains the information content if $\hat{R} = R$. In general, R is not known, so the best possible approximation for R should be used in the transformation (3.7). (It can be shown that any other $\hat{R}$, which is not a multiple of R, reduces the information.) From now on in this section, subscript T will be removed from the variables and Eq. (3.2) with number of measurements equal to number of states and H equal to a unity matrix will be considered to be the general equations. Using (3.8) and (3.1):

$$z(k+1) = \phi z(k) + Gu(k) + w(k) + v(k+1) - \phi v(k) \tag{3.13}$$

Let
$$y(k) \triangleq \begin{bmatrix} z(k) \\ \hline u(k) \end{bmatrix} \tag{3.14}$$

$$\Theta \triangleq [\phi \,|\, G] \tag{3.15}$$

and
$$\eta(k) \triangleq w(k) + v(k+1) - \phi v(k) \tag{3.16}$$

Therefore, (3.11) becomes

$$z(k+1) = \Theta y(k) + \eta(k) \tag{3.17}$$

We now consider the choice of the instrumental variables. Their role has been explained in the previous chapter. To be useful, they must satisfy the following conditions:

(a) Since the noise term in Eq. (3.17) is a linear combination of $w(k)$, $v(k)$, and $v(k+1)$, it is necessary that the instrumental variables

be independent of these noise terms so that the estimates are unbiased. In other words, the selected instrumental variables should be independent of the measurements $y(k)$, $y(k+1)$, $y(k+2)$, $\ldots, y(N)$.

(b) The instrumental variables should be correlated with $y(k)$. In particular, if $y_I(k)$ is the instrumental variables vector

$$\text{plim } \frac{1}{N} \sum_k y(k)\, y_I^T(k) \quad \text{should be nonsingular} \tag{3.18}$$

where "plim" stands for probability in the limit. This condition is necessary for the estimates to be consistent. It will be shown later that the matrix in Eq. (3.18) is related to covariance of parameter estimation errors.

(c) The efficiency of the estimates depends on the cross-correlation between $y(k)$ and $y_I(k)$.

Consider two possible choices of the instrumental variables vector

$$\begin{aligned} &\text{(a)} \quad y_I^{(1)}(k) = \begin{bmatrix} z(k-1) \\ \hline u(k) \end{bmatrix} \\ &\text{(b)} \quad y_I^{(2)}(k) = \begin{bmatrix} \hat{z}(k) \\ \hline u(k) \end{bmatrix} = \begin{bmatrix} \hat{x}(k) \\ \hline u(k) \end{bmatrix} \end{aligned} \tag{3.19}$$

$\hat{x}(k)$ is an estimate of $x(k)$ based on $z(1), z(2), \ldots, z(k-1)$. Note, first, that both choices (a) and (b) satisfy condition (a), ensuring that the estimates are unbiased. Choice (a) does not use parameter values to determine the instruments. On the contrary, choice (b) implicitly implies the use of estimated parameters to obtain the best estimate of the current state from past measurements. In other words, the estimated parameters are "bootstrapped" for state estimation and determination of instrumental variables.

It will now be shown that both choices (a) and (b) have advantages. $y_I^{(1)}(k)$ may not have a high correlation with $y(k)$ because of two reasons: (a) both $z(k-1)$ and $z(k)$ are contaminated with random noise, and (b) the deterministic parts of $z(k-1)$ and $z(k)$ are not the same. However, this choice ensures that the necessary condition (b) for the instrumental variables is satisfied, because

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_k y(k)\, y_I^{(1)T}(k) = \left[\begin{array}{c|c} R_{yy}(1) & R_{yu}(0) \\ \hline R_{uy}(1) & R_{uu}(0) \end{array}\right] \tag{3.20}$$

which is nonzero for persistently exciting systems except in trivial cases. Thus, independent of the initial parameter values or the parameter values at any point in time, the instrumental variables $y_I^{(1)}(k)$ ensure that the resulting parameter estimates are consistent but not necessarily very efficient.

$Y_I^{(2)}(k)$ uses the parameter estimates to predict the state vector which is a part of the instrumental variables. If the parameters are close to the true values, the state estimate would be good and the instrumental variables would be efficient. However, if the parameters are far from the true values, the instrumental variables may not even be consistent. Rowe [17] and Pandaya [12] use instruments of the type $y_I^{(2)}(k)$ for certain ARMA models. Pandaya has observed the divergence problems and uses an adaptive filter to reduce the problem. However, the convergence of his algorithm cannot be ensured.

It is desirable to incorporate the best properties of $y_I^{(1)}(k)$ and $y_I^{(2)}(k)$ in the selection of the instrumental variables. This is done by introducing the Insured Instrumental Variables (IIV). Two choices of IIV are:

1. A linear combination of choices (a) and (b)

$$y_I(k) = \alpha y_I^{(1)}(k) + (1 - \alpha)\, y_I^{(2)}(k)$$

$$0 < \alpha \leq 1 \qquad (3.21)$$

2. After every r data points, set

$$\hat{x}(k) = y(k-1) \qquad 1 \leq r < \infty \qquad (3.22)$$

and in between propagate the state equations to use $y_I^{(2)}(k)$ as the instruments. This way, even if the filter were diverging, the state estimate may not diverge in r sample periods.

Both of the above choices are called insured instrument variables because in each case, a penalty is paid to ensure convergence. By choosing α greater than zero in Eq. (3.21) and r finite in Eq. (3.22), the estimates are not as efficient as they could be if the parameters are close to the true values. However, they are not as bad as they would be if the parameter values used to predict the state vector are very wrong. α could be chosen as one to start with and decreased as convergence occurs. Similarly, starting from value of one, r could be increased as true parameter values are obtained.

Once the instrumental variables vector is chosen, the problem is to use it for determining parameter estimates. If there are $N+1$ observations $z(1), z(2), \ldots, z(N+1)$, we can write Eq. (3.17) compactly as

$$Z_N = \Theta Y_N + \eta_N \qquad (3.23)$$

where $$Z_N \triangleq \{z(2),\ z(3),\ \ldots,\ z(N+1)\} \qquad (3.24)$$

and $$Y_N \triangleq \{y(1),\ y(2),\ \ldots,\ y(N)\} \quad \text{etc.} \qquad (3.25)$$

Let $\hat{Y}_N$ be the instrumental matrix defined as

$$\hat{Y}_N \triangleq [y_I(1), y_I(2), \ldots, y_I(N)] \tag{3.26}$$

Postmultiply (3.23) by $\hat{Y}_N^T$

$$Z_N \hat{Y}_N^T = \Theta Y_N \hat{Y}_N^T + \eta_N \hat{Y}_N^T \tag{3.27}$$

An estimate of Θ based on $N+1$ observations is,

$$\hat{\Theta}_N = (Z_N \hat{Y}_N^T)\ (Y_N \hat{Y}_N^T)^{-1} \tag{3.28}$$

This can be converted into a recursive algorithm

$$\hat{\Theta}_N = \hat{\Theta}_{N-1} + \frac{\{z(N+1) - \hat{\Theta}_{N-1} y(N)\}\ \hat{y}^T(N) V_{N-1}}{(1 + \hat{y}^T(N) V_{N-1} y(N))} \tag{3.29}$$

and

$$V_N = V_{N-1} - \frac{V_{N-1} y(N) \hat{y}^T(N) V_{N-1}}{1 + \hat{y}^T(N) V_{N-1} y(N)} \tag{3.30}$$

where

$$V_N = (Y_N\ \hat{Y}_N^T)^{-1} \tag{3.31}$$

In many cases, the data rate is fast and the parameter estimates are not required after every data point. In that case, it is computationally inefficient to use Eqs. (3.29) and (3.30). Suppose the parameters have to be updated after every k measurements. Equation (3.28) can be written as

$$\hat{\Theta}_{N+k}\ (Y_{N+k}\ \hat{Y}_{N+k}^T) = Z_{N+k}\ \hat{Y}_{N+k}^T \tag{3.32}$$

which gives

$$(\hat{\Theta}_{N+k} - \hat{\Theta}_N)\, Y_{N+k}\, \hat{Y}^T_{N+k} = \sum_{i=N+1}^{N+k} z(i+1)\, \hat{y}^T(i) - \hat{\Theta}_N \sum_{i=N+1}^{N+k} y(i)\hat{y}^T(i) \tag{3.33}$$

Depending on k, various formulae can be used

$$\hat{\Theta}_{N+k} = \hat{\Theta}_N + \left\{ \sum_{i=N+1}^{N+k} z(i+1)\hat{y}^T(i) - \Theta_N \sum_{i=N+1}^{N+k} y(i)\hat{y}^T(i) \right\} V_{N+k}$$

for large k

$$= \hat{\Theta}_N + \sum_{i=N+1}^{N+k} \{z(i+1) - \Theta_N y(i)\}\, \hat{y}^T(i)\, V_{N+k} \tag{3.34}$$

for small k

V_{N+k} can be obtained directly by inversion or Eq. (3.30) can be used to updating. Again, for large k, it is computationally more efficient to perform a direct inversion, while for small k, Eq. (3.30) should be used. These trade offs can be determined by counting the number of computations required per parameter update. The best method depends on the problem at hand.

Many times the parameters of the system are changing slowly. To be able to track time varying parameters, it is necessary to "phase out" the estimate based on old measurements. One method to do this is to use an exponentially fading filter on past measurements leading to the following equations.

$$\Theta_N = \Theta_{N-1} + \frac{\{z(N+1) - \Theta_{N-1} y(N)\}\, \hat{y}^T(N) V_{N-1}}{(\rho + \hat{y}^T(N)\, V_{N-1}\, y(N))} \tag{3.35}$$

and

$$V_N = \frac{1}{\rho}\left[V_{N-1} - \frac{V_{N-1}\, y(N)\, \hat{y}^T(N)\, V_{N-1}}{(\rho + \hat{y}^T(N)\, V_{N-1} y(N))}\right] \qquad 0 < \rho < 1 \qquad (3.36)$$

3.2.3 Instrumental Variables Approach When the Observability Index is Greater Than One (Rank (H) < n)

In many identification problems, the measurements of all state variables are not available. In other words, it is not easy to construct state estimates at a point from the measurements at that point. It is well known that in this case, even if H is known, it is not possible to obtain estimates of all parameters in ϕ and G [18]. Also, it is not easy to determine which set of parameters in (ϕ, G) are identifiable. We will assume a certain parameter set in (ϕ, G) is identifiable.

Consider the case where the system is completely controllable from each input and the p measurements y are the noisy measurements of the last p states. Let ϕ_{ij} and G_{ij} $(i \geq n-p, \text{all } j)$ be the set of unknown parameters. It is known that this set of parameters is identifiable from the measurements. Other parameters in ϕ and G are assumed known. The Type II canonical form of Denham [18] is a special case of this form and the Type I canonical form can be transformed in this form. Substituting the measurements in the equations of motion, we have

$$\begin{bmatrix} x_1(k+1) \\ \hline Z(k+1) \end{bmatrix} = \left[\begin{array}{c|c} \phi_{11} & \phi_{12} \\ \hline \phi_{21} & \phi_{22} \end{array}\right] \begin{bmatrix} x_1(k) \\ \hline Z(k) \end{bmatrix} + \begin{bmatrix} G_1 \\ \hline G_2 \end{bmatrix} u(k) + \begin{bmatrix} \Gamma_1 \\ \hline \Gamma_2 \end{bmatrix} w(k)$$

$$- \begin{bmatrix} \phi_{12} \\ \hline \phi_{22} \end{bmatrix} v(k) + \begin{bmatrix} 0 \\ \hline I \end{bmatrix} v(k+1) \qquad (3.37)$$

Defining,

$$\Theta = [\phi_{21} \mid \phi_{22} \mid G_2]$$

$$y^T(k) = [x_1^T(k),\ z^T(k),\ u^T(k)] \tag{3.38}$$

$$\eta_2(k) = \Gamma_2 w(k) - \phi_{22} v(k) + v(k+1)$$

we get

$$z(k+1) = \Theta y(k) + \eta_2(k) \tag{3.39}$$

Similarly, the first n-p equations in (3.37) can be written as

$$x_1(k+1) = \phi_{11}\, x_1(k) + (\phi_{12} \quad G_1) \begin{bmatrix} z(k) \\ \hline u(k) \end{bmatrix} + \Gamma_1 w(k) - \phi_{12}\, v(k) \tag{3.40}$$

For the purpose of identification, Eq. (3.39) resembles Eq. (3.17) except that $x_1(k)$ is not known exactly. The instrumental variables must be selected based on the same considerations as in the last section. Since $x_1(k)$ is not known exactly and there is no measurement of these states, choices (a) and (b) defined by Eqs. (3.18) and (3.19) will have to be modified somewhat. Therefore $x_1(k)$ is determined by a direct integration of Eq. (3.40) after the noise terms are dropped. Since the parameters in this equation are exactly known, the estimate of $x_1(k)$ is noisy but will not diverge if ϕ_{11} is stable. In choice (b) for the instrumental variables, $x_1(k)$ and $x_2(k)$ are determined from the measurements in the best possible way. After the instrumental variables are chosen, the recursive procedures of Section 3.2.2 can be used.

3.2.4 State Estimation When the Parameters Are Not Known Exactly

The main problem in using the instrumental variables of Eq. (3.19) is the determination of efficient state estimates from the measurements when the system parameters are not known exactly. In other words, we need an algorithm to find a way of determining a good estimate $\hat{x}(k)$ given $z(1)$, $z(2)$, ..., $z(k-1)$, recursively.

If the parameters of the system are known, the best estimate $x(k)$ can be determined using the Kalman filter, i.e.,

$$\hat{x}(k+1) = \phi\hat{x}(k) + Gu(k) + \phi K\nu(k) \tag{3.40}$$

and

$$\nu(k) = z(k) - \hat{x}(k) \tag{3.41}$$

In an actual implementation, the best estimates of Φ and G must be used in Eqs. (3.40) and (3.41) since the true values are not known. This algorithm is good when the parameters are close to the true values. If this is not so, a major contribution in the innovation $\nu(k)$ is due to the incorrect state updates caused by erroneous ϕ and G in Eq. (3.40). This is particularly true when the process noise is small compared to the measurement noise (i.e., the Kalman filter gains K are small). If Eqs. (3.40) and (3.41) are used since the start of the algorithm, the estimates may diverge initially, giving inconsistent estimates. One way to handle this problem is to account for the incorrect parameter values by increasing the covariance of the process noise. In addition, it may be necessary to update the Kalman gain as the parameter estimates improve. An adaptive filter suggested by Pandaya [12] can also be used. It is important that this filter be good, because the efficiency of the instrumental variables depends on the correlation between $y(k)$ and $\hat{y}(k)$.

3.2.5 Parameter Estimation Error and Identifiability

In Sections 3.2.2 and 3.2.3, we discussed the set of parameters which are structurally identifiable. However, the input may be such that it does not excite all the modes of the system. It is clear, then, that it will be impossible to get good estimates of some of the system parameters. To determine this on-line, we need to know the accuracy with which the parameters are estimated. From Eqs. (3.27) and (3.28), it is clear that

$$(\hat{\Theta}_N - \Theta)\, Y_N\, \hat{Y}_N^T = \eta_N\, \hat{Y}_N^T \tag{3.42}$$

$$\hat{\Theta}_N - \Theta = \eta_N\, \hat{Y}_N^T\, (Y_N\, \hat{Y}_N^T)^{-1} \tag{3.43}$$

$\eta(k)$ is not correlated with $\hat{y}(k)$ but is correlated with $y(k)$. However, it is easy to show that

$$\operatorname*{plim}_{N\to\infty}\ \eta_N\, \hat{Y}_N^T\, (Y_N\, \hat{Y}_N^T)^{-1} = 0 \tag{3.44}$$

if the conditions regarding the efficient instrumental variables are satisfied. The covariance of the estimation error is difficult to determine. It can be obtained by neglecting the correlation between η_N, Y_N and $\hat{Y}_N$, approximately, as

$$D_i = (Y_N\, \hat{Y}_N^T)^{-1}\, \hat{Y}_N\, \hat{Y}_N^T\, (Y_N\, \hat{Y}_N^T)^{-1}\ \{Q_{ii} + R_{ii} - (\phi R \phi^T)_{ii}\} \tag{3.45}$$

where D_i is the covariance of the estimation error of parameters in the ith row of Θ. Using the approximation

$$\hat{Y}_N\, \hat{Y}_N^T \cong Y_N\, \hat{Y}_N^T \tag{3.46}$$

we get

$$D_i = (Y_N \hat{Y}_N^T)^{-1} \{Q_{ii} + R_{ii} - (\phi R \phi^T)_{ii}\} \tag{3.47}$$

The approximation of Eq. (3.46) is as good as the estimates of the state variables used as instruments. As the number of data points increase and the estimates of parameters improve, Eq. (3.46) will be a better approximation. If Eq. (3.46) holds, V_N is a good measure of the parameter estimation errors. The diagonal terms of this matrix, when multiplied by the second term on the right hand side of Eq. (3.47), provide covariance of estimation error on each parameter.

3.2.6 Application to Aircraft Parameter Identification

Decoupled longitudinal and lateral aircraft equations of motion are given by Etkin [19]. In the aircraft equations, many terms in the longitudinal and lateral equations are unity, zero or otherwise known. This fact can be used to advantage in the instrumental variable approach.

The state equations for the lateral motions of an aircraft are

$$\frac{d}{dt}\begin{bmatrix} p \\ r \\ \beta \\ \phi \end{bmatrix} = \begin{bmatrix} L_p & L_r & L_\beta & 0 \\ N_p & N_r & N_\beta & 0 \\ \sin\alpha & -\cos\alpha & Y_\beta & \frac{g}{V} \\ 1 & \tan\theta & 0 & 0 \end{bmatrix}\begin{bmatrix} p \\ r \\ \beta \\ \phi \end{bmatrix} + \begin{bmatrix} L_{\delta_a} & L_{\delta_r} \\ N_{\delta_a} & N_{\delta_r} \\ Y_{\delta_a} & Y_{\delta_r} \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \delta_a \\ \delta_r \end{bmatrix} \tag{3.48}$$

Suppose there are noisy measurements of p, r, β and ϕ. Let the sampling interval be Δ. The discrete time equivalent of the above system is, for small Δ,

$$\begin{bmatrix} p(k+1) \\ r(k+1) \\ \beta(k+1) \\ \phi(k+1) \end{bmatrix} = \begin{bmatrix} 1+L_p\Delta & L_r\Delta & L_\beta\Delta & 0 \\ N_p\Delta & 1+N_r\Delta & N_\beta\Delta & 0 \\ (\sin\alpha)\Delta & -(\cos\alpha)\Delta & 1+Y_\beta\Delta & \frac{g}{V}\Delta \\ \Delta & (\tan\theta)\Delta & 0 & 1 \end{bmatrix} \begin{bmatrix} p(k) \\ r(k) \\ \beta(k) \\ \phi(k) \end{bmatrix} + \begin{bmatrix} L_{\delta_a}\Delta & L_{\delta_r}\Delta \\ N_{\delta_a}\Delta & N_{\delta_r}\Delta \\ Y_{\delta_a}\Delta & Y_{\delta_r}\Delta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \delta_a \\ \delta_r \end{bmatrix} \tag{3.49}$$

Since the last column of ϕ and last row of ϕ and G do not contain any unknown parameters, the above equations can be simplified

$$\begin{bmatrix} p(k+1) \\ r(k+1) \\ \beta(k+1) - \frac{g}{V}\Delta\ \phi(k) \end{bmatrix} = \begin{bmatrix} 1+L_p\Delta & L_r\Delta & L_\beta\Delta \\ N_p\Delta & 1+N_r\Delta & N_\beta\Delta \\ \sin\alpha\Delta & -\cos\alpha\Delta & 1+Y_\beta\Delta \end{bmatrix} \begin{bmatrix} p(k) \\ r(k) \\ \beta(k) \end{bmatrix}$$

$$+\Delta \begin{bmatrix} L_{\delta_a} & L_{\delta_r} \\ N_{\delta_a} & N_{\delta_r} \\ Y_{\delta_a} & Y_{\delta_r} \end{bmatrix} \begin{bmatrix} \delta_a \\ \delta_r \end{bmatrix} \tag{3.50}$$

Equation (3.50) can be used for identification, while Eq. (3.49) is used for prediction and, hence, determining the instruments.

The four states (angle-of-attack α, forward speed u, pitch rate q, and pitch angle θ), which describe the longitudinal motions of an aircraft in discrete form are

$$\begin{bmatrix} \alpha(k+1) \\ u(k+1) \\ q(k+1) \\ \theta(k+1) \end{bmatrix} = \begin{bmatrix} 1+Z_\alpha\Delta & Z_u\Delta & Z_q\Delta & 0 \\ X_\alpha\Delta & 1+X_u\Delta & X_q\Delta & -g\Delta \\ M_\alpha\Delta & M_u\Delta & 1+M_q\Delta & 0 \\ 0 & 0 & \Delta & 1 \end{bmatrix} \begin{bmatrix} \alpha(k) \\ u(k) \\ q(k) \\ \theta(k) \end{bmatrix} + \Delta \begin{bmatrix} Z_{\delta_e} \\ X_{\delta_e} \\ M_{\delta_e} \\ 0 \end{bmatrix} \quad \begin{bmatrix} Z_\alpha \\ X_\alpha \\ M_\alpha \\ 0 \end{bmatrix} \alpha_g \tag{3.51}$$

which could be simplified as before

$$\begin{bmatrix} \alpha(k+1) \\ u(k+1)+g\Delta\theta(k) \\ q(k+1) \end{bmatrix} = \begin{bmatrix} 1+Z_\alpha\Delta & Z_u\Delta & Z_q\Delta \\ X_\alpha\Delta & 1+X_u\Delta & X_q\Delta \\ M_\alpha\Delta & M_u\Delta & 1+M_q\Delta \end{bmatrix} \begin{bmatrix} \alpha(k) \\ u(k) \\ q(k) \end{bmatrix} + \Delta \begin{bmatrix} Z_{\delta_e} \\ X_{\delta_e} \\ M_{\delta_e} \end{bmatrix}$$

$$+\Delta \begin{bmatrix} Z_\alpha \\ X_\alpha \\ M_\alpha \end{bmatrix} \alpha_g \tag{3.52}$$

Examples of the application to flight test data of this algorithm are detailed in Chapter IV.

3.3 ON-LINE MAXIMUM LIKELIHOOD METHODS

The likelihood method has been a subject of considerable study in the parameter identification work [13]. The method has been very attractive because the resulting estimates possess desirable statistical properties. It can be shown that the likelihood estimates are asymptotically efficient, unbiased, and consistent. This technique converts the identification problem into an optimization problem of the likelihood function. The likelihood method is basically a batch technique requiring several passes through the data because the likelihood function is nonlinear in parameters. Therefore, it is usually implemented as an off-line technique. The maximum

likelihood method has, however, been recently modified to make it an effective tool for on-line and real time identification purposes. The techniques will try to integrate the requirements of efficiency and convergence and low computation time and storage.

3.3.1 Off-Line Maximum Likelihood Method With Newton-Raphson Optimization

Consider a discrete time system

$$x(k+1) = \phi x(k) + Gu(k) + \Gamma w(k) \tag{3.53}$$

with measurements

$$y(k) = Hx(k) + v(k) \qquad i=1,2,3,\ldots N \tag{3.54}$$

The negative log-likelihood function is

$$J_N = \frac{1}{2} \sum_{k=1}^{N} \{\nu^T(k)\ B^{-1}\ \nu(k) + \log\ |B|\} \tag{3.55}$$

where $\nu(k)$ are the innovations and B is the covariance of the innovations (see Eqs. (3.5) and (3.6)). We assume here that B is a constant. A direct optimization with respect to B gives

$$B = \frac{1}{N} \sum_{k=1}^{N} \nu(k)\ \nu^T(k) \tag{3.56}$$

Substituting B in the negative log-likelihood function

$$J_N = \frac{N}{2} \log |B| + \text{constant} \tag{3.57}$$

The first derivative of the cost function is

$$\frac{\partial J_N}{\partial \theta} = \sum_{k=1}^{N} \frac{\partial \nu^T(k)}{\partial \theta} B^{-1} \nu(k) \tag{3.58}$$

and the second derivative (which is the information matrix) is,

$$M_N = \frac{\partial^2 J_N}{\partial \theta^2} = E\left\{ \frac{\partial J_N}{\partial \theta} \left(\frac{\partial J_N}{\partial \theta} \right)^T \right\}$$

$$\cong \sum_{k=1}^{N} \frac{\partial \nu^T(k)}{\partial \theta} B^{-1} \frac{\partial \nu(k)}{\partial \theta} \tag{3.59}$$

In the Newton-Raphson procedure, the parameter step is determined by solving the linear equations

$$M_N \Delta\theta_N = \frac{\partial J_N}{\partial \theta} \tag{3.60}$$

After updating, the new parameters are now used to recompute the first and the second gradients of the negative log-likelihood function, which determines the next parameter step. This procedure is repeated until the gradient is insignificant, the cost stops decreasing and $\Delta\theta$ becomes small.

Both the first and the second gradients of J_N are functions of $\frac{\partial \nu(k)}{\partial \theta}$. These gradients are computed by differentiating Eqs. (3.5) and (3.6) with respect to each parameter

$$\frac{\partial \hat{x}(k+1)}{\partial \theta_i} = \frac{\partial \phi}{\partial \theta_i} \hat{x}(k) + \phi \frac{\partial \hat{x}(k)}{\partial \theta_i} + \frac{\partial G}{\partial \theta_i} u(k) + \frac{\partial \phi K}{\partial \theta_i} \nu(k)$$

$$+ \phi K \frac{\partial \nu(k)}{\partial \theta_i} \tag{3.61}$$

$$\frac{\partial \nu(k)}{\partial \theta_i} = - \frac{\partial H}{\partial \theta_i} \hat{x}(k) - H \frac{\partial \hat{x}(k)}{\partial \theta_i} - \frac{\partial D}{\partial \theta_i} u(k) \tag{3.62}$$

The gradient of Kalman gain with respect to θ_i is determined from Eq. (3.6). If there are m parameters, the computation of the first and second gradients of the cost functions requires the solution to $(m+1)n$ difference equations. A flow chart of the algorithm is shown in Figure 2.

3.3.2 Modification of the Off-Line Likelihood Method for On-Line and Real Time Applications

There are two basic problems with the off-line maximum likelihood method for on-line and real time applications. These are: (a) too much computation time is required in solving the sensitivity equations, and (b) the entire data record must be stored because several passes are required through the data record. Two major modifications are made to overcome these problems. First, gradient computation must be simplified and secondly, more rapid techniques for parameter update must be implemented. Gradient computations may be simplified by using the requirement that the system is linear. The computation of the first and second gradients of the cost function then does not require the explicit computation of the sensitivities of the innovations with respect to the parameters. This result is shown as follows. The innovation gradient with respect to all parameters can be written in matrix form

$$\nu_\theta(k) = y_\theta(k) - H_\theta x_\theta(k)$$

$$x_\theta(k+1) = \phi_\theta x_\theta(k) + G_\theta u(k) \tag{3.63}$$

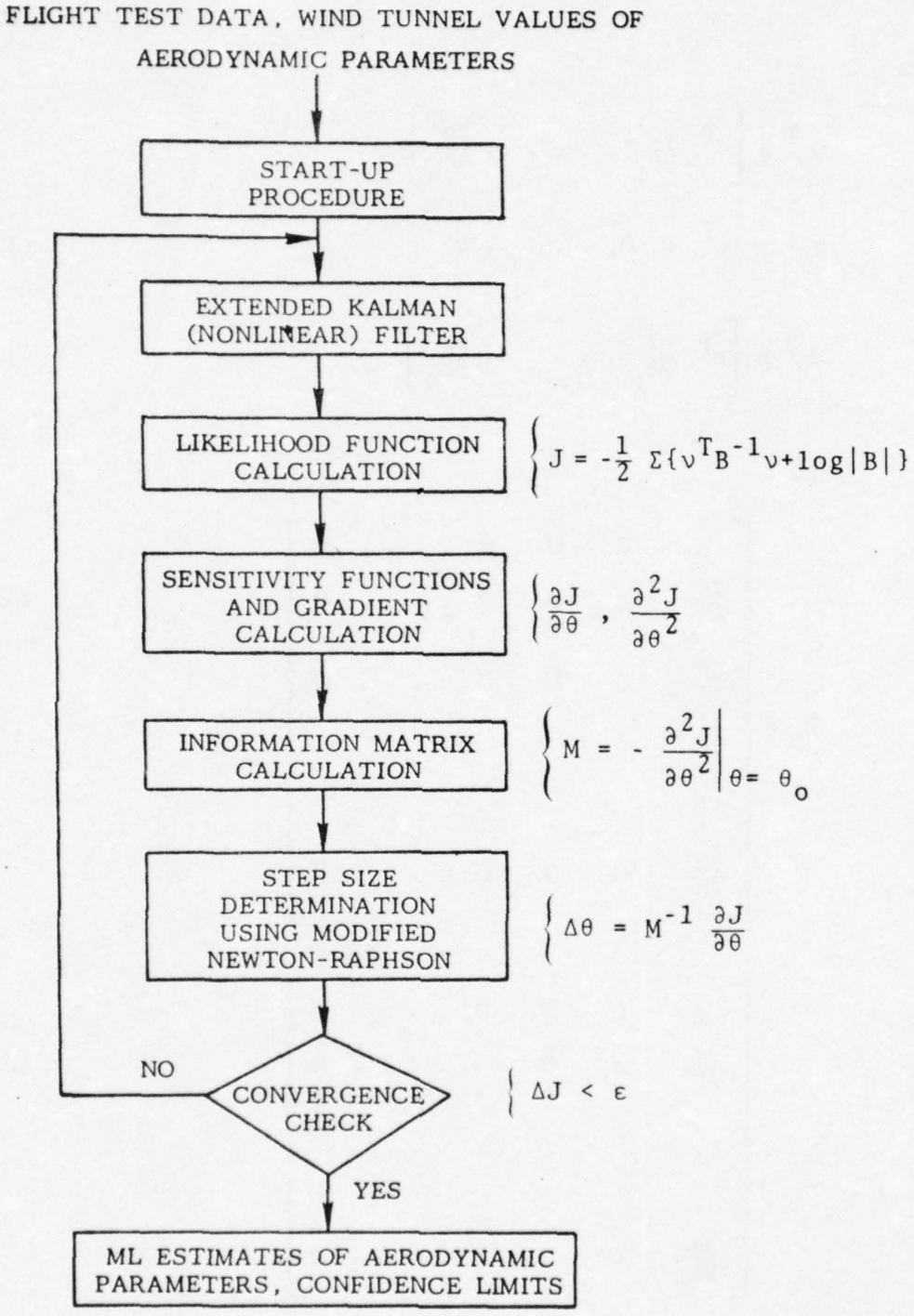


Figure 2 Flow Chart of Maximum Likelihood Identification Program

where

$$\nu_\theta^T \triangleq \left[\nu^T, \frac{\partial \nu^T}{\partial \theta_1}, \ldots, \frac{\partial \nu^T}{\partial \theta_m}\right]$$

$$y_\theta^T \triangleq (y^T, 0, 0, \ldots, 0) \tag{3.65}$$

$$x_\theta^T \triangleq \left[x^T, \frac{\partial x^T}{\partial \theta_1}, \ldots, \frac{\partial x^T}{\partial \theta_m}\right]$$

and

$$H_\theta = \begin{bmatrix} H & 0 & 0 & 0 & \ldots & 0 \\ \frac{\partial H}{\partial \theta_1} & H & 0 & 0 & \ldots & 0 \\ \frac{\partial H}{\partial \theta_2} & 0 & H & 0 & \ldots & 0 \\ \vdots & & & \ddots & & \\ \frac{\partial H}{\partial \theta_m} & 0 & 0 & 0 & \ldots & H \end{bmatrix} \tag{3.66}$$

$$\phi_\theta = \begin{bmatrix} \phi & 0 & 0 & 0 & \ldots & 0 \\ \frac{\partial \phi}{\partial \theta_1} & \phi & 0 & 0 & \ldots & 0 \\ \vdots & & & \ddots & & \\ \frac{\partial \phi}{\partial \theta_m} & 0 & 0 & 0 & \ldots & \phi \end{bmatrix} \tag{3.67}$$

$$G_\theta{}^T = \left[G^T \;\vdots\; \frac{\partial G^T}{\partial \theta_1} \quad \cdots \quad \frac{G^T}{\partial \theta_m} \right] \tag{3.68}$$

Equation (3.63) is the principal reason for high computation time using the maximum likelihood technique. Figure 3 simply illustrates the increase in the number of states plus sensitivity equations as the number of parameters, m, is increased (see Figure 1 to estimate execution time). It is shown in Appendix A, however, that this system of equations can be reduced to a lower order system, specifically

$$x_A(k+1) = \phi_A x_A(k) + G_A\, u(k)$$

and

$$x_\theta(k) = T x_A(k) \tag{3.69}$$

The order r of the system is smaller than or equal to $n(q+1)$. Note that $[\phi_A, G_A]$ *are in controller canonical form.* Substituting $x_\theta(k)$ from Eq. (3.69) into Eq. (3.63), we have

$$\nu_\theta(k) = y_\theta(k) - H\, T\, x_A(k) \triangleq y\,(k) - \overline{T}\, x_A(k) \tag{3.70}$$

The first p elements of ν_θ are the innovations ν, the next p elements are the gradient of the innovations with respect to the first parameter, and so on. Let $\overline{T}_o$ be the first p rows of $\overline{T}$, T_1 the next p rows, and so on. Then

$$\frac{\partial J_N}{\partial \theta_i} = \sum_{k=1}^{N} \nu_i{}^T(k)\; B^{-1}\; \nu(k)$$

$$= \sum_{k=1}^{N} (-\, x_A{}^T(k)\; T_i{}^T)\; B^{-1}\; (y(k) - T_o\, x_A(k)) \tag{3.71}$$

$$= \mathrm{Tr} \left\{ -\sum_{k=1}^{N} y(k)\; x_A{}^T(k)\;\; T_i{}^T B^{-1} + \sum_{k=1}^{N} x_A(k)\; x_A{}^T(k)\;\; T_i^T B^{-1} T_o \right\}$$

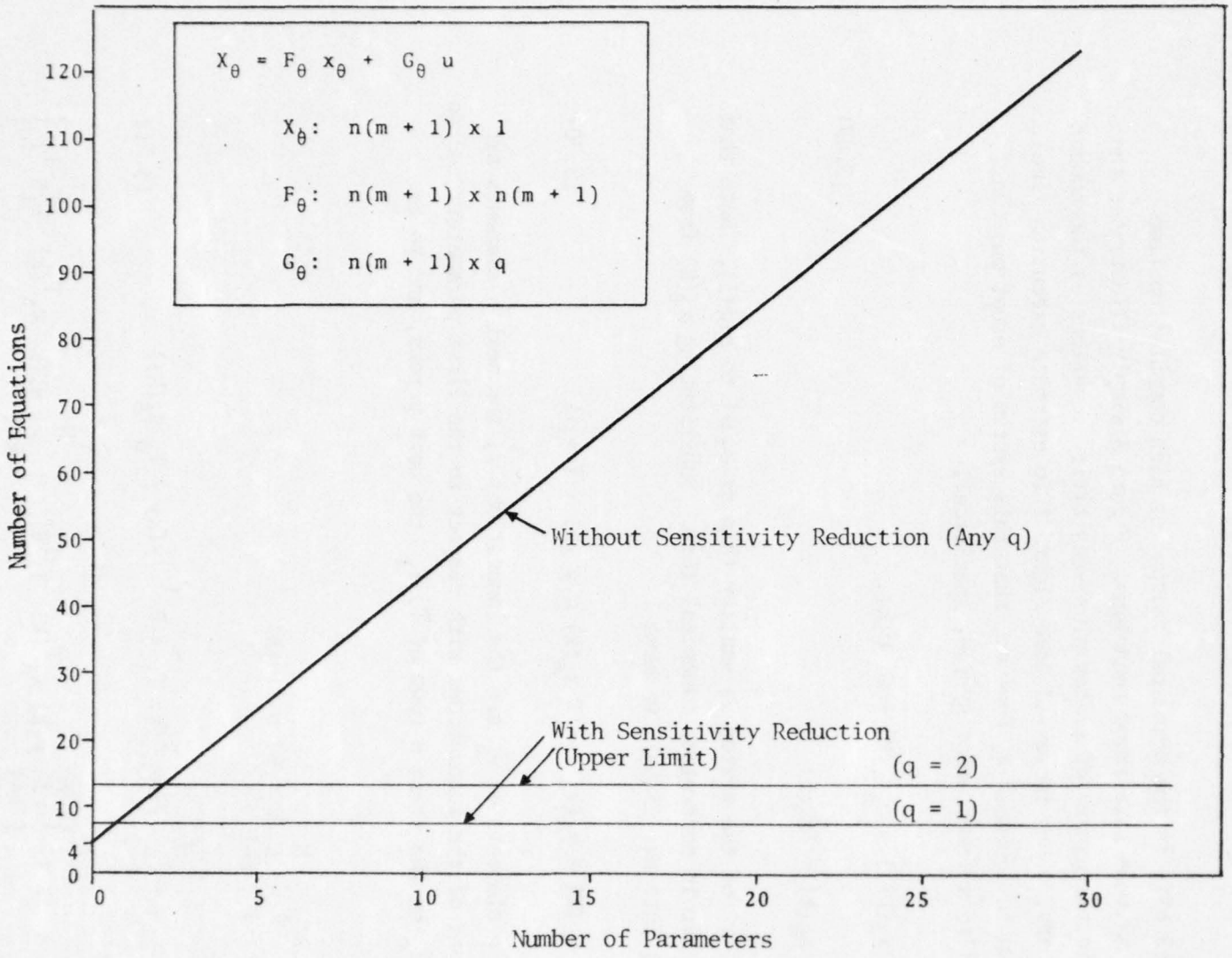


Figure 3 Number of State Plus Sensitivity Equations to be Integrated on Each Iteration of Maximum Likelihood

and

$$\frac{\partial^2 J_N}{\partial\theta_i \partial\theta_j} \cong \sum_{k=1}^{N} \nu_i^T(k)\ B^{-1}\ \nu_j(k)$$

$$= \mathrm{Tr}\left\{ \sum_{k=1}^{N} x_A(k)\ x_A^T(k)\ \ T_i^T\ B^{-1}\ T_j \right\} \qquad (3.72)$$

Thus, the computation of the gradient and the second gradient of the negative log-likelihood function requires the determination of the autocorrelation of x_A and the cross-correlation of x_A and y. This requires the propagation of only the x_A equations.

The estimate of B is

$$B = \frac{1}{N} \sum_{k=1}^{N} \nu(k)\ \nu^T(k)$$

$$= \frac{1}{N}\,\mathrm{Tr}\left\{ \sum_{k=1}^{N} y(k)\ y^T(k) + \sum_{k=1}^{N} x_A(k)\ x_A^T(k)\ \ T_o^T\ B^{-1}\ T_o \right.$$

$$\left. - 2 \sum y(k)\ x_A^T(k)\ \ T_o^T\ B^{-1} \right\} \qquad (3.73)$$

Note that the various matrices involved in the computation of the cost and the first and second gradients of the cost are submatrices of T^* defined as

$$T^* = \begin{bmatrix} I \\ \hline T_1^T \\ \hline \vdots \\ T_m^T \end{bmatrix} B^{-1} \left[I \,\vdots\, T_1 \,\vdots\, T_2 \,\vdots\, \cdots \,\vdots\, T_m \right] \qquad (3.74)$$

T^* can be computed and stored before the on-line identification is begun.

The second modification is related to the first and consists of approximations to the gradients. In general, the derivatives of J_N computed using the a priori values of the derivatives are in error since these values are usually not correct. As mentioned before, the maximum likelihood optimization will seldom converge in one Newton-Raphson step, requiring computation of ϕ_A, G_A and T for parameter values at every step.

Three techniques to reduce this problem are as follows.

1. It may be possible to use the same sensitivity equations even after the parameters are updated until the parameter step is large enough to cause appreciable error in the computation of the gradients. In other words, the same sensitivity equations are not updated nearly as often as the parameter step. This method is simple but is applicable only when the initial parameters are fairly good or the information about the parameters is obtained at a low rate (i.e., low system excitation or high noise-to-signal ratio). In either of these cases, the parameter step will be small over a short period of time.

2. The gradients of ϕ_A, G_A and T with respect to the parameters can be computed and stored. Then ϕ_A, G_A and T can be modified more often than in the last case, thus

$$\phi_A(\theta) = \phi_A(\theta') + \sum_{i=1}^{m} \frac{\partial \phi_A}{\partial \theta_i} (\theta_i - \theta_i')$$

$$G_A(\theta) = G_A(\theta') + \sum_{i=1}^{m} \frac{\partial G_A}{\partial \theta_i} (\theta_i - \theta_i') \qquad (3.75)$$

etc.

Since ϕ_A and G_A are in controller canonical form, only the canonical variables need to be updated. The sensitivities of only these canonical variables need to be precomputed and stored. Notice that this does not require much computation in real time.

3. The state x_A for different parameter values can be computed directly

$$x_A(k,\theta) = x_A(k,\theta') + \sum_{i=1}^{m} \frac{\partial x_A}{\partial \theta_i} (\theta_i - \theta_i') \tag{3.76}$$

$\frac{\partial x_A}{\partial \theta_i}$ is computed by writing the sensitivity equations for the system of equations governing $x_A(k)$. The sensitivity reduction is again used on this system to minimize the number of equations to be solved. Then, the $x_A(k,\theta)$ can be computed directly once θ is known.

The preceding techniques are useful for reducing the computer requirements of an off-line algorithm. Results for flight test data are given in Chapter IV.

3.3.3 Recursive Equations for Real Time Application

The determination of the parameter step size requires the solution of m simultaneous linear equations. It would be computationally infeasible to solve these equations at every time step for a real time algorithm. There are two methods to obtain the parameter step with a reasonable computation time: (a) use recursive equations to update the information matrix and the parameter step without explicitly computing the inverse of the information matrix, or (b) use some approximation technique.

The use of recursive equations is based on the solution for $\Delta\theta_N$ written as

$$\Delta\theta_N = M_N^{-1} \frac{\partial J_N}{\partial \theta} \tag{3.77}$$

The recursive equations to update M_N^{-1} and $\Delta\theta_N$ can be written as

$$M_{N+1}^{-1} = M_N^{-1} - M_N^{-1} \frac{\partial \nu^T(N+1)}{\partial \theta} \left\{ B + \frac{\partial \nu(N+1)}{\partial \theta} M_N^{-1} \frac{\partial \nu^T(N+1)}{\partial \theta} \right\}^{-1} \frac{\partial \nu(N+1)}{\partial \theta} M_N^{-1} \tag{3.78}$$

$$\Delta\theta_{N+1} = \Delta\theta_N + M_N^{-1} \frac{\partial \nu^T(N+1)}{\partial \theta} \left\{ B + \frac{\partial \nu(N+1)}{\partial \theta} M_N^{-1} \frac{\partial \nu^T(N+1)}{\partial \theta} \right\}^{-1} \left\{ \nu(N+1) - \frac{\partial \nu^T}{\partial \theta} \Delta\theta_N \right\} \tag{3.79}$$

These equations are useful only if the number of measurements is small. Otherwise, the computation time is too large.

Approximation methods are superior if the number of measurements is large and the parameter values are not required at every measurement point. Let the parameter updates be required every k data points, then

$$M_{N+k} \Delta\theta_{N+k} = \frac{\partial J_{N+k}}{\partial \theta} \tag{3.80}$$

$$M_{N+k} (\Delta\theta_{N+k} - \Delta\theta_N) = \frac{\partial J_{N+k}}{\partial \theta} - \frac{\partial J_N}{\partial \theta} - (M_{N+k} - M_N) \Delta\theta_N \tag{3.81}$$

$$M_{N+k} (\Delta\theta_{N+k} - \Delta\theta_N) = \sum_{i=N+1}^{N+k} \frac{\partial \nu^T(i)}{\partial \theta} B^{-1} \left\{ \nu(i) - \frac{\partial \nu(i)}{\partial \theta} \Delta\theta_N \right\} \tag{3.82}$$

The second term on the right hand side represents the "true" innovations when the parameter value is $\theta_o + \Delta\theta_N$. The sum on the right hand side of the equation can be computed recursively for any k. M_{N+k} can also be computed recursively. To simplify the solution of the above equation, it is assumed that the process is in progress for a long time. Then

$$M_{N+k} \simeq M_N \tag{3.83}$$

Therefore, the inverse of the information matrix does not have to be computed at every point the parameters are updated. Note that the sum on the right hand side of Eq. (3.82) can be computed using the sensitivity functions reduction.

3.3.4 An On-Line Maximum Likelihood Method for Nonlinear Systems

In many applications, the computation time and computer storage are at a very high premium while the accuracy of the estimates and the time required to get accurate estimates is a secondary consideration. This may be particularly true where the system is defined by a nonlinear system of equations then the instrumental variable methods are not applicable and it is not possible to use reduced sensitivity equations. Then the likelihood method can be used without computing the gradients.

Consider the system following the equations of motion

$$x(k+1) = \phi(x(k), u(k), \theta) + w(k) \tag{3.84}$$

with measurements of nonlinear combinations of states and inputs

$$z(k) = h(x(k), u(k), \theta) + v(k) \tag{3.85}$$

where w(k) and v(k) are white noise sources with covariances Q and R, respectively, and cross-covariance zero and θ are the vectors of unknown parameters.

The negative log-likelihood function can be written in terms of the parameters

$$J(\theta) = \frac{1}{2} \sum_{k=1}^{N} (\nu^T(k)\, B^{-1}\, \nu(k) + \log |B|) \tag{3.86}$$

The problem is to find a θ which minimizes $J(\theta)$. Instead of using a gradient method, a direct search technique will be used. See Powell [20] and Stewart III [21].

The underlying principle of all these direct search methods is basically the same. A certain value is chosen for θ and the likelihood function is computed for this value of θ. The likelihood function is computed for another value of θ. These two likelihood functions are compared and are used to determine a value of the parameter vector at which it would be most useful to compute the likelihood function. This process is repeated until a satisfactory convergence results.

The on-line identification requirements make a direct application of this optimization technique infeasible since it would require storing the entire data and propagating the equations of motion through this data again and again for different parameter values. One method which seems very promising is to determine the likelihood function for any parameter vector using only s points. In short, the likelihood function is determined for different parameter vectors using different sections of the data. This would also obviate the problem of storing the data.

Whittle [22] has shown that the limiting distribution of twice the difference of the negative log-likelihood functions for any parameter value to its expectation for true parameter values can be approximated by a chi-square distribution with one degree-of-freedom. Thus, the smaller the s, the higher the variance of the likelihood function estimate compared to its value. In the beginning, therefore, s can be chosen small since we are looking for

gross changes in the value of the likelihood function. However, as convergence is approached, s should be increased so that the difference in the value of the likelihood function is not lost in the estimation errors. Large s would also reduce the errors caused in the comparison of the likelihood functions based on different input responses.

Several important points are in order here. If more computer time is available, the likelihood function can be evaluated for several parameter vectors simultaneously. This would reduce the amount of data required to get accurate parameter estimates. The optimization routine will have to be modified so that it selects several points in the parameter space where it is most useful to determine the value of the likelihood function. This feature makes the method very attractive for real time application because the identification routine can adapt to the available computer time. If the computation load increases, the technique could start evaluating likelihood functions for one or even no parameter value.

This technique, essentially, treats each parameter value as a separate model and chooses the model for which the negative log-likelihood function is minimum. With some modification, it can be applied when the parameter space is not compact, in particular, for two models with different dimension and parameterization.

3.4 COMPARISON OF THE INSTRUMENTAL VARIABLES METHOD AND THE ON-LINE MAXIMUM LIKELIHOOD APPROACH

In Chapter II, it was shown that under certain circumstances, the least squares and the maximum likelihood methods belong to the same general class. The same relationship holds between the instrumental variables method (least squares or equation error type method), and the on-line maximum likelihood method. The instrumental variables method differs from the least squares in that the instrumental variables are chosen such that the resulting estimates are biased. The on-line maximum likelihood method is a simplified form of the off-line maximum likelihood method involving no storage of past input/output data and simplified computation of the first and the second gradients of the negative log-likelihood function.

From the application viewpoint, there are several differences in the two approaches. These differences, which dictate where either of these techniques should be used, are summarized here:

1. The on-line maximum likelihood method is more general and statistically more efficient than the instrumental variables approach.

2. The convergence of the instrumental variables approach of Section 3.2 is assured, while the maximum likelihood method does not have guaranteed convergence. In other words, the instrumental variables method does not require good starting values of parameters, unlike the on-line maximum likelihood method.

3. The on-line maximum likelihood method requires much more computer time and storage than the instrumental variables approach.

4. The maximum likelihood method can be applied in a variety of circumstances as long as the parameters are identifiable. Moreover, the maximum likelihood method is the only method which can be used with nonlinear systems.

5. The single step of the Newton-Raphson procedure in maximizing the negative log-likelihood function may be inadequate if the parameters are far from the true values.

3.5 SUMMARY

This chapter presented two techniques for on-line identification of parameters of state variable models. The first method, the instrumental variables approach (IVA), can be used with linear systems, is extremely quick and has guaranteed convergence irrespective of starting parameter values and changes in parameter values. The second method uses the maximum likelihood approach

and can be applied to both linear and nonlinear systems. For linear systems, the computation time is reduced considerably by the sensitivity functions reduction technique. This method also requires good starting values for the parameters.

The instrumental variables and the on-line maximum likelihood approaches are two complementary methods for on-line and real time estimation of parameters in dynamic models.

The next chapter presents some results on the application of these methods to simulation and flight test data.

IV. APPLICATION OF ON-LINE ALGORITHMS TO SIMULATED AND FLIGHT TEST DATA

4.1 INTRODUCTION

In the previous chapter, two on-line and real time parameter identification algorithms were developed. Both of those algorithms have excellent statistical properties. It was also shown that the first method, the instrumental variables approach (IVA) has guaranteed convergence irrespective of the starting parameter values and the measurement noise. The on-line maximum likelihood (OLML) method requires a higher computation time and good starting parameter values but produces estimates which are statistically somewhat superior.

This section details the implementation algorithm for the instrumental variables approach. The computer program based on this algorithm is used both with simulation data (which includes process noise and measurement noise effects) and with actual flight test data for various aircraft. In the simulation data, parameters are changed in the middle of the simulation to determine the speed with which new parameter estimates are obtained. The maximum likelihood method is not implemented in a separate on-line computer program because the characteristics of the likelihood method are well known. A study on the computation time requirements of the On-Line Maximum Likelihood method is presented.

Section 4.2 gives the IVA algorithm, its flow chart and computation and storage requirements. Simulation results are presented in Section 4.3. The parameter estimates from flight test data are shown in Section 4.4. Section 4.5 discusses the implementation aspects of the real time and on-line maximum likelihood.

4.2 THE IVA ALGORITHM

Table 2 shows the initial computations required to start the IVA on-line algorithm and the inputs. The actual algorithm for each new sampling point is given in Table 3. The number of computations and storage locations required are summarized in Table 4.

The three tables are detailed and self-explanatory. Some aspects of the algorithm need further comment.

State and Covariance Update: In the implementation of Table 3, the state vector is estimated by using a simple Kalman filter. It could be modified to include an adaptive Kalman filter or a parameter insensitive filter. These could result in some improvement. Another technique is to update the Riccati gain matrix (state estimation error covariance) based on the identified parameter values and a priori or identified process and measurement noise covariance. The covariance update is not recommended because it could cause problems when the parameter estimates are far from the true values and it also requires considerable additional computation time. It may be feasible to update it at regular intervals.

Choice of α: The α time history is chosen a priori. Usually, α should be chosen close to one in the beginning when the parameter values are not known at all. The α is gradually decreased until it reaches a value close to zero (about .1 is sufficient). This low value should be maintained, thereafter. This ensures good convergence in the beginning when the filter is operating poorly. If parameters of the system change in the steady state operation, the method is guaranteed to converge to new parameter values.

Table 2

Inputs and Initial Computation for IVA

INPUT	SYMBOL	STORAGE
1. Initial Estimate of State	$x(1)$	n
2. Covariance of State Estimate	$P(1)$	n^2
3. Kalman Gain (Optional)	$K(1)$	nm
4. Process Noise Covariance (Optional)	Q	n^2
5. Measurement Noise Covariance	R	m^2
6. State Transformation Matrix	ϕ_1	n^2
7. Control Distribution Matrix	G_1	nq
8. Measurement Distribution Matrix	H	mn
9. $(Y_o\ Y_o^T)^{-1}$	V	$(n+q)^2$
10. Past Fading	ρ	1

Initial Computation:

Assign: $\Theta_1 = [\phi_1 \quad G_1]$
$T = H^T R^{-1}$
$R^{-1} = TH$
$K = KH$ (Optional)
$T = RT$
$z(1) = Tz(1)$

T requires storage of p x n

Table 3

IVA Algorithm

	ASSIGN	+ OR −	X	÷	$\sqrt{\ }$	STORE	COMMENT
I. UPDATE V							
1. $\hat{x}^*(i) = \alpha z(i-1) + (1-\alpha)\, x(i)$		(n+1)	Zn				
2. Set $p_i = \begin{bmatrix} \hat{x}^*(i) \\ u(i) \end{bmatrix}$						n+q	
3. Set $p_i^* = \begin{bmatrix} z(i) \\ u(i) \end{bmatrix}^T$						n+q	
4. $q_i = V_i p_i$		(n+q)(n+q-1)	$(n+q)^2$			n+q	
5. $\beta = \rho + p_i^* q_i$		(n+q+1)	(n+q)			1	
6. $q_i^* = \left(p_i^* V_i\right)^T$		(n+q)(n+q-1)	$(n+q)^2$				Overwrite p_i
7. $q_i = q_i/\beta$				(n+q)			
8. $V_{i+1} = \frac{1}{\rho}(V_i - q_i^* q_i^T)$		$(n+q)^2$	$(n+q)^2$	$(n+q)^2$			
II. UPDATE STATE							
9. $\nu(i) = z(i) - \hat{x}(i)$		n				n	
10. $\hat{x}(i+1) = \phi_i(x(i)+K_i\nu(i)) + G_i u(i)$		n(3n+q-2)	n(2n+q)			n	

Table 3 (Concluded)

IVA Algorithm

	ASSIGN	+ OR −	X	÷	✓	STORE	COMMENT
III. COVARIANCE UPDATE (OPTIONAL)							
11. $P(i+1)=\phi_i(I-K_i)P(i)\phi_i^T + \Gamma_i Q_i \Gamma_i^T$		$3n^3-n^2$	$2\frac{1}{2}n^3+\frac{n^2}{2}$				
12. $R^*(i+1)=P(i+1)+R(i+1)$		$\frac{n(n+1)}{2}$					
13. $K(i+1) = P(i+1)B^{-1}(i+1)$		$\frac{3}{2}n^3-\frac{3}{2}n^2$	$\frac{3}{2}n^3+\frac{3}{2}n^2-n$	$2n-1$	n		
IV. PARAMETER UPDATE							
14. $z(i+1) = Tz(i+1)$		$m(p-1)$	np				
15. $\varepsilon_i = -\Theta_i P_i^{*T}+z(i+1)$		$n^2 + nq + q$	$(n+q)n$			n	
16. $\Theta_{i+1} = \Theta_i + \varepsilon_i q_i^T$		$(n+q)n$	$(n+q)n$				
17. Update ϕ_i and G_i	$(n+q)n$						

Table 4

Number of Computations and Storage for Each Step of the IVA

1.	Assign $(n+q)(n+2)$
2.	Add or Subtract: $8n^2 + 3q^2 + 9nq - 2n + np + 2$
3.	Multiply: $7n^2 + 9nq + 3q^2 + np + 3n + q^2$
4.	Divide: $n^2 + q^2 + 2nq + n+q$
5.	Square Root: None
6.	Storage: $4n^2 + m^2 + q^2 + 3nq + 2nm + pn + 7n + 3q + 2$

n = no. of states, p = no. of measurements, q = no. of inputs

Updating ϕ and G: Sometimes the parameters θ change wildly, particularly in the beginning of a run or when parameters change. The matrices ϕ and G also change rapidly causing problems in determining the instruments. Therefore, it is often preferable not to update ϕ and G after each sample point. This produces a certain "damping" in the parameter identification which is good for convergence.

Choice of the Past Fading Factor, ρ: The past fading factor should be chosen close to one. If it is too close to one, the parameter estimates are more accurate when parameter values are constant but are very sluggish to changing parameter values. If ρ is too far from one, the converse holds. In the fault detection situation (possible change in parameter value), the factor ρ is associated with the problem of false alarm and delayed alarm. In the data processing case, ρ should equal one.

4.3 APPLICATION OF IVA TO SIMULATION DATA

The instrumental variables method has been applied to the DC-8 and F-14 simulation data. The equations of motion are summarized in Section 3.2.6 for both the longitudinal and the lateral cases.

4.3.1 DC-8 Simulated Lateral Motions

The state equations for the lateral motions of a DC-8 are given by Equation (3.48). For the purpose of identification, it is assumed that there are noisy measurements of p, r, β and ϕ. The measurement noise covariance matrix is

$$R = \text{diag}[2.5 \times 10^{-5},\ 2.5 \times 10^{-5},\ 2.5 \times 10^{-5},\ 2.5 \times 10^{-5}] \qquad (4.1)$$

The sampling interval Δ is .05 sec. The true values of the parameters and inputs used in the simulation are given in Table 5. The variation of α with time is shown in Figure 4. The parameters are identified using the instrumental variables method. The parameter values at the end of 2.5 sec., 5 sec., 7.5 sec., and 10 sec. are shown in the table.

Parameters L_p, L_r, L_β, L_{δ_a}, N_p, N_r, N_β, N_{δ_r} and Y_β are close to their true values while parameters L_{δ_r}, N_{δ_a}, Y_{δ_a}, Y_{δ_r} are far from their true values. This is the identifiability problem. The instrumental variables method does not check if any parameter is identifiable or not. It simply produces incorrect estimates for unidentifiable or poorly identifiable parameters.

4.3.2 Short Period Motions of an F-14 Aircraft

The short period motions of an F-14 aircraft in wind gust disturbance are simulated using the following equations of motion for a sampling interval of .05 sec. (units ft., deg., sec.).

$$\begin{bmatrix} \alpha(k+1) \\ q(k+1) \end{bmatrix} = \begin{bmatrix} .9757 & .04837 \\ .04295 & .9774 \end{bmatrix} \begin{bmatrix} \alpha(k) \\ q(k) \end{bmatrix} + \begin{bmatrix} -.0067 \\ -.1547 \end{bmatrix} \delta_e + \begin{bmatrix} -1.748 \times 10^{-4} \\ -2.011 \times 10^{-4} \end{bmatrix} u \qquad (4.2)$$

Table 5

DC-8 Simulation Data (Lateral Mode)

PARAMETER	TRUE VALUE	IVA ON-LINE VALUE AFTER:				
		0 SEC.	2.5 SEC.	5 SEC.	7.5 SEC.	10 SEC.
L_p	-2.029	-12.84	13.96	-2.0	-2.106	-2.106
L_r	.4697	-2.7	-.956	2.32	.511	.441
L_β	-2.092	1.628	-10.98	-2.0	-2.076	-1.88
L_{δ_a}	22.02	-1.83	-1.82	20.0	22.4	22.39
L_{δ_r}	.4807	-.260	1.4	1.77	.426	.258
N_p	-.05828	-7.77	-12.0	.032	-.0376	-.0508
N_r	-.2724	-17.54	-2.16	-1.09	-.228	-.3106
N_β	1.655	1.36	-8.05	1.65	1.632	1.716
N_{δ_a}	-.1709	1.0534	1.052	-1.93	-.348	-.28
N_{δ_r}	-1.3680	-.0071	-.48	-.272	-1.26	-1.38
Y_β	-.1274	-20.19	-2.9	.054	.132	-.1288
Y_{δ_a}	0	.2314	.23	.95	-1.31	-1.274
Y_{δ_r}	.03630	.01140	-.74	-1.08	.196	.232

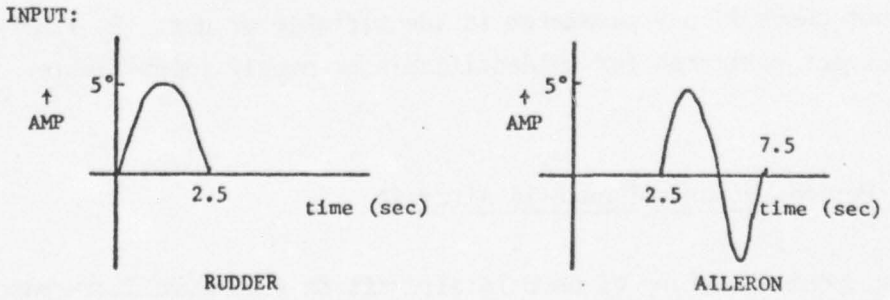


Figure of Rudder and Aileron Inputs for DC-8 Simulation

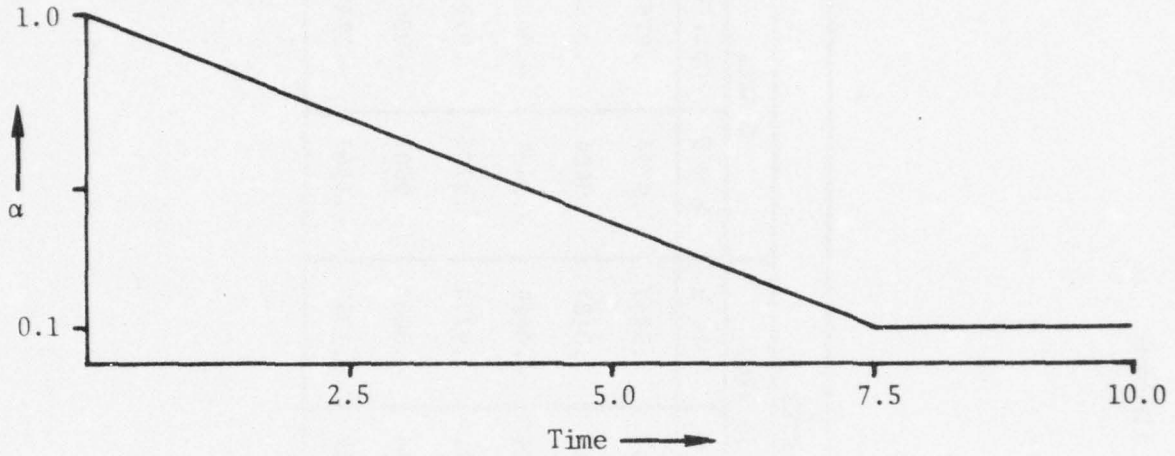


Figure 4 Variation of α With Time

where α is the angle-of-attack, q is the pitch angle, δ_e is the elevator deflection and u is random wind speed. The random wind is assumed to be exponentially correlated with RMS value of 30 ft. $sec.^{-1}$ and a correlation time of 1.1 sec. There are noisy measurements of α and q. The following cases are tried:

(1) Parameters constant, and

$$R = \begin{bmatrix} 2.5 \times 10^{-6} & 0 \\ 0 & 2.5 \times 10^{-6} \end{bmatrix} \tag{4.3}$$

The results for a 40 sec. long experiment with zero K and optimal K are shown in Table 6.

Table 6

IVA For Simulated Short Period Motions of an F-14 Aircraft

$Q_d = 2000.$

$R_{\alpha_d} = 2.5 \times 10^{-6}$

$R_{q_d} = 2.5 \times 10^{-6}$

TRUE VALUE	STARTING VALUE	ESTIMATED VALUES AFTER							
		10 SEC.		20 SEC.		30 SEC.		40 SEC.	
		k = 0	Opt. k	k = 0	Opt. k	k = 0	Opt. k	k = 0	Opt. k
.9757	.9882	.9757	.9757	.9757	.9757	.9757	.9757	.9757	.9757
.0484	.0225	.0484	.0484	.0484	.0484	.0484	.0484	.0484	.0484
-.0430	-.0019	-.0429	-.0429	-.0429	-.0429	-.0429	-.0429	-.0429	-.0429
.9774	.9891	.9774	.9774	.9774	.9774	.9774	.9774	.9774	.9774
-.0067	-.0015	-.0066	-.0066	-.0067	-.0067	-.0066	-.0067	-.0067	-.0067
-.1547	-.0783	-.1547	-.1547	-.1547	-.1547	-.1547	-.1547	-.1547	-.1547

(2) Effect of K: For a 40 sec. long experiment, the measurement noise covariance matrix is taken as

$$R = \begin{bmatrix} .0625 & 0 \\ 0 & .0625 \end{bmatrix}$$

and all the parameters are doubled instantly at 20 sec. Four cases are tried:

(a) K = 0

(b) Optimal K

(c) Nonoptimal K

(d) Updating K at each iteration but delaying it for .50 sec. before use.

The past fading factor is .99 is all these experiments. The results are given in Table 7 and Figure 5. It is clear that using the Kalman filter is a big improvement, while suboptimal Kalman gains do not cause much deterioration.

(3) Effect of ρ: To speed up the convergence when parameters change, it is necessary to reduce ρ. This reduction in ρ increases parameter estimation error in the steady state. The last case with nonoptimal K is tried with $\rho = 0.97$. The results are shown in Table 8 and Figure 6.

4.4 APPLICATION TO FLIGHT TEST DATA

The instrumental variables program has been used extensively with flight test data. To date, lateral flight test data of an advanced Navy variable swept wing fighter and the longitudinal flight data of the T-2B Navy trainer and an advanced Navy fighter have been processed using IVA for the purpose of estimating stability and control coefficients. The program has been very successful in each of these cases. One example is given here to illustrate IVA application on such flight data for the swept wing fighter.

Table 7

IVA For Simulated Short Period Motions of an F-14 Aircraft
(Effect of Prediction Filter)

TRUE VALUE (DISCRETE)	0 SEC	5 SEC	10 SEC	15 SEC	20 SEC	PARAMETERS CHANGE TO	25 SEC	30 SEC	35 SEC	40 SEC
.9757	.9882	.978	.9773	.9771	.977	.9498	.974	.9713	.9608	.96
.0484	.0225	.056	.0486	.0485	.0485	.0945	.0315	.014	-.0417	-.0365
.0430	-.0019	-.0715	-.0459	-.0455	-.0455	-.0839	-.0464	-.0468	-.048	-.0467
.9774	.9891	.9228	.9702	.9695	.9695	.953	.9619	.9579	.9449	.9363
-.0067	-.0015	-.0024	-.0044	-.0045	-.0045	-.0245	-.0739	-.1424	-.375	-.3236
-.1547	-.0783	-.276	-.1700	-.175	-.175	-.304	-.2046	-.2179	-.2645	-.2624
		$k = \text{optimal} = \begin{pmatrix} .0321 & .0365 \\ .0365 & .0577 \end{pmatrix}$								
.9757	.9882	.9746	.9763	.9746	.9753	.9498	.959	.9527	.9504	.9512
.0484	.0225	.0565	.0497	.0494	.0482	.0945	.0609	.087	.0911	.0903
.0430	-.0019	-.0419	-.041	-.044	-.0431	-.0839	-.0678	-.0764	-.0831	-.0834
.9774	.9891	.9773	.9762	.9777	.9776	.953	.9618	.9575	.9604	.9536
-.0067	-.0015	.0024	-.0027	-.0042	-.0073	-.0245	-.0454	-.0225	-.0166	-.0249
-.1547	-.0783	-.1513	-.1605	-.153	-.157	-.304	-.242	-.2734	-.2815	-.303
		$k = \frac{1}{2}$ k optimal								
.9757	.9882	.9746	.9763	.9746	.9753	.9498	.9587	.9524	.9504	.9512
.0484	.0225	.0568	.0498	.0494	.0483	.0945	.0589	.0867	.091	.0904
.0430	-.0019	-.0419	-.0409	-.044	-.0431	-.0839	-.0676	-.0764	-.0829	-.0834
.9774	.9891	.9776	.9761	.9778	.9776	.953	.9607	.9564	.960	.9535
-.0067	-.0015	.0029	-.0026	-.0041	-.0072	-.0245	-.0495	-.0236	-.0169	-.0247
-.1547	-.0783	-.1508	-.1605	-.153	-.1566	-.304	-.2438	-.276	-.2823	-.3037
		k Updated								
.9757	.9882	.9746	.976	.9747	.9753	.9498	.9589	.9521	.9503	.9512
.0484	.0225	.0572	.0498	.0493	.0483	.0945	.0564	.0861	.091	.0904
.0430	-.0019	-.0419	-.0406	-.044	-.0431	-.0839	-.0677	-.0761	-.0828	-.0834
.9774	.9891	.978	.976	.9778	.9776	.953	.9612	.9549	.9598	.953
-.0067	-.0015	.0036	-.0253	-.0044	-.0072	-.0245	-.0540	-.0254	-.0174	-.0246
-.1547	-.0783	-.1506	-.1608	-.1531	-.1566	-.304	-.243	-.278	-.2826	-.304

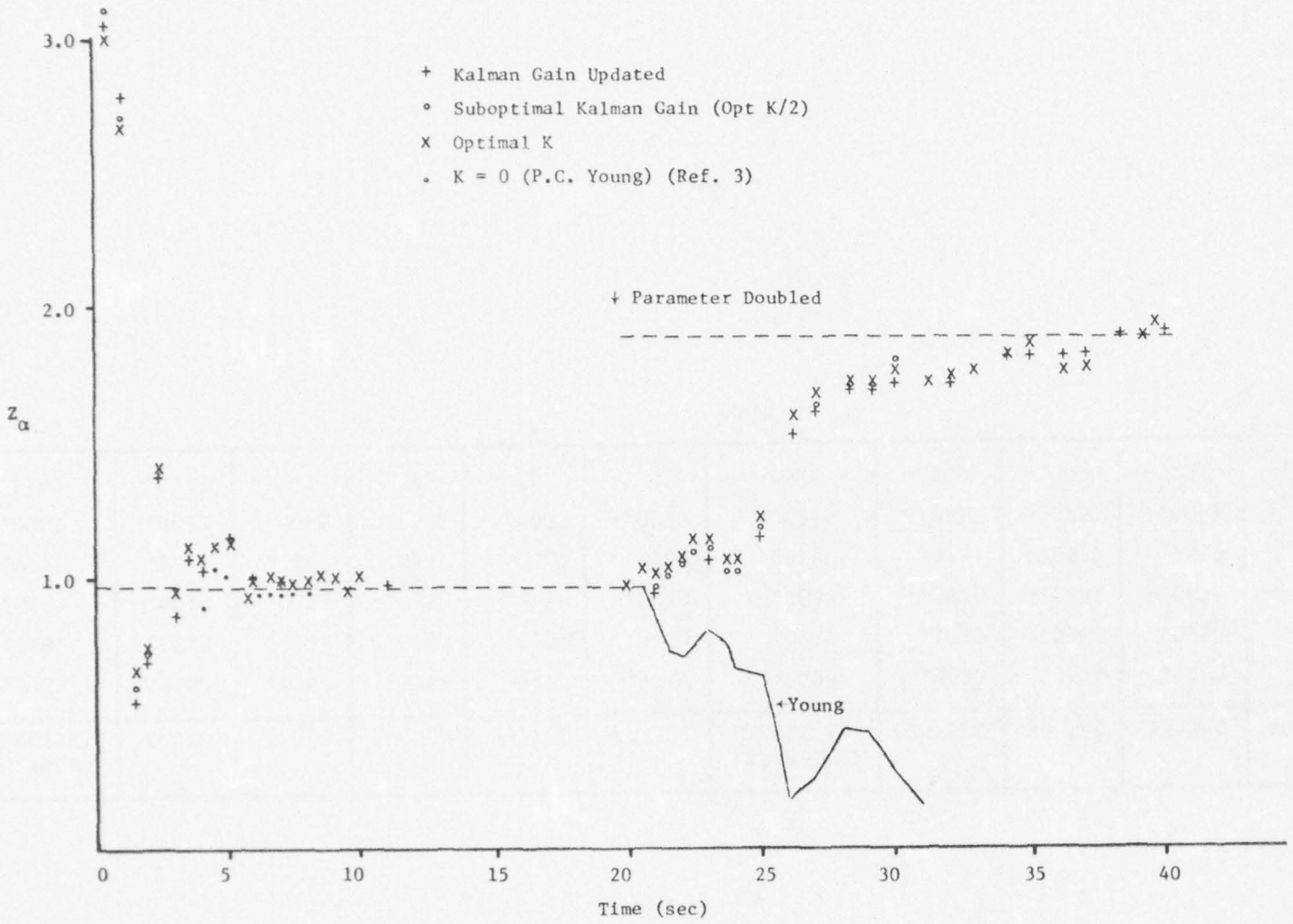


Figure 5 Effect of Kalman Gain K on IVA Estimates

Table 8

Effect of Smaller Fading Factors on Parameter Estimates Using IVA

$\rho = .97$

K Updated

TRUE VALUE (DISCRETE)	0 SEC	5 SEC	10 SEC	15 SEC	20 SEC	PARAMETERS CHANGE TO	25 SEC	30 SEC	35 SEC	40 SEC
.9757	.9882	.9742	.9758	.9723	.9747	.9498	.9433	.9474	.9482	.9509
.0484	.0225	.0608	.0505	.0512	.0473	.0945	.1079	.0983	.0956	.0881
.0430	-.0019	-.0422	-.025	-.0458	-.0429	-.0839	-.0879	-.0841	-.0874	-.0838
.9774	.9891	.981	.965	.9802	.9784	.953	.9788	.9527	.9599	.946
-.0067	-.0015	.0048	.0028	.0035	-.0079	-.0245	-.0115	-.0251	-.0061	-.0349
-.1547	-.0785	-.147	-.196	-.145	-.1586	-.304	-.269	-.3094	-.295	-.3272

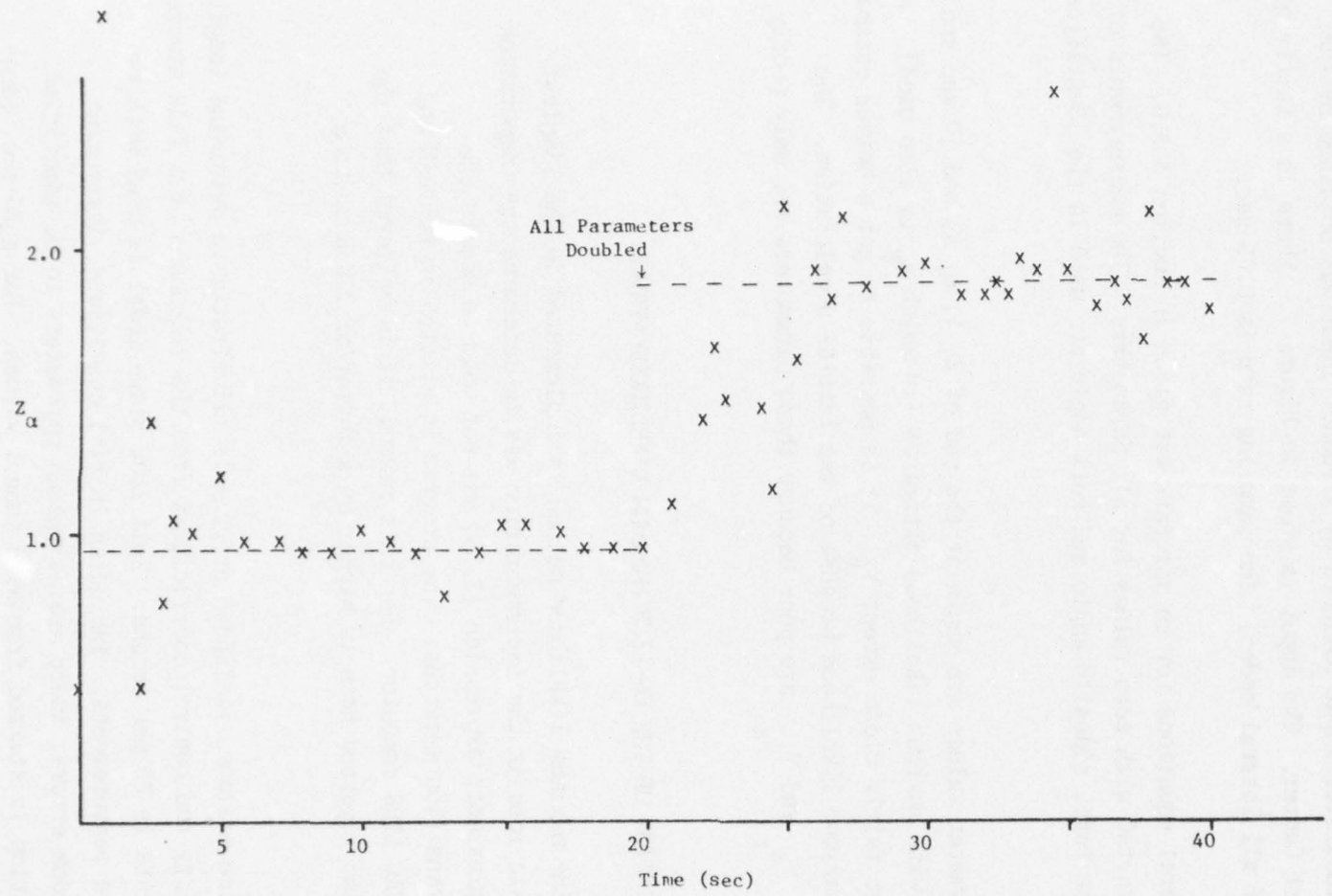


Figure 6 IVA Estimates With ρ = 0.97 and Suboptimal Kalman Filter Gain

The lateral data for an advanced swept wing Navy fighter for a differential horizontal stabilizer input followed by a rudder input was obtained by the Naval Air Test Center. The input is shown in Figure 7. There is a fairly good excitation of all lateral modes. The sampling rate is 0.05 sec.

The lateral equations for an aircraft are given in Section 3.3.6. The program is started with zero values for all parameters. The measurements of roll rate, yaw rate, sideslip angle and roll angle are used in the identification.

The parameter values are shown at the end of 2, 4, 6, 8, and 10 sec and are compared with maximum likelihood estimates (in which a_y is also used). All parameters are fairly close except Y_β. It is possible to get a better estimate of Y_β using maximum likelihood because of the lateral acceleration. The estimates of Y_{δ_r} and Y_{δ_a} are poor because these parameters are only poorly identifiable.

4.5 IMPLEMENTATION OF THE ON-LINE MAXIMUM LIKELIHOOD METHOD

The on-line maximum likelihood method is implemented in the simplest form. The basic idea in the implementation was to determine the computation time. The approximate expression (3.75) was not used to update the sensitivity transition matrices. The program is written in FORTRAN and used on a UNIVAC 1108 computer. For this reason, it is believed that the computation time reported here is higher by a factor of 2 and up to a factor of 5.

The on-line maximum likelihood program is illustrated to determine longitudinal stability and control coefficients from the responses of a T-2B aircraft. The sampling rate is 20 per second. Full four state model is used with two inputs and five measurements. Including initial conditions, instrument biases and random errors, there are seventeen parameters to be identified. The identification is started from wind tunnel values. For a 20 sec. long

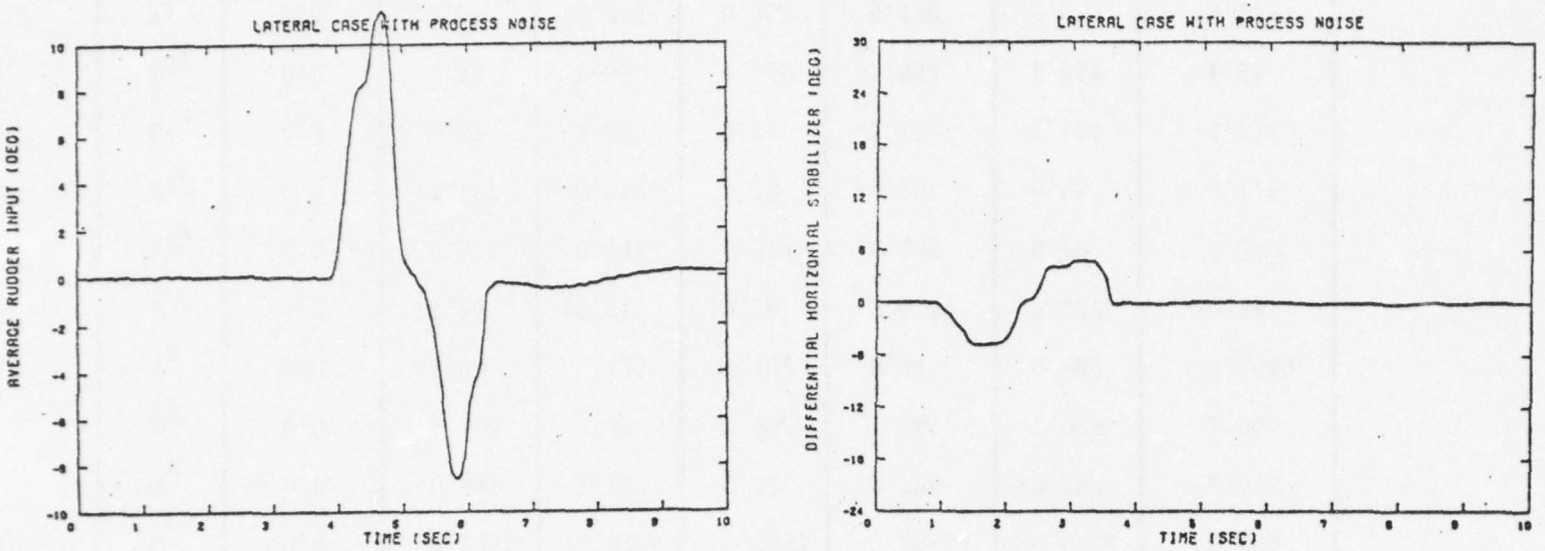


Figure 7 Rudder and Aileron Inputs for Lateral Excitation

Table 9

IVA Estimates for a Swept Wing Navy Fighter
(Dimensional Derivatives)

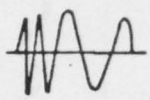
PARA-METER	STARTING VALUE	IDENTIFIED VALUE AT THE END OF:					MAXIMUM LIKELIHOOD VALUE
		2	4	6	8	10	
L_p	0.0	-3.148	-2.844	-1.806	-1.631	-1.604	-2.004
L_r	0.0	26.84	6.32	2.475	2.703	2.795	2.686
L_β	0.0	-20.60	12.98	-7.00	-6.21	-6.04	-8.49
N_p	0.0	-0.1054	-0.0274	-0.0093	-0.00554	-0.0034	-0.004
N_r	0.0	-0.568	-1.12	-0.81	-0.786	-0.784	-0.624
N_β	0.0	4.496	3.45	1.962	1.967	1.975	1.96
Y_β	0.0	-4.84	0.125	0.402	0.44	0.462	-0.0649
L_{δ_a}	0.0	15.76	16.21	10.56	9.664	9.527	10.28
N_{δ_a}	0.0	0.675	0.444	0.314	0.289	0.28	0.240
Y_{δ_a}	0.0	-0.167	-0.0526	-0.20	-0.20	-0.56	N.I.*
L_{δ_r}	0.0	-19.37	-5.68	-0.74	-1.17	-1.124	-3.334
N_{δ_r}	0.0	1.41	1.62	1.569	1.554	1.554	1.55
Y_{δ_r}	0.0	2.6	0.243	0.458	0.430	-0.67	N.I.*

* Not Identifiable

run, (400 data points), the per iteration time is 4 seconds. The program converges in 5 iterations for a total CPU time of 20 seconds. The parameter values are shown in Table 10 and the time history plots in Figure 8. Note that this has been achieved without any loss in accuracy of estimates.

Table 10

Identified Derivatives for T-2B Aircraft
Speed: 679 ft/sec
Altitude: 10,000 ft

INPUT	 random
PARAMETER	
Z_α	- 2.621 *(0.0253)*
Z_u	*
Z_q	1.0
X_α	*
X_u	*
X_q	*
M_α	-25.29 (2.28×10^{-9})
M_u	0.0017
M_q	- 4.656 (9.21×10^{-9})
$Z_{\delta e}$	0.1917 *(0.0154)*
$X_{\delta e}$	*
$M_{\delta e}$	-25.51 *(0.114)*

EQUATION BIAS	
Z_0	-0.00216
M_0	*0.580*
MEASUREMENT BIAS	
b_α rad	-0.00234
b_u ft sec^{-1}	-3.3
b_q rad sec^{-1}	-0.00113
b_θ rad	0.00146
b_{a_z} ft sec^{-2}	-0.207
RANDOM NOISE STANDARD DEVIATIONS	
σ_α rad	0.0054
σ_u ft sec^{-1}	3.17
σ_q rad sec^{-1}	0.0287
σ_θ rad	0.0198
σ_{a_z} ft sec^{-1}	9.35

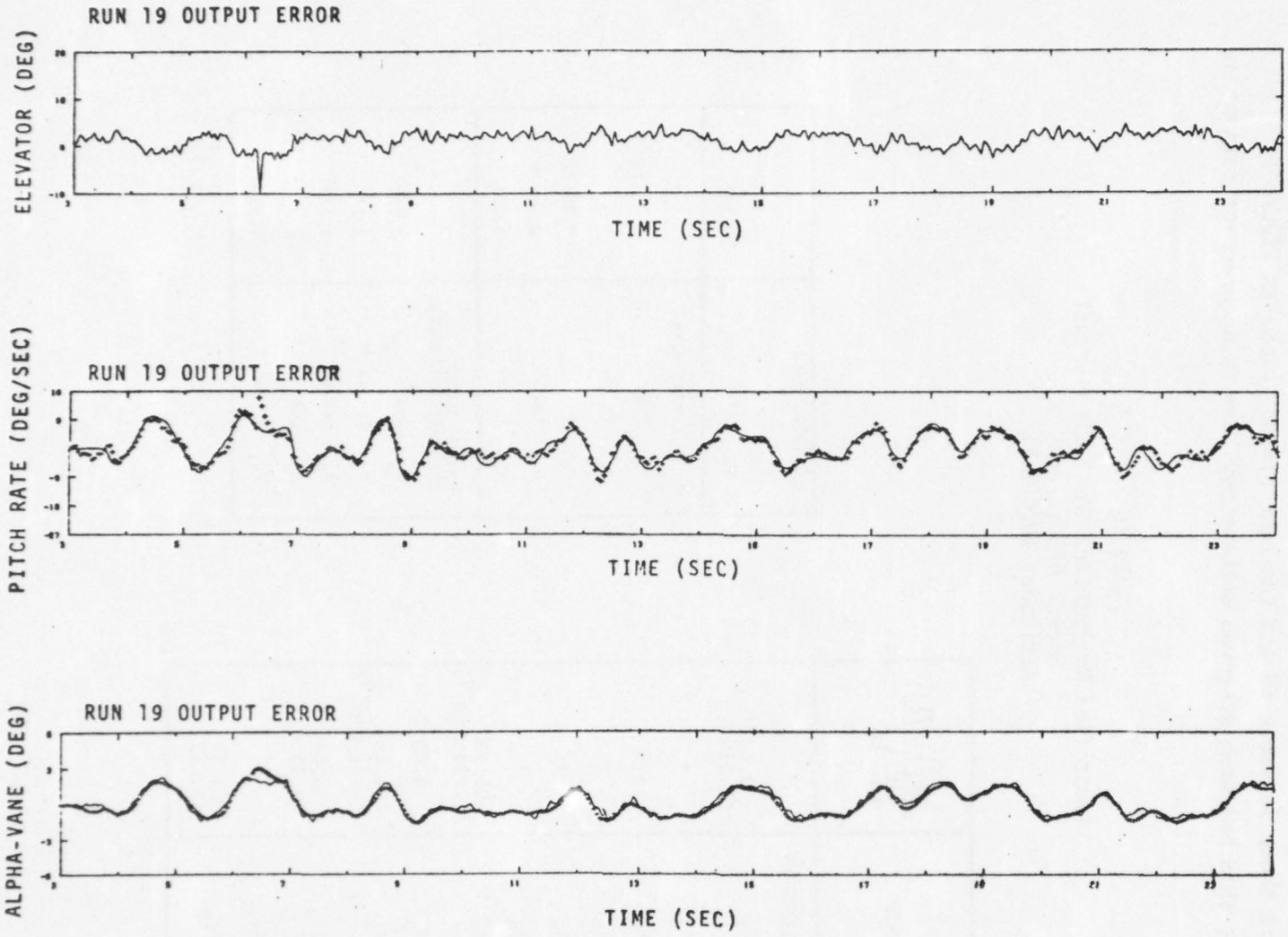


Figure 8 Identification of T-2 Derivatives Using the On-Line Maximum Likelihood Program

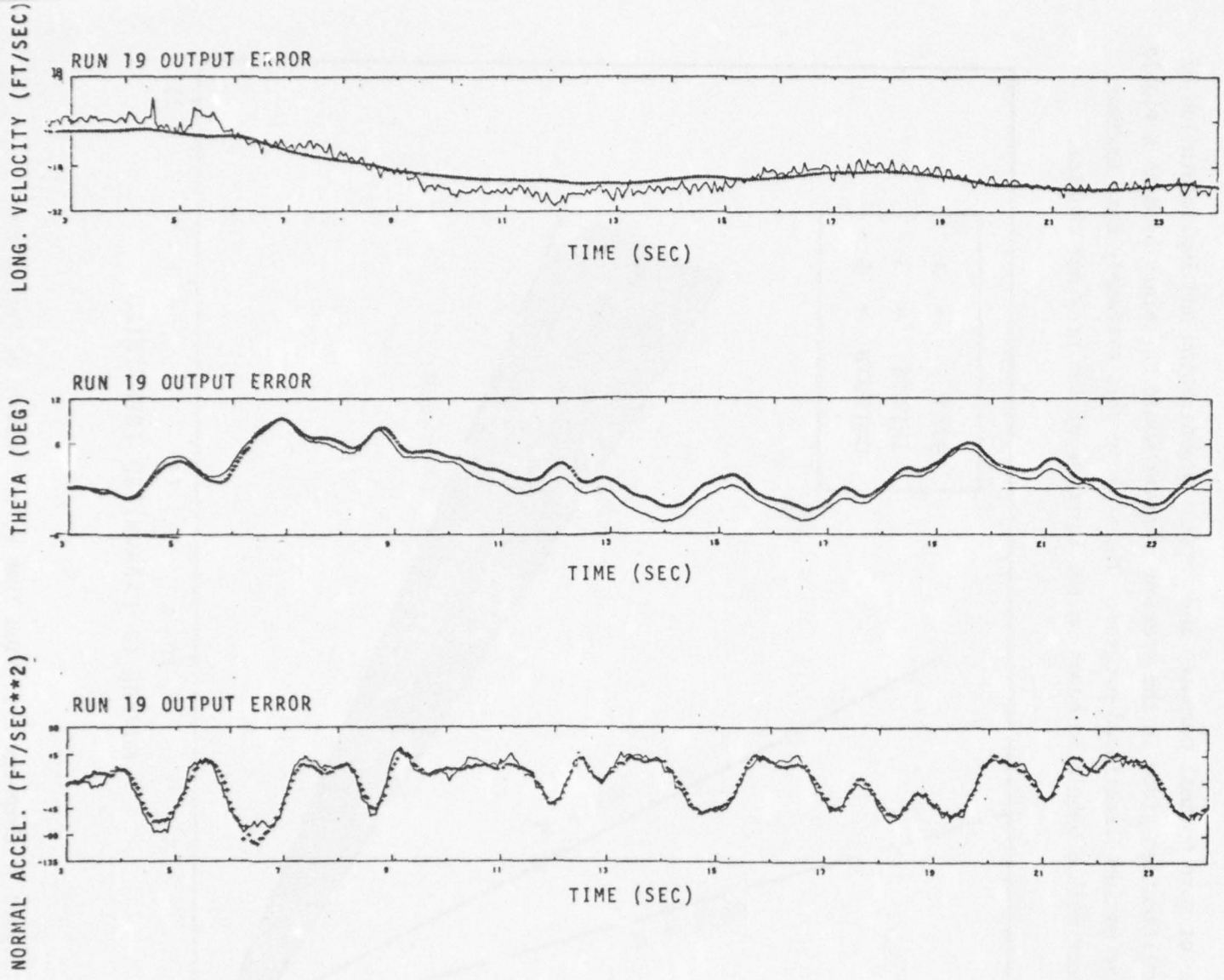


Figure 8 (Concluded) Identification of T-2 Derivatives Using the On-Line Maximum Likelihood Program

Figure 9 shows the computation time required to implement the on-line maximum likelihood program on a UNIVAC 1108 for different numbers of identified parameters and data points.

Use of more optimal computer instruction organization and implementation of the simplification given in the previous chapter would certainly produce a viable real time maximum likelihood program. The work on this extremely fast maximum likelihood method should be based on the formulae of the previous chapter.

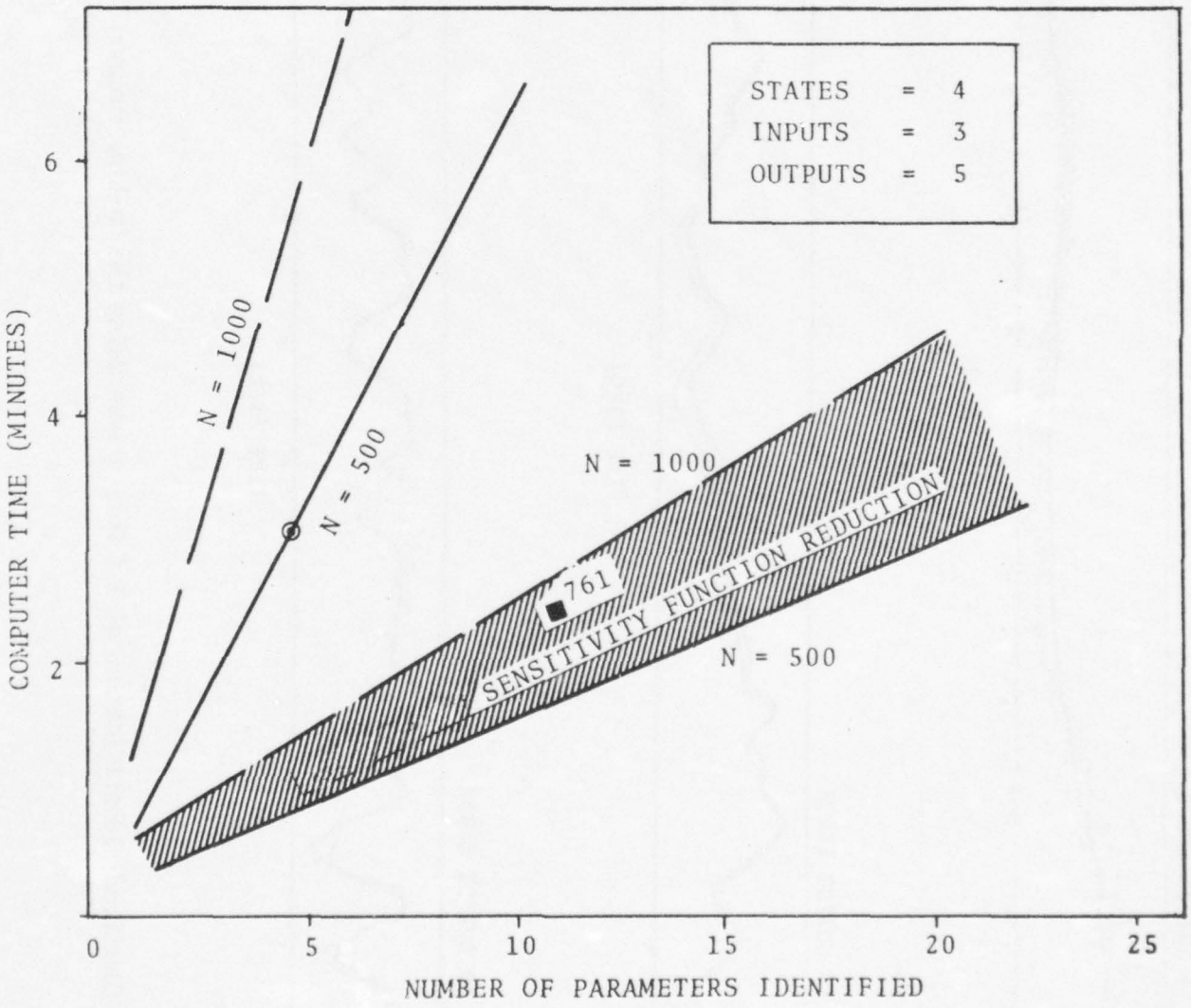


Figure 9 Computer Execution Time Versus Number of Parameters Identified for SCIDNT with Sensitivity Function Reduction

4.6 SUMMARY

The results presented in this chapter clearly demonstrate that the instrumental variables and on-line maximum likelihood approaches are two viable techniques for on-line and real time identification of parameters. By fine tuning the algorithms to the specific problem and the computer, the computation time and storage could be reduced to meet the constraints of the real time application.

V. ON-LINE EVALUTION OF DATA QUALITY FOR IDENTIFICATION

5.1 INTRODUCTION

A considerable amount of flight test data may have to be discarded because of its inability to identify stability and control derivatives to the desired accuracy. This happens when instruments deteriorate or fail, the input signal is unsatisfactory, or there is intermittent loss of the telemetry link. Often the flight tests have to be repeated to obtain this data again causing additional delay and increased cost. Some flight testing facilities around the country have aquired modern equipment to monitor the flight test in real time. An example of such equipment is the Real Time Processing System (RTPS) installed by the Navy at the Naval Air Test Center, (NATC), Patuxent River, Maryland.

Real time flight monitoring systems like the RTPS have vastly increased the capabilities for quick evaluation of flight tests. This could have a tremendous payoff because if a certain maneuver is unlikely to give reliable results, that maneuver could be repeated immediately. This would obviate the need for waiting for future flight tests to repeat this maneuver. Secondly, most of the "bad" data will be isolated before it is passed through the more expensive model structure determination and parameter identification algorithms. Thus, to make full use of this capability, there is need for on-line evaluation of data quality.

On-line and real time data quality analysis is a broad subject and may be called upon to answer many different questions. In this chapter, attention will be focused on how to determine if certain data is useful for the purpose of system identification. The next section will discuss the various requirements for a viable real time data quality analysis algorithm. The techniques for determining instrument deteriorations and failures is given in Section 5.3. Section 5.4 deals with the

determination of the effect of deteriorated or failed instruments on parameter identifiability. The case of poor system excitation is discussed in Section 5.5. The chapter concludes with a discussion of useful criterion for evaluating data quality in Section 5.6.

5.2 REQUIREMENTS FOR ON-LINE EVALUATION OF DATA QUALITY

There are three considerations which define the requirements for an on-line and real time data quality analyzer. These are: (a) a priori information, (b) desired accuracy on data quality evaluation, and (c) implementation environment.

The complexity of the on-line data quality analysis algorithm depends strongly on what can be assumed known about the system before the flight test is conducted. One of the most important among them is the knowledge about the system structure and approximate parameter values. If this is unknown, it is essential to obtain approximate parameter estimates using techniques of Chapters II and III. Of course, this would result in a considerable increase in the computation burden. The second factor is the knowledge of errors in the instruments and their failure and deterioration modes. A priori information about the control surface actuators, telemetry link, control and aircraft non-linearities are also useful.

The second consideration in the real time and on-line data quality evaluation is the accuracy with which it must be determined. In most practical applications, it is not necessary to compute the entire information matrix. Usually, since a yes/no answer is all that is desired, a measure of the dispersion matrix (inverse of the information matrix and Cramer-Rao lower bound on parameter estimation errors) is sufficient. Two suitable measures are the weighted trace and the determinant of the dispersion matrix, or an upper bound on either of these two measures. The upper bounds would ensure that at least a certain accuracy can be achieved in parameter estimates.

The implementation environment determines if the data quality evaluation can be performed in real time or if it is necessary to wait after a maneuver is completed. The structuring of the algorithm also depends on the constraints of a particular situation. However, the real time and on-line application clearly requires that the guidelines for computer implementation of on-line identification methods detailed in Chapter II be followed here.

The data quality will be analyzed in the light of three main causes of unsatisfactory identification results from a given data: (a) instrument deterioration and failures, (b) insufficient system excitation or poor inputs, and (c) loss of communication link, i.e., missed sample points. The techniques for detecting instrument degradations and failures are discussed first. The next sections deal with algorithms for isolating poor inputs. These functions can be performed simultaneously or sequentially.

5.3 DETERMINATION OF INSTRUMENT FAILURES AND DEGRADATION

The tests which can be used to determine instrument accuracy and check for their failures can be divided into two classes: (a) those which explicitly or implicitly use the system model, and (b) those which are solely dependent on single instrument temporal or inter-instrument consistency checks. We discuss two tests from the second class here.

5.3.1 Temporal Correlations

Let the system equations be

$$x(n+1) = \phi x(n) + Gu(n) + \Gamma w(n) \tag{5.1}$$

and there are noisy measurements of all state variables.

$$y(n) = x(n) + v(n) \tag{5.2}$$

A recursive equation can be written for the measurements

$$y(n+1) = \phi y(n) + Gu(n) + v(n+1) - \phi v(n) + \Gamma w(n) \tag{5.3}$$

Let ϕ_n and G_n be the best estimate of ϕ and G

$$y(n+1) - \phi_n y(n) - G_n u(n) = (\phi-\phi_n)\, y(n) + (G-G_n)\, u(n)$$

$$+ v(n+1) - \phi v(n) + \Gamma w(n) \tag{5.4}$$

Define

$$z(n) \triangleq y(n+1) - \phi_n y(n) - G_n u(n) \tag{5.5}$$

For high sampling rate, $(\phi-\phi_n)$ and $(G-G_n)$ are small compared to the noise terms even when the parameters are far from true values. The autocorrelation of $z(i)$ becomes

$$Z(0) = E\{z(i)\, z^T(i)\} = R + \phi R \phi^T + \Gamma Q \Gamma^T \tag{5.6}$$

$$Z(1) = E\{z(i+1)\, z^T(i)\} = -\phi R \tag{5.7}$$

If the sampling rate is very high, ϕ can be approximated by identity to give

$$Z(0) = 2R + \Gamma Q \Gamma^T \tag{5.8}$$

$$Z(1) = -R$$

and

$$z(n) = y(n+1) - y(n) - G_n u(n)$$

Estimates of $Z(0)$ and $Z(1)$ are obtained from the measurements

$$\hat{Z}(0) = \sum_{i=1}^{N} z(i)\, z^T(i) \tag{5.10}$$

$$\hat{Z}(1) = \sum_{i=1}^{N} z(i+1)\, z^T(i) \tag{5.11}$$

The reason we can replace ensemble average by time average is that for high sampling rate, z is a stationary process. $\hat{Z}(1)$ can now be used as a simple check for the instrument accuracy and correct functioning. If

$$H_0 : \quad Z_{jj}(1) = -R_{jj}$$

$$H_1 : \quad Z_{jj}(1) \neq R_{jj} \tag{5.12}$$

the hypothesis H_0 is tested against the alternate hypothesis H_1. A low Z_{jj} could result from dead instrument, base connection, incorrect gain, etc. A high Z_{jj} would signify instrument failure and degradation and noisy channel.

Notice that this technique, though simple, can detect only few of the many faults that can occur. Its effectiveness in isolating the failure would often be limited.

5.3.2 Inter-Instrument Comparisons

In the absence of good system models, this method can be applied only if two or more instrument outputs are related by kinematic equations. An example is the measurements of pitch angle, pitch rate and pitch acceleration. The technique is illustrated by the following example.

The equations governing θ, q and $\dot{q}$ are approximately

$$\theta(n+1) = \theta(n) + \Delta q(n) \tag{5.14}$$

$$q(n+1) = q(n) + \Delta\dot{q}(n)$$

Since there are discrete noisy measurements y_1, y_2 and y_3 of θ, q, and $\dot{q}$, we get

$$\begin{pmatrix}\theta(n+1)\\ q(n+1)\end{pmatrix} = \begin{pmatrix}1 & \Delta\\ 0 & 1\end{pmatrix}\begin{pmatrix}\theta(n)\\ q(n)\end{pmatrix} + \begin{pmatrix}0\\ \Delta\end{pmatrix} y_3(n) - \begin{pmatrix}0\\ \Delta\end{pmatrix} v_3 \tag{5.16}$$

and

$$\begin{pmatrix}y_1(n)\\ y_2(n)\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}\theta(n)\\ q(n)\end{pmatrix} + \begin{pmatrix}v_1\\ v_2\end{pmatrix} \tag{5.17}$$

y_3 is, in this formulation, the input signal for the θ, q system. Since all state definition matrices are known a Kalman filter (adaptive or nonadaptive) can be propagated in real time. The innovations covariance is related to the covariance of the random noises v_1, v_2 and v_3. The instrument failures can be detected quickly. This method has been used for data smoothing by Molusis [23].

Another simple and effective procedure can be used when the kinematic relations are as simple as this case.

$$y_1(n+1) - y_1(n) = \Delta y_2(n) + v_1(n+1) - v_1(n) - \Delta v_2(n) \tag{5.18}$$

$$y_2(n+1) - y_2(n) = \Delta y_3(n) + v_2(n+1) - v_2(n) - \Delta v_3(n) \tag{5.19}$$

Instrumental variables technique is used to estimate the coefficients of $y_2(n)$ in Equation (5.18) and $y_3(n)$ in Equation (5.19). Let the estimated values be Δ_1 and Δ_2, respectively. They are compared with the true values

of Δ. If this is done, it is simple to detect and isolate instrument failures, if only one instrument fails. The following "truth" table can be used.

	$\Delta_1 = \Delta$	$\Delta_1 \neq \Delta$
$\Delta_2 = \Delta$	All Instruments Working	Pitch Angle Incorrect
$\Delta_2 \neq \Delta$	Pitch Acceleration Incorrect	Pitch Rate Gyro Failure

The instrumental variables method is simple to use because the model is not exactly as long as there are no failures. Note that the instrument noises can also be determined. This procedure can be modified for more complicated kinematic models.

An on-line adaptive Kalman filter can also be used to determine various noise covariances. This technique would be useful in determining instrument degradations rather than catastrophic failures.

5.4 EFFECT OF INSTRUMENT FAILURES AND DEGRADATIONS ON PARAMETER IDENTIFIABILITY

The evaluation of the reduction in identifiability of parameters resulting from the failure or degradation of one or more instruments, is facilitated by isolating the contribution of each instrument towards the estimation accuracy of the parameter values. Therefore, the information matrix, a measure of the information about parameter values, is decomposed for various instruments.

Consider a continuous time representation of a linear time varying system

$$\dot{x} = Fx + Gu \tag{5.20}$$

with measurements

$$y = Hx + v \tag{5.21}$$

where x is a $n \times 1$ state vector, y is a $p \times 1$ output vector and u is a $q \times 1$ input vector. F, G and H are matrices of appropriate dimensions and depend upon m unknown parameters θ. The information matrix for parameters θ, for a test of length T, is

$$M = \int_0^T \frac{\partial (Hx)^T}{\partial \theta} R^{-1} \frac{\partial (Hx)}{\partial \theta} dt \tag{5.22}$$

where R is the power spectral density of the measurement noise v. Equation (5.22) can be written as

$$M = \sum_{i,j=1}^{p} \int_0^T \frac{\partial (H_i x)^T}{\partial \theta} (R^{-1})_{ij} \frac{\partial (H_j x)}{\partial \theta} dt$$

$$\underline{\Delta} \sum_{i,j=1}^{p} M(i,j) \tag{5.23}$$

The single subscripts denote a row or column of a matrix and the double subscripts an element of the matrix

$$M = M(1,1) + (M(1,2) + M(2,1) + M(2,2)) \ldots \tag{5.24}$$

$$= M^{(1)} + M^{(2/1)} \ldots + M^{(p/p-1,p-2\ldots1)} \tag{5.25}$$

The information matrix decomposition of Equation (5.25) is very useful. The first matrix on the right hand side is the information matrix about the parameters if only the first instrument is available. The second matrix is the additional information provided by the second instrument,

given that the first instrument is already present. If the measurement noise in different channels is independent, this decomposition of the information matrix simplifies

$$M(i,j) = 0 \qquad i \neq j \tag{5.26}$$

giving

$$M = M^{(1)} + M^{(2)} + M^{(3)} + \ldots + M^{(p)} \tag{5.27}$$

Notice that, in this case, the information supplied by each measurement is independent of the presence of other instruments. We consider this case from now on. The generalization to correlated measurement noise is obvious.

The dispersion matrix (inverse of the information matrix and Cramer-Rao lower bound on standard deviations of parameter estimation errors) is

$$D = \left(M^{(1)} + M^{(2)} + M^{(3)} \ldots M^{(p)}\right)^{-1} \tag{5.28}$$

5.4.1 Instrument Failures

If the above formulation is used, the effect of instrument failures on parameter estimation accuracy can be evaluated conveniently. The increase in dispersion matrix because of failure of the k^{th} instrument is

$$D_F^{(k)} - D = \left(D^{-1} - M^{(k)}\right)^{-1} - D = D\left(I - M^{(k)}D\right)^{-1} M^{(k)}D \tag{5.29}$$

$(I - M^{(k)}D)$ is a positive semidefinite matrix. If it is not positive definite, the loss of k^{th} instrument would make one or more parameters unidentifiable. (The number of parameters which are unidentifiable equals the number of zero eigenvalues of $(I - M^{(k)}D)$). Assume that no parameter becomes unidentifiable from the failure of the k^{th} instrument.

Then $D_F^{(k)} - D$ is positive semidefinite with rank equal to that of $M^{(k)}$. In other words, the failure of the k^{th} instrument will reduce the identifiability of all parameters. Equation (5.29) gives a quantitative measure of the increase in covariance of parameter estimation errors.

5.4.2 Instrument Deterioration

The information matrix decomposition is useful in determining the increase in estimation errors when the errors in instruments increase. Suppose the power spectral density, R_{kk}, in the k^{th} instrument increases. Differentiating Equation (5.28) with respect to R_{kk} and using Equations (5.23) and (5.27)

$$\frac{\partial D}{\partial R_{kk}} = - D \frac{\partial M}{\partial R_{kk}} D$$

$$= \frac{1}{R_{kk}} D \; M^{(k)} D \tag{5.30}$$

This can be written as

$$\frac{\Delta D}{(\Delta R_{kk}/R_{kk})} = D \, M^{(k)} D \tag{5.31}$$

Thus, the increase in dispersion matrix for a certain relative decrease in R_{kk} is directly proportional to $M^{(k)}$.

5.5 DETERMINATION OF INSUFFICIENT EXCITATION AND POOR INPUTS

The determination of insufficient excitation from poor inputs can be done through evaluation of the information matrix or a measure of the information matrix. The development of the sensitivity covariance accumulation technique, detailed in Chapter III, makes it possible to compute the information matrix on a real time basis. This will be computed for the operating instruments at the a priori parameter value or an approximate estimate of the parameters.

There are several problems with the direct computation of the information matrix, of Chapter III, for the present application. In the output error mode, the information matrix depends only on the input and the instrument errors, but not on actually observed measurements. This makes it very important that the instrument errors be known and also whether the instruments are functioning normally. Therefore, the techniques of the last two sections have to be used to determine instrument accuracies. The second and sometimes a more important problem is incorrect parameter values. The computed information matrix depends on the a priori parameter values and could be in error because of incorrect parameters. This happends when the predicted state variables are significantly different from the true states. There are two methods to handle this problem.

In the first technique, the $x(k)$ on the right hand side of Equation (3.61) is replaced by the measured value of the state variables at that instant of time, i.e., the sensitivity equations are written as

$$\frac{\partial}{\partial\theta_j}\hat{x}(k+1) = \frac{\partial\phi}{\partial\theta_j}y(k) + \hat{\phi}\frac{\partial\hat{x}(k)}{\partial\theta_j} + \frac{\partial G}{\partial\theta_j}u(k) \tag{5.32}$$

The error in state sensitivity follows the equation

$$\frac{\partial}{\partial\theta_j}\tilde{x}(n+1) = -\frac{\partial\phi}{\partial\theta_j}v(n) + \frac{\hat{\phi}\partial\tilde{x}(n)}{\partial\theta_j} + (\phi-\hat{\phi})\frac{\partial x(n)}{\partial\theta_j} \tag{5.33}$$

The last term is small when the sampling rate is high and the parameters are not too far from the true values. With this assumption, we can get a difference equation for the covariance of the error

$$P_j(n+1) = \hat{\phi}\,P_j(n)\hat{\phi} + \frac{\partial\phi}{\partial\theta_j}R\frac{\partial\phi}{\partial\theta_j}^T \tag{5.34}$$

The number of sensitivity equations can be reduced as before considering $y(n)$ and $u(n)$ as inputs to the system. This technique can be generalized to the case where measurements of all state variables are not available.

The second approach is to estimate the state vector using a Kalman filter and the best available parameter values. The process noise covariance may have to be increased artificially to account for the modeling error. The estimated state vector, in this case, will have both random and systematic errors. In general, this will be much closer to the true state than a purely propagated value. By incorporating the measurements, the system outputs are directly used in the determination of the information matrix. The equations are

$$\frac{\partial x(n+1)}{\partial\theta_j} = \phi\frac{\partial x(n)}{\partial\theta_j} + \frac{\partial\phi}{\partial\theta_j}\hat{x}(n) + \frac{\partial G}{\partial\theta_j}u(n) \tag{5.35}$$

$$\hat{x}(n+1) = \phi\hat{x}(n) + Gu(n) + \phi k(y(n) - H\hat{x}(n)) \tag{5.36}$$

An advantage of this method is that it is easy to use when measurements of all state variables are not available.

Example 1:

Consider the continuous system representation

$$\dot{x} = - ax + u \tag{5.37}$$

with continuous measurements

$$y = x + v \tag{5.38}$$

The "true value" of a is 1 and the a priori value is 1.5. Consider the information about parameters resulting from an input

$$u(t) = \delta(t) \qquad 0 \leq t \leq 2 \tag{5.39}$$

The sensitivity equation for parameter a is

$$\frac{d}{dt}\frac{\partial x}{\partial a} = - x - a\frac{\partial x}{\partial a} \tag{5.40}$$

True Information Matrix

$$x(t) = e^{-t}$$

$$\frac{\partial x}{\partial a} = - te^{-t} \tag{5.41}$$

$$M = \frac{1}{r}\int_0^z \left(\frac{\partial x}{\partial a}\right)^2 dt \tag{5.42}$$

$$= \frac{1}{r}\left[-\frac{t^2e^{-2t}}{2} - \frac{te^{-2t}}{2} - \frac{e^{-2t}}{4}\right]_0^2$$

$$= \frac{1}{r}\left[\frac{1}{4} - \frac{9}{4}e^{-4}\right] = \frac{.2088}{r} \tag{5.43}$$

Information Matrix Computed Using A Priori Value of the Variable

It is straightforward to show that

$$M_1 = \frac{1}{r}\ \frac{2}{27} - 1\ \frac{23}{27}\ e^{-6} \quad = \frac{.06948}{r} \tag{5.44}$$

Information Matrix Using the Measured State Variable

Since

$$x(t) = e^{-t} \tag{5.45}$$

$$y(t) = e^{-t} + v \tag{5.46}$$

The equation governing the sensitivity of the state to parameter a is

$$\frac{d}{dt}\ \frac{\partial x}{\partial a} = -\ 1.5\ \frac{\partial x}{\partial a} - e^{-t} - v \tag{5.47}$$

The expected value of $(\partial x/\partial a)$ is

$$\frac{\partial x}{\partial a} = 2(e^{-1.5t} - e^{-t}) \tag{5.48}$$

and its variance is

$$X_a = \frac{r}{3}\ (1 - e^{-3t}) \tag{5.49}$$

The expected value of the information matrix is

$$\bar{M}_2 = \frac{.115}{r} + .555 \tag{5.50}$$

The second term in Equation (5.50) is a bias due to the noise term driving the sensitivity Equation (5.47). For a reasonable signal-to-noise ratio, r is much smaller than one. If this is not so, the bias may produce a serious error in the computed information matrix. Notice also that the bias is independent of the input and the response and therefore can be corrected for, as long as the measurement noise covariance is known. It is clear that the second approach gives a better approximation to the information matrix.

Remarks:

1. Many methods (e.g., Denery [24] fall under the general class of methods presented in the previous sections. Any of these could also be used for the on-line evaluation of data quality.

2. Sensitivity function method gives a more accurate information matrix, at the cost of increased computation time. This method is also more general and can be used with slight loss of efficiency when all measurements are not available.

3. The estimate of the state based on previous measurements need not be optimal. However, there may be a serious error in the information matrix if there are large errors in state estimates.

4. The instrument checks and information matrix evaluation can be carried out simultaneously on-line. The information matrix can be determined as if all instruments are functioning normally. The kinematic Kalman filters can be propagated in parallel to determine if the instruments are functioning.

5.6 SCALAR CRITERIA

In the last three sections, we developed algorithms for quick determination of the dispersion matrix for on-line and real time use. To make more effective use of real time data quality analysis, there is a need to translate the dispersion matrix into a scalar criterion, upon which the engineer could base the success or failure of a flight test.

The trace, the weighted trace and the determinant of the dispersion matrix are possible measures. The following recursive relations can be used to update the trace and the determinant of the dispersion matrix.

$$\mathrm{Tr}(D_N) = \mathrm{Tr}(D_{N-1}) - \mathrm{Tr}\left[D_{N-1} H_\theta^T(t_N) \{R^{-1} + H_\theta(t_N) D_{N-1} H_\theta^T(t_N)\}^{-1} H_\theta(t_N) D_{N-1}\right] \tag{5.51}$$

$$\left|M_N\right| = \left|M_{N-1}\right| \times \left|I + M_{N-1}^{-1} H_\theta^T(t_N) R^{-1} H_\theta(t_N)\right| \tag{5.52}$$

Updating the trace or the determinant of the dispersion matrix requires storing the entire dispersion matrix and updating it. A lower bound can be obtained for the determinant of the dispersion matrix which does not require that the entire matrix be stored

$$\left|M_{N+k}\right| \geq \left\{\left|M_N\right|^{\frac{1}{m}} + \left|M_{N,N+k}\right|^{\frac{1}{m}}\right\}^m \tag{5.53}$$

$$M_{N,N+k} = \sum_{i=1}^{k} H_\theta^T(t_{N+i}) R^{-1} H_\theta(t_{N+i}) \tag{5.54}$$

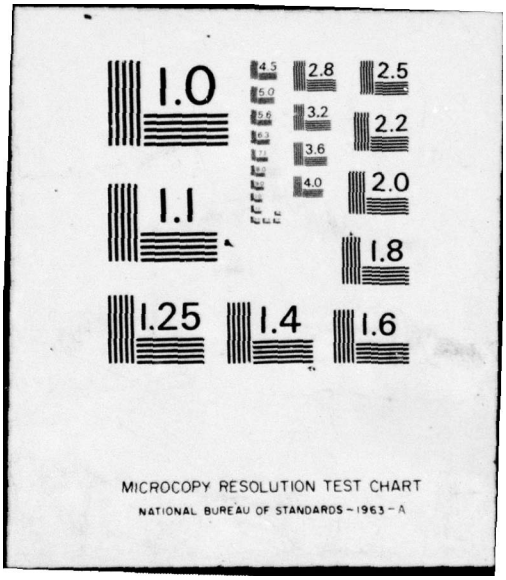
The inequality of Equation (5.53) is useful when $M_{N,N+k}$ is not too small.

AD-A032 145 SYSTEMS CONTROL INC PALO ALTO CALIF F/G 12/1
METHODS FOR THE REAL TIME IDENTIFICATION OF VEHICLE PARAMETERS.(U)
FEB 75 N K GUPTA, W E HALL N00014-72-C-0328
UNCLASSIFIED NL

2 OF 2
AD
A032145

END

DATE
FILMED
1-77


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS - 1963 - A

5.7 SUMMARY

This chapter presented some results on the techniques for on-line evaluation of data quality. The detection of instrument linear failures and degradations are discussed and their effects on reducing parameter identifiability are considered. The problem of insufficient system excitation is also treated in detail. The use of these techniques would lead to a considerable saving in flight test time.

VI. CONCLUSIONS

A review of parameter identification methods reveals that the accuracy and convergence properties of off-line algorithms is not generally characteristic of on-line and real time programs. This work has produced techniques which attempt to bridge the gap between such requirements by developing a very accurate on-line algorithm and a fast off-line algorithm. These new techniques are basically as follows:

(a) A State Variable Instrumental Method which gives statistically efficient and computationally convergent algorithms.

(b) A Linear Maximum Likelihood Algorithm which incorporates a new sensitivity functions reduction method, producing a near-real time program with no loss of accuracy.

Both of these techniques have been evaluated on simulated and flight test data. Results for the instrumental variable approach (IVA) showed notable improvement over previous implementation. No loss in accuracy occurs because the sensitivity functions technique upon which the time reduction is based, is an exact algebraic reduction of the calculation requirements.

Further reductions in time are possible by use of several methods, including the following:

(1) Use of preprocessing data consistency and evaluation algorithms to optimally filter spurious errors.

(2) Use of particular approximations to various stages of the maximum likelihood algorithm.

The analytical, computational, and evaluation results of this work lead to the following principal conclusions:

(1) Significant savings in computational time can be achieved by using on-line or specialized off-line programs to estimate stability and control derivatives. Such a capability can be applied to effect reduction in total flight test time.

(2) The demonstrated capability of on-line algorithms can be extrapolated to other real time applications, including advanced control system and fault detection objectives. Such extrapolation is consistent with the computer hardware advances now being implemented on advanced Naval systems.

APPENDIX A

REDUCTION IN SENSITIVITY FUNCTION COMPUTATION FOR LINEAR TIME INVARIANT SYSTEMS*

A.1 INTRODUCTION

The problem of computing state sensitivities using reduced order models has become very important in parameter estimation involving high order models and many unknown parameters. These techniques allow a considerable saving in computation time which makes the determination of optimal inputs and estimation of parameters from real data feasible for practical systems. Most efforts to date [26-28] have concentrated on finding bounds on the order of the model which can generate state sensitivities for all system parameters. Very little attention has been paid to the formulation of practical techniques leading to these lowest order models. Formulations by Wilkie and Perkins [26], Denery [27], and Neuman and Sood [28] lead to fairly complicated transformations and are not capable of exploiting the characteristics of the system in most cases.

This appendix develops a practical method for obtaining lowest order models for sensitivity functions computations. The technique makes full use of special system characteristics and has general application to high order systems with a large number of unknown parameters. The problem of computing sensitivities for initial conditions is generalized to computing sensitivities for an additional input distribution vector.

---

* This appendix is taken from Reference 25, a paper published under this contract.

This appendix is organized as follows. Section A.2 gives a statement of the problem, assumptions and notation. The problem of single input is treated in Section A.3 and is generalized for multi-input systems in Section A.4. Section A.5 presents the algorithm to implement this technique. The conclusions and some further work is indicated in Section A.6.

A.2 PROBLEM STATEMENT

Consider a system

$$\dot{x} = Fx + Gu \quad x(0) = x_o \tag{1}$$

where x is an n x 1 state vector, u is a q x 1 control vector. F and G are matrices of appropriate dimensions and are functions of m parameters θ. The system starts from the initial state x_o. An alternate representation of the system is obtained by adding one more input and converting the initial condition to zero, i.e.,

$$x = Fx + Gu + x_o u_{q+1}$$

$$\underline{\Delta}\ Fx + G'u' \quad x(0) = 0 \tag{2}$$

and

$$u_{q+1} = \delta(t)$$

where δ is the dirac delta function. This shows that the sensitivities for the initial conditions can be computed in much the same way as the sensitivities for other parameters in G. The initial condition vector must be adjoined to the control distribution matrix even while computing sensitivities for θ with known but non-zero initial condition. Since x_o is not different from parameters in G in this representation, the primes will be removed in Equation (2). Thus, Equation (1) with zero initial condition can be considered without loss of generality. The

problem is to compute the state sensitivites for parameters θ for all time in an efficient manner.

A heretofore uncited property of systems, which depends on the parameters θ, is important in sensitivity computation.

Definition 1--Structural Controllability: A system is said to be structurally controllable if it is controllable for almost all values of parameters. The system may be uncontrollable if certain relations hold among the parameters.

Definition 2--Structural Linear Dependence: A set of vectors have structural linear dependence if a linear combination of these vectors is zero for almost all values of parameters. The particular linear combination may depend on the values of the parameters.

Example 1: Consider the system

$$\dot{x} = \begin{pmatrix} \theta_1 & \theta_2 \\ \theta_3 & \theta_4 \end{pmatrix} x + \begin{pmatrix} 1 \\ 1 \end{pmatrix} u \tag{3}$$

the controllability matrix is

$$\begin{pmatrix} 1 & \theta_1 + \theta_2 \\ 1 & \theta_3 + \theta_4 \end{pmatrix}$$

The system is controllable unless $\theta_1 + \theta_2 = \theta_3 + \theta_4$. Thus, if $\theta_1 = \theta_4 = -1$ and $\theta_2 = \theta_3 = -5$, the system is uncontrollable in the classical sense but structurally controllable.

Initially, the following simplifications can be made:

a. The system is made structurally controllable (including x_o in G) by dropping uncontrollable states. Since the initial condition is zero, the system never moves into the uncontrollable subspace. This reduces the order of the system. Note that the states which are uncontrollable only for the given values of the parameters but which are structurally controllable should not be dropped.

b. All structurally linearly dependent columns of G matrix are lumped with other columns. This reduces the number of effective controls.

These two simplifications reduce computations later. However, the order of the system required to general sensitivity functions will be the same, even if these simplifications are not made.

The state sensitivity for parameter θ_j follows the differential equation

$$\frac{d}{dt}\frac{\partial x}{\partial \theta_j} = F\frac{\partial x}{\partial \theta_j} + \frac{\partial F}{\partial \theta_j}x + \frac{\partial G}{\partial \theta_j}u \qquad (4)$$

$$\frac{\partial x}{\partial \theta_j}(0) = 0$$

The state sensitivities for all parameters θ can be written as

$$\dot{x}_\theta = F_\theta x_\theta + G_\theta u \qquad (5)$$

$$x_\theta(0) = 0$$

where

$$x_\theta = \begin{bmatrix} x \\ \frac{\partial x}{\partial \theta_1} \\ \vdots \\ \frac{\partial x}{\partial \theta_m} \end{bmatrix} \qquad n(m+1) \times 1$$

$$F_\theta = \begin{bmatrix} F & & & 0 \\ \frac{\partial F}{\partial \theta_1} & F & & \\ \vdots & 0 & F & \\ \frac{\partial F}{\partial \theta_m} & 0 & 0 & F \end{bmatrix} \qquad G_\theta = \begin{bmatrix} G \\ \frac{\partial G}{\partial \theta_1} \\ \vdots \\ \frac{\partial G}{\partial \theta_m} \end{bmatrix} \tag{6}$$

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

If (F_θ, G_θ) is uncontrollable, the corresponding controllability matrix is of rank less than $(m+1)n$, say r. Let Q_1 be the set of r independent columns in the controllability matrix. If Q_2 is such that Q_1 and Q_2 form a set of $n(m+1)$ linearly independent vectors, then

$$x'_\theta \triangleq (Q_1 : Q_2)^{-1} x_\theta \triangleq \begin{pmatrix} Q_1^\dagger \\ \cdots \\ Q_2^\dagger \end{pmatrix} x_\theta \tag{7}$$

follows the differential equation (see Chen [15])

$$\dot{x}'_\theta = \begin{pmatrix} F'_{11} & F'_{12} \\ 0 & F'_{22} \end{pmatrix} x'_\theta + \begin{pmatrix} G'_1 \\ 0 \end{pmatrix} u \tag{8}$$

Since the initial condition in (8) is zero, the last (m+1)n-r uncontrollable states remain zero throughout. The remaining r states, x_c follow the differential equation

$$\dot{x}_c = F_c\, x_c + G_c u$$

$$x_c(0) = 0 \tag{9}$$

where

$$F_c \triangleq F_{11}' = Q_1^{\dagger}\, F_\theta Q_1$$

$$G_c \triangleq G_1' = Q_1^{\dagger}\, G_\theta \tag{10}$$

Also

$$x_\theta = (Q_1 : Q_2)\, x_\theta'$$

$$= Q_1 x_c \tag{11}$$

since other states in x_θ' are zero. Note that $Q_1^{\dagger}$ is a pseudo-inverse of Q_1 depending on Q_2. The transformation from F_θ, G_θ, to F_c, G_c and from x_c to x_θ does not involve Q_2 explicitly. Therefore, Q_1 can be chosen to be <u>any</u> pseudo-inverse of Q_1, for example,

$$Q_1^{\dagger} = (Q_1^T\, Q_1)^{-1}\, Q_1^T \tag{12}$$

It is assumed here that the inputs are linearly dependent. If this is not so, the number of inputs can be reduced until they are linearly independent. This will usually result in a reduction in the controllable subspace of (F_θ, G_θ) as shown in Section A.4. Under the assumption of linear independence of inputs, it is necessary and sufficient to solve a system of r

linear equations (9) to determine the state and its sensitivities at all times. The next sections investigate the nature and dimension of the controllable subspace of (F_θ, G_θ) and explore efficient methods for finding Q_1.

A.3 SINGLE INPUT SYSTEMS

In single input systems, F_θ is a $(mx1)n \times (m+1)n$ matrix and G_θ is a $(m+1)n$ vector. The following theorem holds.

Theorem 1: For a single input system, the rank of the controllability matrix of (F_θ, G_θ) cannot exceed $2n$.

Proof: The controllability matrix of (F_θ, G_θ) is

$$C_\theta = G_\theta \vdots F_\theta G_\theta \vdots \ldots F_\theta^{(m+1)n-1} G_\theta \tag{13}$$

It is easy to show that

$$F_\theta^p G_\theta = \begin{bmatrix} F^p G \\ \cdots\cdots \\ \frac{\partial}{\partial \theta_1} F^p G \\ \cdots\cdots \\ \vdots \\ \cdots\cdots \\ \frac{\partial}{\partial \theta_m} F^p G \end{bmatrix} \triangleq D(F^p G) \triangleq \left(\frac{F^p G}{0}\right) + D^*(F^p G) \tag{14}$$

If

$$F^n = \alpha_o I + \alpha_1 F \ldots \alpha_{n-1} F^{n-1}, \tag{15}$$

the (n+k)th column of C_θ is

$$D(F^{n+k-1}G) = \sum_{i=0}^{n-1} \{D^*(\alpha_i I)F^{k+i-1}G + \alpha_i D(F^{k+i-1}G)\} \quad (16)$$

The second term is a linear combination of n preceding columns of C_θ. Thus,

$$\text{rank } C_\theta = \text{rank}\{D(G) \vdots \cdots\cdots \vdots D(F^{n-1}G) \vdots \sum_{i=0}^{n-1} D^*(\alpha_i I)F^i G \vdots \cdots\cdots$$

$$\cdots\cdots \vdots \sum_{i=0}^{n-1} D^*(\alpha_i I)F^{i+n+k-1}G \vdots \sum_{i=0}^{n-1} D^*(\alpha_i I)F^{n+1}G \vdots \cdots\cdots\} \quad (17)$$

The (2n+k)th column of the right hand side matrix is

$$\sum_{i=0}^{n-1} D^*(\alpha_i I)F^{n+1+k-1}G = \sum_{i=0}^{n-1} D^*(\alpha_i I) \sum_{j=0}^{n-1} \alpha_j F^{i+j+k-1}G$$

$$= \sum_{j=0}^{n-1} \alpha_j \sum_{i=0}^{n-1} D^*(\alpha_i I)F^{i+(j+k)-1}G \quad (18)$$

which is a linear combination of n previous columns for $k \geq 0$. Therefore,

$$\text{rank } C = \text{rank}[G_\theta \vdots F_\theta G_\theta \vdots \cdots\cdots F_\theta{}^{2n-1}G_\theta] \leq 2n \quad (19)$$

Thus, the order of the system required to compute all state sensitivities for a single input system cannot exceed 2n. In many practical cases, it is smaller as shown in Example 1.

Corollary 1: If the structurally controllable subspace of (F, G) is of the order p, the maximal order of the controllable subspace of (F_θ, G_θ) is 2p.

Since F^pG is a linear combination of G, FG, ..., $F^{p-1}G$, the corollary follows immediately.

Example 2: Consider the following system

$$\dot{x} = \begin{pmatrix} \theta_1 & \theta_2 \\ 0 & -1 \end{pmatrix} x + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u \qquad x(0) = 0 \tag{20}$$

The state vector and its sensitivities for θ_1 and θ_2 form a set of six differential equations. Since the number of states is two and the number of controls is one, only the first four columns can be independent in the controllability matrix of (F_θ, G_θ). These columns are

$$\text{rank}(C_\theta) = \text{rank}\left[D\begin{pmatrix}0\\1\end{pmatrix}, D\begin{pmatrix}\theta_2\\-1\end{pmatrix}, D\begin{pmatrix}\theta_1\theta_2 - \theta_2\\1\end{pmatrix}, D\begin{pmatrix}\theta_1^2\theta_2 - \theta_1\theta_2 + \theta_2\\-1\end{pmatrix}\right] \tag{21}$$

$$= \text{rank}\begin{bmatrix} 0 & \theta_2 & \theta_1\theta_2 - \theta_2 & \theta_1^2\theta_2 - \theta_1\theta_2 + \theta_2 \\ 1 & -1 & 1 & -1 \\ 0 & 0 & \theta_2 & 2\theta_1\theta_2 - \theta_2 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & \theta_1 - 1 & \theta_1^2 - \theta_1 + 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{22}$$

$$\sum_{i=1}^{q} k_i = n \tag{26}$$

This set of linearly independent vectors in the controllability matrix span the complete n-dimensional space. So any vector can be represented as a linear combination of these vectors. In particular,

$$F^{j-1}G_k = \beta_1 G_1 + \ldots, + \beta_{k_1} F^{k_1-1} G_1, \ldots, \beta_n F^{k_q-1} G^q$$

$$1 \leq j \leq n, \quad 1 \leq k \leq q \tag{27}$$

Therefore,

$$\sum_{i=0}^{n-1} D^*(\alpha_i I) F^{i+j-1} G_k = \beta_1 \sum_{i=0}^{n-1} D^*(\alpha_i I) F^i G_1 + \beta_2 \sum_{i=0}^{n-1} D^*(\alpha_i I) F^{i+1} G_1 \ldots.$$

$$\ldots., + \beta_n \sum_{i=0}^{n-1} D^*(\alpha_i I) F^{i+k_q-1} G_q \tag{28}$$

This is a linear combination of n vectors in the right hand side matrix of (27) for all j and k (the values of β_i depend on j and k). Thus

$$\rho = \text{rank}[D(G_1, FG_1, \ldots, F^{n-1}G_1) \vdots D(G_2, FG_2, \ldots, F^{n-1}G_2) \vdots \ldots \vdots D(G_q, FG_q, \ldots, F^{n-1}G_q)$$

$$\Sigma D^*(\alpha_i I) F^i G_1, \Sigma D^*(\alpha_i I) F^{i+1} G, \ldots, \Sigma D^*(\alpha_i I) F^i G_j, \ldots, \Sigma D^*(\alpha_i I) F^{i+k_q-1} G_q]$$

$$\leq (q+1)n \tag{29}$$

Thus, the procedure for finding independent columns of C_θ consists of finding structurally independent columns of the controllability matrix of (F, G), choosing (q+1)n appropriate columns from C_θ and checking to see if there is any further linear dependence.

Another simplification is possible in large order systems in which each input controls only a small number of states. If k_i' is the dimension of the controllable subspace of the i^{th} input, no more than $2k_i'$ columns involving G_i can be linearly independent in the right hand side matrix of (29), as shown in corollary 1 and theorem 1.

Corollary 2: If for any single input u_j the system is completely controllable,

$$\rho = \text{rank}[D(G_1,\ldots,F^{n-1}G_1),D(G_2,\ldots,F^{n-1}G_2),\ldots,D(G_q,\ldots,F^{n-1}G_q),$$

$$\ldots,\Sigma D^*(\alpha_1 I)F^i G_j,\ldots,\Sigma D^*(\alpha_i I)F^{n+i-1}G_j] \tag{30}$$

The proof is obvious since G_j, ..., $F^{n-1}G_j$ form a set of n linearly independent columns.

A.5 COMPUTATION PROCEDURE

A computer program has been written to carry out this sensitivity functions reduction in linear constant-coefficient systems. The following procedure is used.

a. The initial condition, if nonzero, is appended to the input distribution matrix and the number of inputs is increased by one. The linearly dependent columns in G are merged. Then the structurally uncontrollable states in (F, G) are dropped.

b. Matrices F_θ and G_θ are formed. $k_1, k_2, \ldots, k_q$ of Equation (25) are determined and are used to choose $(q+1)n$ appropriate columns from the controllability matrix of (F_θ, G_θ).

c. The dimension k_i' of state space controllable from each input u_i alone is determined. If for any i

$$2k_i' < n + k_1 \tag{31}$$

the last $n+k_i-2k_i'$ columns involving G_i in the right hand side matrix of (29) are dropped.

d. The remaining columns are checked for linear independence. Gram-Schmidt procedure is used to drop columns, which are linearly dependent on other columns. The set of remaining columns is Q_1.

e. Any pseudo-inverse of Q_1 is determined. Equation (12) is used to compute F_c and G_c.

f. Equation (9) is solved for $x_c(t)$ and Equation (11) is used to find x_θ at the desired points.

A.6 CONCLUSIONS

This appendix presents a technique for determining state sensitivity functions for parameters in state transition matrix, input distribution matrix, and initial conditions for linear time-invariant systems. The input distribution matrix is augmented with the initial condition vector by addition of a new virtual input. The state sensitivities for initial condition is computed in the same way as the state sensitivities for other parameters in input distribution matrix.

A systematic method for finding the controllable subspace of the augmented system, in which the state vector is the system state and its sensitivities, is presented. It is necessary to start with no more than (q+1)n columns of the controllability matrix of the augmented system. These columns can be selected quickly by inspection of the controllability matrix of the initial system. If r is the dimension of the controllable subspace of augmented system, it is necessary to solve r linear equations to evaluate the state vector and its sensitivities.

This method of sensitivity functions reduction fully exploits the characteristics of the system and that the sensitivity to all parameters in the system may not be required. It leads to the minimal order model under the circumstances.

REFERENCES

[1] Gupta, N.K., and Hall, W.E., Jr., "SCIDNT I: Theory and Applications," Systems Control Report to ONR, Technical Report #3, December 1974.

[2] Goldberger, Econometric Theory, Wiley, New York, 1963.

[3] Saridis, G.N., "Comparison of Six On-Line Identification Algorithms," Automatica, January 1974.

[4] Isermann, R., "Test Cases for Evaluation and Comparison of Different Identification and Parameter Estimation Methods," IFAC Symposium on Identification, The Hague, 1973 (also Automatica, January 1974).

[5] Eykhoff, P., Editor, Proceedings of the Identification and System Parameter Estimation Conference, The Hague, June 1973.

[6] Clark, D.W., "Generalized Least Squares Estimation of the Parameters of a Dynamic Model," preprints IFAC Symposium on Identification, Prague, 1970.

[7] Hastings-James, R., and Sage, M.W., "Recursive Generalized Least Squares Procedure for On-Line Identification of Process Parameters," Proc. IEE, Vol. 116, 1969.

[8] Saridis, G.N., and Stein, G., "Stochastic Approximation Algorithms For Linear System Identification," IEEE Trans. Aut. Control, AC-13, 515-523, 1968.

[9] Chen, R.T.N., et al., "Development of the Advanced Techniques for the Identification of V/STOL Aircraft Stability and Control Parameters," CAL Report No. BM-2820-F-1, Final Report to NAVAIR, August 1971.

[10] Wong, K.Y., and Polak, E., "Identification of Linear Discrete Time Systems Using an Instrumental Variable Method," IEEE Trans. Aut. Control, AC-12, 1967.

[11] Young, P.C., "An Instrumental Variable Method for Real Time Identification of a Noisy Process," IFAC Automatica, Vol. 6, 1970.

[12] Pandaya, R., "Instrumental Variable Technique for System Identification," submitted to IEEE Trans. Aut. Control, Special Issue on Identification and Time Series Analysis, December 1974.

[13] Kashyap, R.L., "Maximum Likelihood Identification of Stochastic Linear Systems," IEEE Trans. Aut. Control, AC-15, 25-34, 1970.

[14] Gertler, J., and Bányácz, C.S., "A Recursive (On-Line) Maximum Likelihood Identification Method," IEEE Trans. Aut. Control, 816-819, December 1974.

[15] Chen, C.T., "Introduction to Linear Systems Theory," Holt, Rinehart and Winston, New York, 1970.

[16] Kaminski, P.G., "Square Root Filtering and Smoothing for Discrete Processes," SUDAAR No. 427, Dept. of Aeronautics and Astronautics, Stanford University, July 1971.

[17] Rowe, I.H., "A Bootstrap Model for the Statistical Estimation of Model Parameters," in Proc. Joint Automatic Control Conference, 966-967, Boulder, Colorado, August 1969.

[18] Denham, M.J., "Canonical Forms for the Identification of Multivariable Linear Systems," IEEE Trans. on Auto. Control, Special Issue on Identification and Time Series Analysis, December 1974.

[19] Etkin, B., Dynamics of Atmospheric Flight, Wiley, New York, 1972.

[20] Powell, M.J.D., "An Efficient Method for Finding the Minimum of a Function of Several Variables Without Calculating Derivatives," Computer Journal, 7, pp. 155-162, 1964.

[21] Stewart III, G.W., "A Modification of Davidon's Minimization Method to Accept Difference Approximation of Derivatives," JACM, Vol. 14, p. 72, 1967.

[22] Whittle, P, "On the Fitting of Multivariable Autoregression and the Approximate Canonical Factorization of a Spectral Density Matrix," Biometrika, Vol. 50, pp. 129-134, 1963.

[23] Molusis, J., "Helicopter Stability Derivative Extraction and Data Processing Using Kalman Filtering Techniques," Presented at the 28th Annual National Forum of the American Helicopter Society, Washington, D.C., Prepring No. 641, May 1972.

[24] Denery, D.G., "Identification of Systems From Input-Output Data With Application to Air Vehicles," NASA TN D-6468, August 1971.

[25] Gupta, N.K. and Mehra, R.K., "Computational Aspects of Maximum Likelihood Estimation and Reduction in Sensitivity Function Computations," IEEE Transactions on Auto. Cont., pp. 774-783, December 1974.

[26] Wilkie, D.F. and Perkins, W.R., "Generation of Sensitivity Functions for Linear Systems Using Low Order Models," IEEE Trans. on Auto. Cont., April 1969.

[27] Denery, D.G., "Simplification in the Computation of the Sensitivity Functions for Constant Coefficient Linear System," IEEE Trans. on Auto. Cont., August 1971.

[28] Neuman, C.P. and Sood, A.K., "Sensitivity Functions for Multi-Input Linear Time-Invariant Systems - II, Minimum-Order Models," International Journal of Control, 1972, Vol. 15, No. 3, pp. 451-463.

Distribution List for Technical Report #4
to Contract N00014-72-C-0328

Office of Naval Research
Department of the Navy
Arlington, Virginia 22217
ATTN: Mr. D.S. Siegel, Code 211 (4)
Mr. M. Cooper, Code 430B (1)
Dr. S. Brodsky, Code 432 (1)

Director
Office of Naval Research Branch Office
1030 East Green Street
Pasadena, California 91106
ATTN: Mr. R.F. Lawson (1)
CDR R.L. Breckon (1)

San Francisco Area Office
Office of Naval Research
50 Fell Street
San Francisco, California 94102
ATTN: Mr. J. Froman (1)

Director
U.S. Naval Research Laboratory
Washington, D.C. 20390
ATTN: Tech. Information Div. (3)
Library, Code 2029 (1)

Chief of Naval Development
Washington, D.C. 20360
ATTN: W.H. Young, NMAT-0334 (1)

Commander
Naval Air Systems Command
Washington, D.C. 20360
ATTN: R.F. Siewert, NAIR-3600 (1)
H. Andrews, NAIR-5301 (1)
R. A'Harrah, NAIR-53014 (1)
J.C. Taylor, NAIR-530141A (1)

Commanding Officer
Naval Air Development Center
Warminster, Penn. 18974
ATTN: A. Piranian (1)
R. Fortenbaugh (1)

Commanding Officer
Naval Air Test Facility
Patuxent River, MD 20670
ATTN: Bob Traskos (1)
Roger Burton (1)

Commanding Officer
Naval Ship Research and
Development Center
Bethesda, MD 20034
ATTN: W.E. Smith, Code 1576 (1)

Air Force Flight Dynamics Laboratory
Air Force Systems Command
Wright Patterson Air Force Base
Ohio 45433
ATTN: D. Bowser, AFFDL/FGC (1)
F. Thomas, AFFDL/FGC (1)
P. Blatt, AFFDL/FGL (1)

National Aeronautics & Space Admin.
Langley Research Center
Hampton, Virginia 23365
ATTN: J.R. Chambers (1)
Dr. R.L. Bowles (1)
Dr. M.J. Queijo (1)
W. Reed III (1)
Dr. J. Creedon (1)

Director
U.S. Army Air Mobility Research &
Development Laboratory
Langley Directorate
Hampton, Virginia 23365
ATTN: Robert Tomain (1)

Director
U.S. Air Mobility Research &
Development Laboratory
Ames Directorate
Moffett Field, California 94303
ATTN: Dr. I. Statler (1)
Dr. R. Carlson (1)
Dr. R.T. Chen (1)
David Key (1)

The Analytic Sciences Corp.
6 Jacob Way
Reading, Mass. 01867
ATTN: Dr. C. Price (1)

Air Force Office of Scientific Research
1400 Wilson Blvd.
Arlington, Virginia 22209
ATTN: Lt. Col. W.J. Rabe (1)

National Aeronautics & Space Admin.
Ames Research Center
Moffett Field, California 94035
ATTN: Dr. George Meyer (1)
Dr. Wayne Johnson (1)

United States Air Force
Air Force Systems Command
Aeronautical Systems Division
Wright Patterson AFB, Ohio 45433
ATTN: Mr. C.A. Skira, AFAPL/TBC (1)
Mr. L. Small, AFAPL/TV2C (1)