Advanced Methods of Model Structure Determination from Test Data

Narendra K. Gupta,* W. Earl Hall Jr.,† and Thomas L. Trankle‡
Systems Control, Inc., Palo Alto, Calif.

The technology of system identification from input/output test data has been developed into an operational procedure for improving and validating mathematical models upon which state-of-the-art guidance and control systems are based. A wide application of this technology has focused attention on one particular phase of the system identification methodology, the model structure determination phase, as an essential step. This step consists of processing the input/output data to determine the significant linear and nonlinear equations and associated parameters that are necessary to represent an observed system response. Although the model structure determination problem has been recognized for a long time, it has been attacked systematically only recently. This paper summarizes various aspects of this problem and results of some recent research. A unified approach is developed for this problem and is applied to test data from a submarine, a missile, and an aircraft engine.

Introduction

SYSTEM identification is the process of testing systems to obtain a mathematical model of their behavior. The complete system identification procedure consists of three major steps, which include test design, data processing, and model verification.[1-4] A number of techniques must be used in the data-processing stage, particularly for nonlinear operating regimes, to obtain the maximum information from the data. In the first stage, unmeasured or failed channels of data are reconstructed based on available measurements. This also gives preliminary force and moment coefficient time histories of interest. In the model structure determination (MSD) stage, the dependent forces and moments are related to the independent variables (e.g., angle of attack, Mach number, etc.) to provide the best model over the range of interest. In most cases, particularly those encountered with experimental data, the overall computation burden requires the use of a simplified method in the model structure determination step, leading to the necessity of improving the parameter estimation accuracy and refining the model structure using a nonlinear maximum likelihood method.

Although the fundamental problem of model structure estimation has been recognized since the origin of physical science, the utilization of statistical methods to determine systematically significant characteristics in nonlinear dynamic systems was presented first in 1974.[1,2] Such methods had been applied to static systems[5] (e.g., in econometrics and biometrics) and to linear systems.[6] From a more fundamental approach, however, the work of Akaike[7] and Kullback[8] provided the significant information theoretical formulation concepts for both linear and nonlinear systems. The theoretical basis of the methods discussed in this paper is found in the cited work of Thiel,[5] Akaike,[7] and Kullback.[8] The application to aerospace systems can be traced to the work of Gerlach,[9] in the Netherlands, who demonstrated the use of a linear technique on actual test data.

Our research into the model structural determination process for nonlinear systems has been isolated into three significant issues: 1) specification of classes of a priori models, 2) criteria for hypothesis testing of data against models, and 3) numerical procedures. The following sections of this paper discuss these issues one by one, followed by examples that demonstrate how system characteristics affect these issues in specific cases.

Specification of Classes of a Priori Models

The first major step in the model structure estimation process is the selection of mathematical forms to be correlated and tested against data. The specification of these models must be sufficiently broad to include the most probable relationship without attempting to consider all models. In order to form a tractable a priori model base, the following general considerations have been found to be relevant.

1) It is useful to recognize two aspects of mathematical models of physical systems. Basic models are derived from fundamental physical laws. Useful engineering models are, to some extent, empirical, wherein many of the system interrelationships are defined by tests and appropriate interpretation of the results. This obviously means that the role of physical engineering analysis in the model structure estimation process is essential.

2) The ultimate validation criteria, which include ability to explain existing data, consistency of empirical results with phenomenological considerations, and ability to predict data, dictate model form.

3) The objective for which the model is to be used will indicate the levels of complexity required and the regime over which the model should be valid. This means that there may not be a unique model to describe a particular vehicle. In fact, model selection depends mainly on the use to which the model is put.

In order to optimize the initial model for computational speed and numerical accuracy, it must include elements of known phenomenological behavior supplemented by various classes of functional forms, depending on the particular problem. In the absence of strong phenomenological evidence to the contrary, polynomials are well know as simple and versatile choices for mathematical models. Indeed, polynomials serve exceedingly well in modeling the errors of inertial quality gyros and accelerometers, lift and drag curves for aircraft, and gravitational anomalies, to cite only a few

Received July 15, 1977; presented as Paper 77-1170 at the AIAA Guidance and Control Conference, Hollywood, Fla., Aug. 8-10, 1977; revision received Nov. 14, 1977.

Index categories: Testing, Flight and Ground; Handling Qualities, Stability and Control; Computer Communications, Information Processing and Software.

*Senior Engineer, System Identification and Control Division. Member AIAA.

†Program Manager, System Identification and Control Division. Member AIAA.

‡Engineer, System Identification and Control Division. Member AIAA.

examples. If, however, it is known that the process is susceptible to a growth phenomenon, such as vortex buildup from aircraft control deflections, exponential or logarithmic functions are desirable. The use of polynomials therefore should be undertaken with a reasonably clear understanding of the limitations.

Given a set of p independent variables $x(1)$, $x(2)$, $x(3),\ldots,x(p)$ (e.g., angle of attack, control deflection, pitch rate, etc.) and a variable y that is expected to depend on these variables, the problem is to select a general polynomial form(s) to represent y in terms of the $x(i)$, $i=1,2,\ldots,p$. Factors that affect the particular polynomial form include the following: 1) a priori phenomenological information about the process; 2) range of variation of the dependent variables (and possibility of differences in physical phenomena over that range); 3) specification of whether the polynomial is to be used as is, is to be differentiated (requiring valid slope representation), or is to be integrated (requiring high reliability in the initial value); and 4) computational resources available to use the model (in terms of speed and memory).

To meet these requirements, we have used two basic polynomial formulations, as shown in Table 1:

1) A regular polynomial is used to represent a simple continuous phenomenon. This polynomial is usually linear in the unknown parameters C_i (although many cases arise in which the polynomial is nonlinear in the unknown parameters). For improved numerical conditioning, orthogonal polynomials are desired. The classical orthogonal polynomials of Legendre, Laguerre, and Hermite may be generated by three term recurrence relations that are programmed easily. The Tschebycheff polynomial is known to demonstrate the properties of both the Fourier series and the orthogonal polynomials.[10]

2) A spline of order n and continuity ν is used to represent a simple heterogeneous phenomenon. The spline function consists of piecewise polynomials wherein the derivatives are continuous. Hence the spline preserves the continuity of lower-order derivatives across function discontinuities. There are $n+1$ regions defined by $x<x_1$, $x_1\le x<x_2,\ldots,x_n\le x$. At each of the transition points x_i, which are called knots, the

Table 1 Selection of functional forms for model structure determination

Type of phenomenon	Basic form	Improved numerical conditioning
	Polynomial	Orthogonal polynomial
Homogeneous	$y=\sum_{i=0}^{m} C_i x^i$	$y=\sum_{i=0}^{m} c_i p_i(x) \quad 0\le x\le 1$ where $\int_a^b p_m p_n \, dx=\delta_{mn}$
	Spline	*B* Spline
Heterogeneous	$y=\sum_{i=0}^{m} C_i x^i+\sum_{j=1}^{n}\sum_{i=\nu+1}^{m} C_{ij}(x-x_j)^i$	$y_j=\sum_{i=0}^{m+1}\frac{(m+1)(x-x_{i+1})^q}{w_j(x_{i+j})}$ $\nu=m-1$
	where $(x-x_j)^i_+=\begin{cases}0 & x<x_j\\ (x-x_j)^i & x\ge x_j\end{cases}$	$w_j(x)=(x-x_j)(x-x_{j+1})\cdots(x-x_{j+q+1})$

Table 2 General polynomials for function of several variables

Type of phenomenon	Basic form	Hierarchical form
Homogeneous	General polynomial	Multiple subset polynomials
	$y=\sum_{i=0}^{m1}\sum_{j=0}^{m2} C_{ij}x^i(1)\,x^j(2)$	$y=\sum_{i=0}^{m1} C_i x^i(1)$ $C_i=\sum_{j=1}^{m2i} C_{ij}x^j(2)$
Heterogeneous	Polynomial spline	Multiple subset spline
	$y=\sum_{i_1=0}^{m1}\sum_{i_2=0}^{m2} C_{i_1 i_2}x^{i_1}(1)x^{i_2}(2)+\sum_{j_1=1}^{n1}\sum_{j_2=1}^{n2}\sum_{i_1=\nu_1+1}^{m1}\sum_{i_2=\nu_2+1}^{m2} C_{i_1 j_1 i_2 j_2}[x(1)-x_{j_1}(1)]^{i_1}[x(2)-x_{j_2}(s)]^{i_2}+\cdots$	$y=\sum_{i=0}^{m1} C_i x^i(1)+\sum_{j=0}^{m1}\sum_{i=\nu+1}^{m1} C_{ij}[x(1)-x_j(2)]^i_+$ $C=\sum_{i=0}^{m2} d_i x^i(2)+\sum_{j=1}^{m2}\sum_{i=\nu_2+1}^{m2} d_{ij}[x(2)-x_j(2)]^i_+$

first ν derivatives of the function y are continuous. The independent variable y is a linear function of parameters C_i but is a nonlinear function of knot locations. B splines are used for improved numerical conditioning (see Table 1).

The representations for a single independent variable can be generalized to many independent variables in several ways. Straightforward generalizations of the polynomial and spline forms for two variables are given in Table 2. Such an approach is useful for cases where the number of independent variables and/or terms in approximating polynomials is small. For several independent variables, a more organized procedure has been found necessary. The method that has been implemented to achieve this is an extension of Ivakhnenko's group concept.[11] Table 2 shows the typical equations that are used to evaluate large levels of subsets of model terms. For example, the equation of Table 2 (for regular polynomials) is written in terms of the second variable for all of the unknown variables in the first equation.

The models specified using the foregoing procedure will, in general, be too complex. The next section discusses statistical criteria that are used to simplify the preceding representations such that useful models are obtained from a limited set of data.

Quantitative Criteria for Comparison of Competitive Models Against Test Data

The representations of the previous sections are quite flexible and allow for a very large number of model structures. In certain applications, several competing representations based on different independent variables may be hypothesized, leading to a further increase in the number of plausible model structures. The use of measured data isolates the most likely model and indicates model adequacy in explaining the observed behavior. This section discusses several classes of criteria used for this purpose. The tradeoffs to be considered in the particular selection of the criteria are 1) the distribution function of the noise (certain criteria are applicable when the noise is white Gaussian, whereas others are more general), 2) a priori knowledge about the noise distribution function, and 3) number of models to be compared.

Let there be N sets of measurements (or reconstructed values) represented by y_i and $x_i(1)$, $x_i(2)$, ..., $x_i(p)$, $i=1,2,...,N$. Quantitative criteria used for model substantiation based on these data may be divided into four broad categories, shown in Table 3. Note that all of these criteria can be used with both the equation error and the dynamic model formulations.

Fit Error Statistics

Fit error is a measure of the difference between the measured response and its estimate based on the model. Suppose that two models, M_1 (with m_1 parameters θ_1) and M_2 (with m_2 parameters θ_2), are to be compared. Let $\hat{y}_i(M_1,\hat{\theta}_1)$ and $\hat{y}_i(M_2,\hat{\theta}_2)$ be the estimated values of y_i based on models M_1 and M_2, respectively ($\hat{\theta}_1$ and $\hat{\theta}_2$ are the corresponding parameter estimates). Then the fit error leads to the following criterion:

$$\left.\frac{\frac{1}{N}\sum_{i=1}^{N}[y_i-\hat{y}_i(M_1,\hat{\theta}_1)]^2}{\frac{1}{N}\sum_{i=1}^{N}[y_i-\hat{y}_i(M_2,\hat{\theta}_2)]^2}\right\}\begin{matrix}<1 \text{ select model } M_1\\ >1 \text{ select model } M_2\end{matrix} \quad (1)$$

Table 3 Comparison of criteria for validation of models against test data

Class	Equation	Comments
Fit error		
Error covariance	$(1/N)\Sigma(y_i-\hat{y}_i)^2$	Since fit error always increases with number of parameters, subjective termination criteria required
Whiteness test	$(1/N)\Sigma(y_i-\hat{y}_i)(y_{i+1}-\hat{y}_{i+1})$ etc.	
Fit error corrected for degrees of freedom	$[1/(N-m_1)]\Sigma(y_i-\hat{y}_i)^2$	
Likelihood approach		
Likelihood ratio	$p(Y\|M_1,\hat{\theta}_1)/p(Y\|M_2,\hat{\theta}_2)$	Equations given for two models (see text for generalization); requires knowledge of probability distributions, a priori; works with any noise distribution
Log likelihood ratio corrected for degrees of freedom	$\log\{p(Y\|M_1,\hat{\theta}_1)/p(Y\|M_2,\hat{\theta}_2)\}$ $-2m_1+2m_2$	
Prediction error		
Direct determination over an independent data set	$\frac{1}{N_p}\sum_{i=1}^{N_p}(y_i-y_{ip})^2$	Excellent when model is used for prediction; automatically incorporates degrees of freedom
Estimate over same region as data	$\frac{N+m_1}{N-m_1}\cdot\frac{1}{N}\cdot\sum_{i=1}^{N}(y_i-\hat{y}_i)^2$	
Estimate over a new region	$\frac{1}{N_p}\sum_{i=1}^{N_p}[\sigma_i^2+\hat{\sigma}_i^2(M_1)]$	
F ratio		
Equation F ratio	$\frac{R^2/m}{(1-R^2)/(N-m)}$	Most-used test in econometrics and biometrics; easy to implement; excellent for a first cut
Parameter F ratio	$\frac{(R_2^2-R_1^2)/m_2}{(1-R_2^2)/(N-m_1-m_2)}$	
Parzen's test	See ref. 13	

Improved results are obtained if the fit error is corrected for the number of unknown parameters in the model. The adjusted fit error for model M_l is

$$\frac{1}{(N-m_l)} \sum_{i=1}^{N} \{y_i - \hat{y}_i(M_l, \hat{\theta}_l)\}^2 \tag{2}$$

The fit error rarely should be used directly for comparison of models. It is, however, very useful in establishing the validity of the model selected by other approaches.

Likelihood Ratio Statistics

The likelihood approach has been used extensively for both parameter estimation and comparison of competitive models based on test data. The central concept is the likelihood function, which defines the probability that the measured data were generated by any specific model or any set of parameter values. The likelihood function for model M_l is $p(Y|M_l, \hat{\theta}_l)$, where Y is the set of measurements $y_1, y_2, \ldots, y_N$, and $\hat{\theta}_l$ is the maximum likelihood estimate of θ_l assuming that model M_l holds. The model selection criterion is the ratio of the likelihood functions for models M_l and M_2. The likelihood ratio must, however, be corrected for the degrees of freedom. Otherwise the results will always favor more complex models. When there are p competitive models to be compared, the following procedure may be used.

Step 1: compute J_i, $i=1,2,\ldots,p$: $J_i = \ell n\{p(Y|M_i, \hat{\theta}_i)\} - 2m_i$

Step 2: let J_j be such that $J_j > J_i$, $i \neq j$. Then J_j is the "most likely" model. For two candidate models, the method simplifies to

$$\ell n\left\{\frac{p(Y|M_1, \hat{\theta}_1)}{p(Y|M_2, \hat{\theta}_2)}\right\} - 2m_1 + 2m_2 \begin{cases} > 0 \text{ select } M_1 \\ < 0 \text{ select } M_2 \end{cases} \tag{3}$$

Important generalizations of the likelihood ratio result when one model is the subset of the other model, or, in general, all models considered are subsets of the same maximum model. Let the maximal model be characterized by parameters θ. When certain parameters are zero, a lower-order subset is obtained. (Note that this may be used to model a wide variety of situations.) The log likelihood function then may be expanded in a multidimensional Taylor series about our best estimate $\hat{\theta}$ of θ:

$$\ell n\ p(Y|\theta) = \ell n\ p(Y|\hat{\theta}) + \frac{\partial\, \ell n\, p(Y|\hat{\theta})}{\partial \hat{\theta}} (\theta - \hat{\theta}) + \tfrac{1}{2} (\theta - \hat{\theta})^T M (\theta - \hat{\theta}) + \cdots \tag{4}$$

where M is the information matrix. Since $\hat{\theta}$ is the best estimate of θ, the first gradient of the likelihood has mean zero and covariance M^{-1}. To test the hypothesis that a lower-order model is valid, the estimated $\hat{\theta}$ must be compared with M^{-1} to test if setting a component of $\hat{\theta}$ equal to zero will decrease the likelihood function significantly. A model structure determination principle based on this concept was detailed by Gupta.[12] The likelihood method is optimal under a variety of circumstances. Its rigorous applicability is limited by the theoretical requirement to know the probability density of the measurements for each model.

Prediction Error Statistic

The capability of a model to predict system responses for a class of inputs is a desirable quality. Therefore, prediction error of models may be used as the quantitative criterion to select the model that best substantiates the measured data. There are two methods to determine the prediction error of a model. In the first method, the measured data are divided into two parts. The first part of the data is used to estimate unknown parameters in each model. The estimated models then are used to predict the response for the second data set. The model that predicts the response most accurately is the desired model. The second method is indirect, where the prediction error is estimated statistically. The prediction error either is based on the same input as the one used to estimate the models or covers a different region. (The choice depends on the ultimate application of the model.) A good estimate of prediction error for a model M_l with m_l parameters over the data region may be shown to be[7]

$$\frac{N+m_l}{N-m_l} \left\{ \frac{1}{N} \sum_{i=1}^{N} [y_i - \hat{y}_i(M_l, \hat{\theta}_l)]^2 \right\} \tag{5}$$

The prediction over a data region y_{pi}, $i=1,2,\ldots,N_p$ is computed as follows. Let $\hat{y}_{pi}(M_l, \hat{\theta}_l)$ be the estimated value for y_{pi} with mean square error $\hat{\sigma}_i^2(M_l)$ for model M_l. If σ_i^2 is the variance of noise in y_{pi}, the prediction error is

$$\frac{1}{N_p} \sum_{i=1}^{N_p} [\sigma_i^2 + \hat{\sigma}_i^2(M_l)] \tag{6}$$

Prediction error criteria are most useful when the estimated model is to be used for simulation. They automatically incorporate the degree-of-freedom correction for the number of unknown parameters in each model.

***F*-Ratio Statistic**

The F-ratio is perhaps the most widely used statistic for model hypothesis testing, particularly in econometrics and biometrics. The test is based on the assumption of normally distributed random disturbances and requires a priori specification of acceptance-rejection boundary.

These assumptions are restrictive, and, as noted in Ref. 1, F tests should not be used as the only criterion for the adequacy of a model. On the other hand, they do have compensating attributes, which include the following:

1) Standard algorithms have been optimized for computer implementation of F-ratio statistical hypothesis testing. They allow the consideration of an extremely large number of models.

2) A relative maximum of the F-ratio with the number of parameters often is found in practice. This maximum is cited in Ref. 1 as an experimental result, which, to the authors' knowledge, has not been investigated yet by other researchers.

3) For many practical cases, it gives the same result as more sophisticated approaches.

The desirable performance of the F-ratio, in spite of the assumptions, is presumably due to the robustness of the statistic and the particular self-check features used in the implementation. In particular, application of the F test criterion on properly prefiltered data has been found to improve the utility of the statistics. Three approaches have been used based on 1) equation F-ratio, b) parameter F-ratio, and c) Parzen's test. The equation F-ratio tests the validity of the entire model. It is given by

$$F = \frac{R^2/m}{(1-R^2)/(N-m)} \tag{7}$$

where R is the equation multiple correlation coefficient, N is the number of data points, and m is the number of parameters in the equation.

The parameter F-ratio statistic is applied as follows. Suppose that the complete set of parameters θ is divided into two subsets of θ_1 and θ_2 of size m_1 and m_2. An F test can be used to test the hypothesis that θ_2 is equal to any specific value (usually zero), whereas parameters θ_1 are chosen to minimize

fit error. The F-ratio of parameters θ_2 is

$$F(\theta_2) = \frac{(R_2^2) - R_1^2)/m_2}{(1-R_2^2)/(N-m_1-m_2)} \quad (8)$$

where R_1 is the multiple correlation coefficient with parameters θ_1, and R_2 is the multiple correlation coefficient with parameters θ_1 and θ_2. If the F-ratio is small compared to a threshold, parameters θ_2 may be set to zero. F-ratio for parameters in the equation is computed in a similar manner. In practice, the F test is performed on single parameters rather than on sets of parameters.

Parzen[13] devised a unique method based on the F statistic to develop models for a physical system. The method assumes that the "true" model is very complex and has several degrees of freedom. Since the measured data contain noise and do not encompass the entire operation regime, the estimated model must be a simplified version of the true model. Each simplified model then is compared with a high-order model using an F statistic.

Summary

Quantitative criteria presented in this section look at model structure optimality from different viewpoints. Extensive experience has indicated that, for well-behaved systems, all of the tests (except fit error) give similar model structures if the noise has a normal distribution and the specified a priori minimal model includes all of the effects observed in the data. If the distribution of the noise is significantly different from normal or if the noise is not white, the likelihood ratio test gives the best results.

Implementation Considerations

Implementation of methods to develop models from test data must consider the following factors:

1) The number of hypothesized models may be very large (number in billions or more is common).

2) All models contain some level of modeling error (i.e., no model of a physical system is perfect). The models are "good" or "poor," not "right" or "wrong."

3) Many physical systems are dynamic. The modeling of system dynamics may improve the model structure estimates.

4) Distribution functions usually are not known and must be approximated.

5) It should be possible to incorporate the analyst's opinion.

6) Computation time should be reasonable (because of factor 1).

These factors indicate that proper implementation is a key part of the successful model structure determination process. Three formulations have been used successfully: 1) equation error, 2) Kalman filter (or extended Kalman filter), and 3) maximum likelihood (or its special case, output error).

Equation Error Formulation

In the equation error formulation, measurements of dependent and independent variables are related by

$$\bar{y}_i = f(\bar{X}_i, \theta) + \epsilon_i; \qquad i=1,2,\ldots,N \quad (9)$$

$\bar{X}_i$ is the set of p independent variables at the ith point, and θ is the $k\times 1$ vector of unknown parameters. $\bar{y}_i$ and $\bar{X}_i$ are obtained by reconstruction from available measurements. For example, $\bar{y}_i$ may represent angular accelerations obtained by differentiating measurements of angular rates. Such reconstruction often is necessary in equation error formulations. The results shown here are for uncorrelated and Guassian noise, although extension to other cases is straightforward. Equation (9) is linearized about the nominal value θ_0 of the parameter vector

$$\bar{y}_i - f(\bar{X}_i, \theta_0) \approx \frac{\partial f(\bar{X}_i, \theta_0)}{\partial \theta_0}(\theta-\theta_0) + \epsilon_i; \quad i=1,2,\ldots N \quad (10)$$

which may be written as

$$\Delta Y = A\Delta\theta + \epsilon \quad (11)$$

ΔY is an $N\times 1$ vector, and A is an $N\times k$ matrix. Note that Eq. (11) usually will be an overdetermined system ($N>k$). $\Delta\theta$ is chosen to minimize $\|\Delta Y - A\Delta\theta\|^2$.

Nominal parameter values θ_0 are obtained either from a priori estimates or from a previous iteration. Equation (11) must be solved a number of times until convergence occurs. Most nonlinear equations of engineering significance can be solved by iteratively forming and solving a series of linear equations [Eq. (11)].

A straightforward way to evaluate all of the possible models is to solve the complete least-squares problem for each of the parameter subsets. This is unpractical even for systems with 25 unknown parameters (34×10^6 models). One method to make the procedure feasible is to use the stepwise regression procedure.[1] By considering one parameter at a time, only a small fraction of all of the subsets is tested. This allows the analysis of models having up to 400 candidate parameters. One disadvantage of the method is that it usually finds only one subset of each size, unlike the complete search methods.

Kalman Filter

The Kalman filter and extended Kalman filter have been applied to many problems in state estimation, parameter identification, and fault detection. Like the least-squares method, the parameter estimates resulting from a Kalman filter are biased. To use this approach for model structure estimation, one set of dynamic equations is written for each model proposed, i.e., for the ith model,

$$\dot{x}_i = f(M_i, x_i) + w(t); \qquad 0\le t\le T$$
$$y(k) = h(M_i, x_i, k) + v(k); \qquad k=1,2,\ldots,N$$

The states x_i, of course, include the unknown parameters in each model. An extended Kalman filter is developed for these equations. One or more of the following quantities then is used to compare the models.

1) Innovations: The innovation for model i is $y(k) - \hat{y}(k\,|k-1,M_i)$. Its bias and covariance is a good measure of one-step-ahead prediction error. It also could be used to compute certain likelihood ratios by making suitable assumptions. The problem with innovation is that, although the parameters are being adjusted in the initial portion of data, innovations are large, and more complex models may be penalized unnecessarily.

2) Fit error: The fit error for model i is $y(k) - \hat{y}(k\,|N,M_i)$ and can be used like innovations to compare models. Residuals also may be used.

3) Parameter estimates and covariances: This is a poor basis of a test because predicted covariances are often grossly in error when determined using a Kalman filter.

Problems with using a Kalman filter for model structure determination are 1) choice of measurement and process noise covariance have a strong influence on model (in general, increasing process noise covariance will result in less complex models); 2) filter divergence because of poor starting values may invalidate an otherwise good model; and 3) since one extended Kalman filter is required for each model (34×10^6 Kalman filters for a problem with 25 parameters, each of which could be zero), the method often requires unacceptable computation time in practical systems.

Maximum Likelihood Method

Several of the problems associated with an extended Kalman filter may be solved by using the maximum likelihood method. This method has been described extensively[1,2] in the literature. It gives accurate estimates of parameters, as well as associated error covariances. Although all of the quantities

used for model comparison with a Kalman filter may be applied, the likelihood ratio and prediction error tests are most appropriate. In certain cases, log likelihood ratio expansions of the kind shown in the likelihood statistic test section can simplify the implementation significantly.

Examples

Example 1: Advanced Turbofan Engine: Real-Time Model Structure Determination Using a Kalman Filtering Approach

The modern aircraft turbine engine is a sophisticated, multi-input/multi-output device whose aerothermodynamic mathematical models are highly nonlinear. As part of the reliability studies of a digital, quadratic regulator for an advanced turbofan, we used a Kalman filter approach to determine system model structure in a real-time application. Different models were isolated using a likelihood ratio technique, with test statistics formulated from the output of a bank of Kalman filters. (Montgomery[14] has applied a similar technique for fault-detection applications.)

Figure 1 shows the overall configuration of the system with the model determination functions. The technique consists of the following steps (a detailed algorithm may be found in Ref. 14):

1) A series of possible models is hypothesized, and a Kalman filter (or extended Kalman filter for a nonlinear model) for each is constructed.

2) The output of the system is input to each Kalman filter and a likelihood ratio calculated from the filter innovations and innovation covariance.

3) The likelihood corresponding to the incorrect models will go to zero, whereas that corresponding to the correct model will become large.

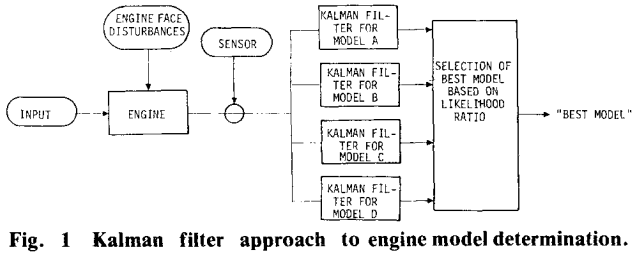

Fig. 1 Kalman filter approach to engine model determination.

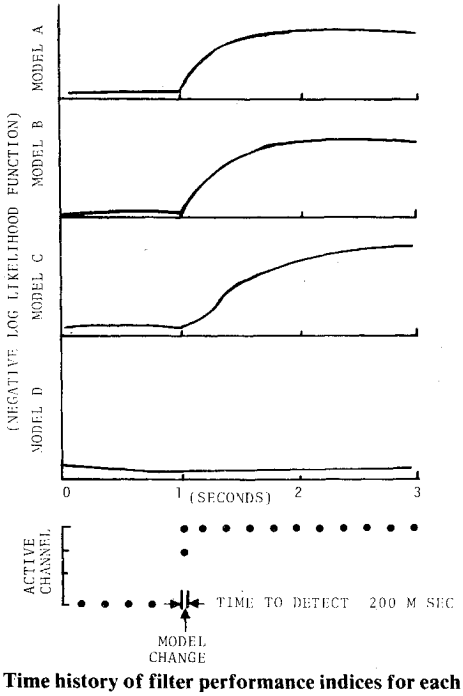

Fig. 2 Time history of filter performance indices for each model.

4) For this application, a filter performance index defined as *1* (likelihood ratio) was used to implement the algorithm in the controller.

Figure 2 shows the simulated filter performance index, indicating that a model change occurred and was detected 200 ms later (time delay of the Kalman filter). The simulation used is a complete nonlinear representation of the particular engine being analyzed.

This example illustrates the tradeoff between computational time and model complexity. Since it was desired to have an on-line, real-time model structure determination algorithm, for which the number of possible models was known to be small, the Kalman filter approach was considered. Indeed, the approach worked well on a complex nonlinear digital simulation. It must be noted, however, that there is a practical limit to the number of models which can be used in the algorithm, and that the Kalman filter itself is highly susceptible to numerical divergences when the assumed process and measurement statistics are chosen incorrectly. Utilization of approximate nonlinear update equations with piecewise constant Kalman gain reduces this problem in the engine application. However, this approach is not considered realistic for the type of model structure determination where a large number of models must be postulated.

Example 2: Submersible Vehicle Pitch Moment Equation

A proposed structure for an equation representing the hydrodynamic pitch moment on a submersible vehicle is

$$MY = (\rho/2)\{\ell^5 M_{rp} rp + \ell^4 M_q uq + \ell^3 [M_{\cdot} u^2 + M_w uw + M_{w|w|}|w|(v^2+w^2)^{1/2} + M_{\delta s} u^2 \delta s + M_{i|w|} u|w| + M_{ww} w(v^2+w^2)^{1/2}] + C_L I_1(v,w,q,r) + C_d I_2[v,r,v(t-\tau)]\}$$

This structure is hypothesized from considerations of high Reynolds number flow theory. I_1 represents the integral over the vehicle hull of the pressure field due to the leeward side vortex flow. I_2 represents the integrated effects of the circulation generated by lateral motions of the fairwater structure.

We wish to determine the relative significance of the 10 candidate dimensionless hydrodynamic parameters during a turning maneuver. We first perform a digital six-degree-of-freedom simulation of the maneuver, recording the state and state derivative time histories. We use the recorded time histories to calculate, at each sampling time, the total pitch moment *MY* (*MY* is calculated using only rigid body kinematic relations) and each of the unknown parameters on the right-hand side. A hierarchical approach determines the best model having a single parameter, the best model having two parameters, and so on, to the model having all 10 parameters. The method implicitly examines 1024 distinct models during this process but requires less than 1.0 s of computer time on a Univac 1108 computer. The example uses data generated by a digital simulation of a full-scale vehicle.

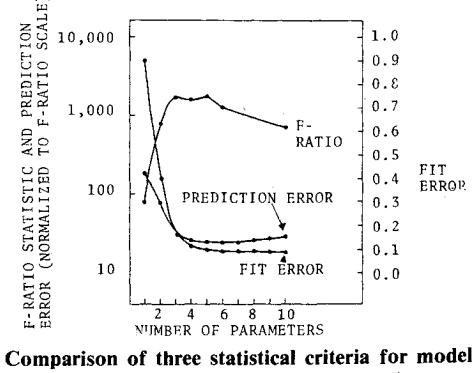

Fig. 3 Comparison of three statistical criteria for model structure determination (using hierarchical polynomial model).

Table 4 MY equation parameters for five-and six parameter models

Five-parameter model	Six-parameter model
C_L	C_L
M_{rp}	M_{rp}
M_q	M_q
M_w	M_w
M_{δ_s}	M_{δ_s}
	C_d

Table 5 Selection of a priori models to describe flowfield around a slender body of revolution

Phenomenological information
Selection of axis system
Dependence of wind-frame moment on angle of attack
Erratic changes in moments with Reynolds number
Stochastic phenomenon
Empirical model
$C_m = C_{m_0} + C_{m_\alpha}\alpha + C_{m_{\alpha^2}}\alpha^2$
$C_{m_0} = \sum_{i=1}^{m_0} \{a_{i0}(t-t_i)_+^0 + a_{i1}(t-t_i)_+\}$
Similar equations for C_{m_α} and $C_{m_{\alpha^2}}$

Table 6 Final estimates of parameters for slender body of hypersonic velocity

Parameter	Estimated value (normalized)
C_{m_0}	− 0.22
C_{m_α}	−17.48
Ramp in C_{m_α} at 0.37	2.68
Jump in C_{m_0} at 2.25	0.454
Jump in C_{m_0} at 2.45	− 0.244
Jump in C_{m_0} at 3.20	0.568

The technique also has been applied to data from tank tests of a free-running model vehicle.

Figure 3 shows plots of three test statistics vs the number of parameters in the model: 1) normalized fit error, 2) F-ratio, and 3) prediction error. The evaluation of these statistics assumes that the total hydrodynamic moment can be measured experimentally to within only 10% accuracy.

Table 4 gives the parameters selected for the five-variable model (recommended by the F statistic) and for the six-variable model (recommended by the prediction error statistic). Subsequent processing of these data with the maximum likelihood algorithm further validated the adequacy of this model.

Example 3: Unstable Flow over a Slender Body of Revolution

This example deals with the estimation of a mathematical model for flow characteristics on a slender spinning body of revolution at hypersonic speeds. The model is based on actual test data. The physics associated with this unstable phenomenon are complex. The aerodynamic effects are known to be stochastic. Several hypotheses have been proposed to explain the phenomenon of boundary-layer transition which occurs when Reynolds number crosses a certain critical value and the flow over a body becomes turbulent.

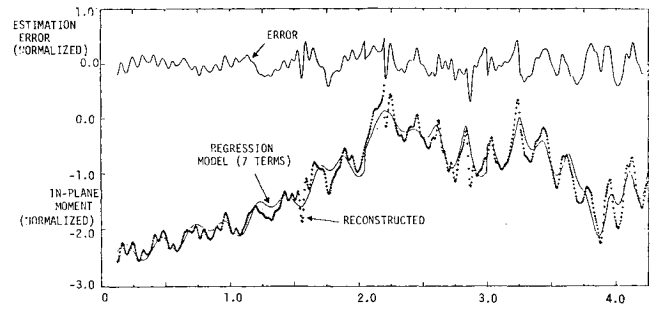

Fig. 4 Reconstructed and estimated in-plane moment.

The relevant axis system in this case is wind-fixed; i.e., the z axis always lies in the plane containing the relative wind vector and the vehicle x axis. Since the data consist of the region prior to the boundary-layer transition and during transition, the spline models are most appropriate. The spline model allows jumps and ramp changes in moment and its derivative with respect to angle of attack. Table 5 summarizes the a priori model selection.

Only the model for the wind in-plane moment will be presented here. The stochastic nature of the body environment indicates that the process noise is very high. The vehicle has excellent instrumentation, resulting in small measurement noise. Because of the predominance of the state uncertainty, the equation error method can be used not only for model structure estimation but also for parameter identification. The following procedure is used to derive the model:

1) Knots are placed initially every 0.5 s in the spline function.

2) The pitching moment time history is reconstructed from available measurements of rates and accelerations.

3) The equation error formulation with equation F-ratio criteria gives a model with six terms: a) C_{m_0}, b) C_{m_α}, c) ramp change in C_{m_α} at 0.5, d) jump in C_{m_0} at 2, e) jump in C_{m_0} at 2.5, and f) jump in C_{m_0} at 3.5 s.

4) Ten knots are placed around each of the identified knots and the equation error formulation reapplied, giving the model of Table 6. Note the changes in knot locations.

5) Knot locations were iterated using a gradient procedure. Final knot locations were the same as in Table 6, with the same variation explained.

The gradient procedure converged in two iterations. (In the second iteration, the improvement in fit was less than 0.1%). Figure 4 shows a comparison of the reconstructed C_m and estimated C_m.

Summary

A comprehensive advanced model structure determination for dynamic systems is an essential aspect of system identification for aerospace systems. The critical elements of such a procedure are 1) specification of functional forms for useful models, 2) criteria for selecting an adequate model against competing models, and 3) efficient numerical algorithms for integrating the conflicting requirements of adequate functional forms and accurate criteria.

Polynomials are often an appropriate choice for functional form. Physical considerations, however, always should be included in this initial specification. Within the class of polynomials, further optimization of functional form is provided by the utilization of splines and various orthogonalization procedures.

Statistical criteria for model selection must include provision for incorporating the effects of reduced degrees of freedom as more parameters are added. Criteria should be selected for both theoretical and practical evaluation op-

timality. More complex criteria should be used when computationally feasible, but simpler criteria should not be disregarded if an adequate model is, in fact, to be selected from many possible models.

Numerical procedures can be either off-line or on-line, depending on the application. Equation error, Kalman filter, and maximum likelihood methods are well suited for many practical vehicles, and examples for each are given.

Acknowledgments

This work has been sponsored by the Office of Naval Research under Contract N00014-76-C-0420. David S. Siegel is the ONR Scientific Officer.

References

[1]Hall, W.E., Gupta, N.K., and Smith, R.G., "Identification of Aircraft Stability and Control Derivatives for the High Angle-of-Attack Regime," Systems Control, Inc., TR 2, for Office of Naval Research under Contract N00014-72-C-0328, 1974.

[2]Hall, W.E., Gupta, N.K., and Tyler, J.S., "Model Structure Determination and Parameter Identification for Nonlinear Aerodynamic Regimes," *AGARD Specialist Meeting on Methods for Aircraft State and Parameter Identification,* NASA, Nov. 1974.

[3]Hall, W.E. and Gupta, N.K., "System Identification for Nonlinear Aerodynamic Flight Regimes," *Journal of Spacecraft and Rockets,* Vol. 14, Feb. 1977, pp. 73-80.

[4]Gupta, N.K. and Hall, W.E., "Model Structure Determination and Test Input Selection for Identification of Nonlinear Regimes," Office of Naval Research, Rept. ONR-CR215-213-5, Feb. 1976.

[5]Thiel, H., *Principles of Econometrics,* Wiley, New York, 1971.

[6]Lee, R.C.K., *Optimal Estimation, Identification, and Control,* MIT Research Monograph 28, MIT Press, Cambridge, Mass., 1964.

[7]Akaike, H., "Statistical Predictor Identification," *Annals of the Institute of Statistics,* Vol. 20, Feb. 1970, pp. 203-217.

[8]Kullback, S., *Information Theory and Statistics,* Wiley, New York, 1959.

[9]Gerlach, O.H., "The Determination of Performance and Stability Derivatives from Dynamic Maneuvers," Delft Univ. of Technology, Netherlands, Rept. VTH-163, March 1971.

[10]Hamming, R., *Numerical Methods for Engineers and Scientists,* McGraw-Hill, New York, 1962, p. 249.

[11]Ivakhnenko, A.G., "Polynomial Theory of Complex Systems," *IEEE Transactions on Systems, Man, and Cybernetics,* Vol. SMC-1, Oct. 1971, pp. 364-378.

[12]Gupta, N.K., "Bias and Mean Square Error Properties of General Estimators," *Decision and Control Conference,* Clearwater Beach, Fla., 1976.

[13]Parzen, E., "A New Look at the Statistical Model Identification," *IEEE Transactions on Automation and Control,* Vol. AC-19, Dec. 1974, pp. 723-729.

[14]Montgomery, R.K. and Price, D.B., "Failure Accommodation in Digital Flight Control Systems Accounting for Nonlinear Aircraft Dynamics," *Journal of Aircraft,* Vol. 13, Feb. 1976, pp. 76-82.