JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 15, No. 4, July–August 1992

Parameter Identification of Linear Systems Based on Smoothing

M. Idan* and A. E. Bryson†
Stanford University, Stanford, California 94305

A parameter identification algorithm for linear systems is presented. It is based on smoothing test data with successively improved sets of system model parameters. The smoothing pass through the data provides all of the information needed to compute the gradients of the smoothing performance measure with respect to the parameters. The parameters are updated using a quasi-Newton procedure, until convergence is achieved. The advantage of this algorithm over standard maximum likelihood identification algorithms is the computational savings in calculating the gradients. This approach is extended to identify one set of parameters from several test runs. The algorithm is compared to other time-domain algorithms using a simple example with computer simulated data. The performance of this algorithm is demonstrated in identifying the parameters of a linear model describing the rigid body dynamics of the DLR BO-105 research helicopter from flight test data. The identification results are presented and compared to recently published models using maximum likelihood and frequency-domain algorithm. The identified models are in good agreement with each other.

Nomenclature

D,H	= measurement equation matrices
J,J_j	= performance measure, single maneuver
J_T	= total multimaneuver performance measure
M	= number of maneuvers
P_0	= initial conditions weighting matrix
Q	= process noise weighting matrix
R	= output error weighting matrix
u	= known input vector, dimension c
v	= measurement noise vector, dimension m
w	= unknown input, the process noise vector, dimension q
x	= state vector, dimension n
x_0	= a priori estimate of the state initial conditions $x(0)$
z	= measurement vector, dimension m
z_m	= experimental data, actual measurements
α_j	= maneuver weighting factor
Γ_u,Γ_w	= discrete control matrices
Θ	= parameters weighting matrix
θ	= vector of model parameters, dimension p
θ_0	= a priori estimate of the parameters θ
λ	= Lagrange multiplier
ν	= difference between the measured outputs and the model outputs, $z_m - z$
Φ	= discrete state dynamics matrix

I. Introduction

THE parameter identification task can be defined as follows: Given a system model that contains unknown constant parameters, determine these parameters so that the response of the model to a given set of inputs best fits the experimental system response to the same input signals. The quality of the fit is measured by a scalar function of the difference between the experimental output data and the model output, called the performance measure. A commonly used performance measure is a weighted least squares function of this difference.

Parameter identification algorithms can be categorized in many ways. The main distinction is between time- and frequency-domain algorithms. In this paper, mainly time-domain algorithms are discussed and compared. However, when the proposed time-domain identification algorithm is tested, it will be compared to results obtained from both time- and frequency-domain algorithms.

Maximum likelihood estimation has become a standard approach for parameter identification in the presence of random signals in the system, widely used in many fields of engineering, e.g., aircraft dynamic modeling from flight test data.[1-3] The resulting algorithms can be divided into three classes. Equation error algorithms can be used when noise-free measurements of all of the states are available and process noise (unknown random inputs) are the only random signals present in the system. These are the simplest maximum likelihood algorithms, often referred to as regression algorithms. Output error algorithms can be used when the process noise is negligible but the measurement noise is not.

In many cases, both the process noise and the measurement errors have to be accounted for while modeling the actual system. For example, a ship is moved by the unknown motions of the sea, a ground vehicle vibrates because of the nonsmooth terrain it is moving on, and an aircraft is disturbed by the turbulent air. Even for flight through smooth air, it is often advisable to include process noise to account for modeling errors.[1]

To reduce the effects of random signals on the identification, filtering and/or smoothing techniques are used. In fact, it has been shown[4-6] that these algorithms are equivalent. When process noise can be neglected, the performance measures are actually equal. The main differences among the algorithms lie in the optimization processes associated with the different approaches. The majority of maximum likelihood identification algorithms are based on filtering techniques[1-3,7,8]; however, there is a growing interest in algorithms that combine filtering and smoothing[9-13] and algorithms based completely on smoothing ideas.[14-17]

The commonly used parameter identification algorithms can be divided into two main groups that are characterized by the way they treat the unknown parameters:

1) Algorithms that identify the system parameters by computing the sensitivities or gradients of the performance meas-

Presented as Paper 90-2800 at the AIAA Atmospheric Flight Mechanics Conference, Portland, OR, Aug. 20–22, 1990; received Nov. 20, 1990; revision received July 1, 1991; accepted for publication Aug. 1, 1991. Copyright © 1991 by the American Institute of Aeronautics and Astronautics, Inc. All rights reserved.

*Graduate Student, Department of Aeronautics and Astronautics; currently Lecturer, Lady Davis Fellow, Technion, Israel Institute of Technology, Haifa, Israel. Member AIAA.

†Professor, Department of Aeronautics and Astronautics. Fellow AIAA.

ure to changes in the parameters. Using filtering techniques to compute the sensitivity functions requires one additional filtering pass through the data for each parameter that is identified. Thus, to identify p parameters (or initial conditions), the experimental data are filtered $p + 1$ times each iteration.

2) Algorithms that treat the parameters as additional constant states of the system, resulting in a higher-order system model that has no unknown parameters. Usually, the new system model is nonlinear because in most engineering applications the new states (the parameters) are related nonlinearly to the original states: e.g., in linear systems, the parameters appear in the system matrices that multiply the states. Also, the problem to be solved is of a much higher dimension when many parameters have to be identified. Thus, the identification problem is converted here into a nonlinear filtering or smoothing problem.

A combined smoothing and identification algorithm was presented by Cox and Bryson.[16] In this approach, the experimental data are smoothed using successively improved sets of system parameters. As part of the smoothing solution, the adjoint variables or the sensitivities of the performance measure to the states of the system are computed. These sensitivities are used to compute the gradients of the performance measure with respect to the unknown parameters. The parameters are updated using a quasi-Newton gradient algorithm. The data are then smoothed with the new set of parameters. This procedure is repeated until a minimum of the performance measure is obtained.

The main advantage of this algorithm is computational. Using the standard maximum likelihood approach, the computations of the sensitivity functions required to compute the gradient of the performance measure with respect to the parameters involve one additional filtering pass for each parameter.[8] Using the smoothing approach, only one smoothing pass through the data is performed to obtain all of the information required to compute the gradients for any number of parameters. The computations involved to determine the gradients of the performance measure with respect to the parameters are similar in both cases.

The advantage of the smoothing identification algorithm compared to the extended model, which treats the parameters as additional states and results in a nonlinear filter or smoother problem, is that the smoother implemented here is of the same dimension as the original system. The nonlinear filtering or smoothing problem of large dimension is replaced here by smoothing of a smaller system together with a quasi-Newton update for the parameters.

In this paper, the parameter identification algorithm presented by Cox and Bryson[16] is expanded and evaluated using actual flight test data. It is often required to identify one set of system parameters from several different sequences of experimental data. This greatly improves the identification quality.[2] For example, during flight tests, for a given flight condition, maneuvers with different inputs are performed with data being recorded for each of them separately, i.e., the data are not continuous. However, one model is required to describe all of these maneuvers since the flight condition is the same for all of them. The proposed algorithm is extended to treat such data in a consistent manner, without any ad hoc procedures.

In Sec. II, the smoothing identification algorithm for linear systems is derived. This smoothing identification algorithm is compared to other time-domain identification procedures using a simple example in Sec. III. A multimaneuver version of this procedure is then presented in Sec. IV. This algorithm was used to identify the six-degree-of-freedom rigid-body model of the BO-105 helicopter. The identification results are presented in Sec. V and compared to recently published rigid-body models for the BO-105 using maximum likelihood and frequency-domain techniques.[18,19] The good agreement among the three models provides a validation check of the suggested identification algorithm. Some concluding remarks are given in Sec. VI.

II. Identification Algorithm

Identification is the process of selecting parameters of a model to minimize some measure of the difference between the measured outputs and the model outputs. A linear system model can be described by state space and measurement equations of the form:

$$\begin{aligned} x(i+1) &= \Phi(i,\theta)x(i) + \Gamma_u(i,\theta)u(i) + \Gamma_w(i,\theta)w(i) \\ z(i) &= H(i,\theta)x(i) + D(i,\theta)u(i) + v(i) \end{aligned} \tag{1}$$

The performance measure is usually a weighted-least-squares error. The identification task is to find the parameters θ, initial conditions $x(0)$, and the sequence $[w(i)]$, $i = 0,\ldots,N-1$ that minimize this performance measure:

$$J = \frac{1}{2}(\theta-\theta_0)^T\Theta^{-1}(\theta-\theta_0) + \frac{1}{2}[x(0)-x_0]^T P_0^{-1}[x(0)-x_0] + \frac{1}{2}\sum_{i=0}^{N-1}[w^T(i)Q^{-1}w(i) + \nu^T(i+1)R^{-1}\nu(i+1)] \tag{2}$$

The first term in Eq. (2) allows for the inclusion of prior knowledge of the parameters. For example, in the case of aircraft model identification from flight test data, the prior information can include wind-tunnel data; ν is the difference between the measured outputs and the model outputs, $z_m - z$.

The minimization is subject to constraint equations (1). To satisfy these equations, the performance measure J is modified by adjoining the constraints using Lagrange multipliers $\lambda(i+1)$:

$$\bar{J} = J + \sum_{i=0}^{N-1}\lambda^T(i+1)[\Phi(i,\theta)x(i) + \Gamma_u(i,\theta)u(i) + \Gamma_w(i,\theta)w(i) - x(i+1)] \tag{3}$$

Minimizing $\bar{J}$ is equivalent to minimizing J subject to constraints (1).

Variational techniques are used to minimize $\bar{J}$. The first variation of $\bar{J}$ due to small changes of the unknowns of the problem is given by

$$\delta\bar{J} = \sum_{i=0}^{N}\frac{\partial\bar{J}}{\partial x(i)}\delta x(i) + \sum_{i=0}^{N-1}\frac{\partial\bar{J}}{\partial w(i)}\delta w(i) + \frac{\partial\bar{J}}{\partial\theta}\delta\theta \tag{4}$$

The relations of the first term in Eq. (4) are

$$\frac{\partial\bar{J}}{\partial x(i)} = \begin{cases} [x(0)-x_0]^T P_0^{-1} + \lambda^T(1)\Phi(0,\theta) & \text{for } i = 0 \\ -\nu^{-T}(i)R^{-1}H(i,\theta) + \lambda^T(i+1)\Phi(i,\theta) - \lambda^T(i) & \text{for } i = 1,\ldots,N \end{cases} \tag{5}$$

where we define $\lambda(N+1) = 0$. The relations of the second term in Eq. (4) are

$$\frac{\partial\bar{J}}{\partial w(i)} = w^T(i)Q^{-1} + \lambda^T(i+1)\Gamma_w(i,\theta), \qquad \text{for } i = 0,\ldots,N-1 \tag{6}$$

The third term in Eq. (4) depends on the functional relationships among the system matrices and the parameters. In general, all of the matrices can be a function of all of the parameters, however, it is uncommon in practice. In the case of the linear rigid-body aircraft model, each parameter (i.e., stability or control derivative) is an entry in one of the system matrices and thus affects only that matrix. Thus, depending on which matrix the parameter is in, the relations of the third term in Eq. (4) are

$$\frac{\partial \bar{J}}{\partial \theta(i)} = \sum_{j=1}^{p} \frac{\theta(j) - \theta_0(j)}{\Theta(i,j)} + \begin{cases} -\sum_{k=1}^{N} \nu^T(k) R^{-1} \frac{\partial H(k,\theta)}{\partial \theta(i)} x(k) \\ -\sum_{k=1}^{N} \nu^T(k) R^{-1} \frac{\partial D(k,\theta)}{\partial \theta(i)} u(k) \\ \sum_{k=0}^{N} \lambda^T(k+1) \frac{\partial \Phi(k,\theta)}{\partial \theta(i)} x(k) \\ \sum_{k=0}^{N-1} \lambda^T(k+1) \frac{\partial \Gamma_u(k,\theta)}{\partial \theta(i)} u(k) \\ \sum_{k=0}^{N-1} \lambda^T(k+1) \frac{\partial \Gamma_w(k,\theta)}{\partial \theta(i)} w(k) \end{cases} \quad \text{for } i = 1,\ldots,p \tag{7}$$

In the current work, the initial Hessian matrix is set to the identity matrix, causing the first iteration to be in the steepest descent direction.

At a stationary point of $\bar{J}$, the first variation $\delta\bar{J}$ vanishes for arbitrary variations of $\delta x(i)$, $\delta w(i)$, and $\delta\theta$, and so the coefficients of each one of the three terms in Eq. (4) must vanish. For a given set of parameters θ, setting Eqs. (5) and (6) to zero and incorporating the constraint equations (1) leads to a smoothing problem described by

$$x(i+1) = \Phi(i,\theta)x(i) + \Gamma_u(i,\theta)u(i) + \Gamma_w(i,\theta)w(i)$$
$$\lambda(i) = \Phi^T(i,\theta)\lambda(i+1) - H^T(i,\theta)R^{-1}\nu(i) \tag{8}$$
$$w(i) = -Q\Gamma_w^T(i,\theta)\lambda(i+1)$$

and the boundary conditions:

$$x(0) = x_0 - P_0\Phi(0,\theta)^T\lambda(1), \qquad \lambda(N+1) = 0 \tag{9}$$

This is a standard linear two-point boundary-value problem (LTPBVP) and it can be solved using one of the well-known smoothing algorithms, e.g., forward covariance filter plus backward smoother or backward information filter plus forward smoother. The latter is advocated in Refs. 15 and 20 and is attractive when there are no good estimates of the initial and final conditions of the state vector. This algorithm was implemented in this work (see Appendix A).

Once the smoothing problem is solved, if follows from Eq. (4) that

$$\delta\bar{J} = \frac{\partial \bar{J}}{\partial \theta}\delta\theta \equiv \bar{J}_\theta \delta\theta$$

which also has to vanish at the stationary point. In general, this will not be the case. Thus, using the solution of the smoothing problem, the gradient of $\bar{J}$ with respect to the parameters θ, $\bar{J}_\theta$ is evaluated from Eq. (7). It is important to note that once the smoothing problem of Eqs. (8) and (9) is solved, the solution contains all of the information required to compute the gradient of $\bar{J}$ for *any* number of parameters. This is an important characteristic of this algorithm.

The parameters θ are updated using a quasi-Newton procedure to minimize $\bar{J}$:

$$\theta_{\text{new}} = \theta_{\text{old}} - (\bar{J}_{\theta\theta})^{-1}\bar{J}_\theta \tag{10}$$

The inverse of the Hessian matrix $(\bar{J}_{\theta\theta})^{-1}$ in Eq. (10) is estimated numerically from successive values of the gradient vector $\bar{J}_\theta$ using a rank-two update procedure[21,22] (see Appendix B). This quasi-Newton method has good numerical characteristics even when inaccurate line search procedures are performed to minimize $\bar{J}$ along a search direction.[21] The initial value of $(\bar{J}_{\theta\theta})^{-1}$ can be chosen as any positive definite matrix.

An outline of the algorithm is the following:

1) With initial estimates of $x(0) = x_0$ and a set of parameters θ obtained from the preceding iteration (or an initial guess θ_0), compute the smoothed states and the forcing functions [Eqs. (8) and (9)] and evaluate the performance measure J using Eq. (2).

2) Evaluate the gradient of $\bar{J}$ with respect to the parameters, using Eq. (7).

3) Update the parameters θ using a quasi-Newton procedure, where the inverse of the Hessian of $\bar{J}$ is estimated using a rank-two update algorithm.

4) Repeat until the changes in θ are small and the performance measure is minimized.

III. Comparison to Existing Time-Domain Algorithms

In this section, we compare our smoothing identification algorithm with existing maximum likelihood and nonlinear smoothing algorithms. The differences among these approaches are discussed. Then the algorithms are compared using computer simulated data for a first-order system with two unknown parameters.

Comparison to Maximum Likelihood Algorithms

The maximum likelihood performance measure is given by

$$J_{\text{ml}} = \frac{1}{2}\sum_{i=1}^{N} \nu^T(i)\bar{R}^{-1}\nu(i) + \frac{1}{2}N \ln |\bar{R}| \tag{11}$$

where $\bar{R}$ is the innovation covariance matrix. Although the performance measures are closely related and the identification results are similar, the maximum likelihood and the smoothing identification algorithms are not exactly the same. The gradient procedures used to update the parameters are different, which leads to different intermediate values of the parameters at each iteration. Thus, the computations involved in determining the gradients of the performance measure with respect to the parameters, which is the main part of the algorithms, are compared. A comparison of the overall performance of the algorithms is then presented using a first-order example.

Maximum likelihood algorithms based on filtering techniques update the parameters by computing output sensitivity functions. The sensitivity functions are the gradients of the innovation signal (the difference between the experimental data and the a priori estimates of these measurements) with respect to the unknown parameters. The dimension of each sensitivity function is the dimension of the measurement vector. These sensitivity functions are used to compute the gradient of the performance measure with respect to the parameters. Thus, for identifying p parameters or unknown initial conditions, p sensitivity vector functions must be computed.

The sensitivity functions can be computed by propagating an analytically derived set of recursive equations[1] or by computing a numerical approximation of these functions (numerical perturbation method[8]). The exact analytical formulation is complicated, especially for time-varying filters. When a steady-state filter is used, some approximations can be made to construct the analytical recursive set of equations to compute the sensitivity functions. In both analytical and numerical perturbation implementations, the computational load associated with the computation of each sensitivity function is equivalent to that of the nominal filtering of the data.[1,8] Thus, at each iteration, $p + 1$ filtering passes though the data are performed.

The gradients of the maximum likelihood performance measure with respect to the parameters $\nabla_\theta J_{\text{ml}}$ are then computed by

$$\nabla_{\theta_j} J_{\text{ml}} = \sum_{i=1}^{N} \nu^T(i) \bar{R}^{-1} \nabla_{\theta_j} \nu(i) \tag{12}$$

where $[\nabla_{\theta_j} \nu(i)]$ is the sensitivity function of the innovation sequence $[\nu(i)]$ associated with the parameter θ_j. The innovation covariance matrix $\bar{R}$ is estimated by[1]

$$\bar{R} = \frac{1}{N} \sum_{i=1}^{N} \nu(i) \nu^T(i) \tag{13}$$

Using the sensitivity functions, the second gradient matrix, the Hessian, of the performance measure J_{ml} is well approximated by

$$\nabla_{\theta_i \theta_j} J_{\text{ml}} = \sum_{i=1}^{N} \nabla_{\theta_i} \nu^T(i) \bar{R}^{-1} \nabla_{\theta_j} \nu(i) \tag{14}$$

In our smoothing identification algorithm, the gradient of the performance measure is computed using the smoothed states and the Lagrange multipliers, where the latter can be interpreted as sensitivities of the performance measure to variations of the state vector. Compared to filtering, smoothing requires one pass through the data, which computationally is less intensive than the filtering pass, especially if a time-varying smoother is used (see Appendix A). If a steady-state smoother is used, the smoothing pass is computationally comparable to the filtering pass. However, only one smoothing pass is performed for any number of identified parameters.

The gradients of the performance measure with respect to the parameters are computed using Eqs. (7). For linear systems, each parameter usually affects only one system matrix. Thus, the gradients of the performance measure with respect to the parameters of a linear system $\nabla_{\theta_j} J_s$ are of the form:

$$\nabla_{\theta_j} J_s = \sum_{i=0}^{N-1} \lambda^T(i+1) \frac{\partial \Phi(i,\theta)}{\partial \theta_j} x(i)$$

or

$$\nabla_{\theta_j} J_s = -\sum_{i=1}^{N} \nu^T(i) R^{-1} E \frac{\partial H(i,\theta)}{\partial \theta_j} x(i) \tag{15}$$

In Eqs. (15), $[x(i)]$ and $[w(i)]$ are the smoothed states and process noises, $[\nu(i)]$ are the differences between the experimental data and the smoothed estimates, and $[\lambda(i)]$ are the Lagrange multipliers.

Comparing Eq. (12) to Eqs. (15), we see that gradient computations are comparable for the two methods after the sensitivity functions and the smoothed estimates are computed. However, in the first case, an equivalent of filtering is performed to compute each sensitivity function for *each* parameter, whereas in the smoothing case, only *one* smoothing pass is performed for *any number* of parameters. Thus, overall, the smoothing identification algorithm shows considerable computational savings, especially when many parameters are identified. This will be now demonstrated by a discrete first-order example.

A discrete first-order system with two unknown parameters will be used to demonstrate the differences between computing the gradients of the performance measure to the parameters using maximum likelihood and the smoothing identification algorithms. The system state and measurement equations are

$$x(i+1) = ax(i) + u(i) + w(i) \tag{16}$$

$$z(i) = hx(i) + v(i) \tag{17}$$

where x is the state of the system, u the known input, w a white zero mean Gaussian process noise with covariance q, z the measurement, and v a white zero mean Gaussian measurement noise with covariance r. Both q and r are assumed to be known; a and h are the system parameters to be identified.

For simplicity, a steady-state filter will be used for both the maximum likelihood and the smoothing identification algorithms. The filter parameters are computed by solving a first-order Riccati equation:

$$(ah)^2 P^2 + (qh^2 - ra^2 + r)P - qr = 0 \tag{18}$$

The Kalman filter gain is then given by:

$$K = \frac{h}{r} P \tag{19}$$

For known covariances q and r, the innovation covariance $\bar{r}$ can be computed by

$$\bar{r} = r + h^2(a^2 P + q) \tag{20}$$

Computing the gradient of the performance with respect to the parameters using the maximum likelihood approach involves the following:

1) Propagating the filter equation:

$$\bar{x}(i+1) = a_f \bar{x}(i) + u(i) + aKz(i+1) \tag{21}$$

where $a_f = a(1 - hK)$, and $[\bar{x}(i)]$ is the sequence of the a priori estimates of the state.

2) With the initial conditions $\bar{x}_a(0) = 0$ and $\bar{x}_h(0) = 0$, propagating the sensitivity functions:

$$\bar{x}_a(i+1) = a_f \bar{x}_a(i) + \frac{\partial a_f}{\partial a} \bar{x}(i) + \left(K + a\frac{\partial K}{\partial a}\right) z(i+1)$$
$$\nabla_a \nu(i) = -h\bar{x}_a(i) \tag{22}$$

$$\bar{x}_h(i+1) = a_f \bar{x}_h(i) + \frac{\partial a_f}{\partial h} \bar{x}(i) + a\frac{\partial K}{\partial h} z(i+1)$$
$$\nabla_h \nu(i) = -[\bar{x}(i) + h\bar{x}_h(i)] \tag{23}$$

Note that for this scalar case the partial derivatives in Eqs. (22) and (23) can be easily determined analytically from Eqs. (18) and (19).

3) Computing the gradients

$$\nabla_a J_{\text{ml}} = \frac{1}{\bar{r}} \sum_{i=1}^{N} \nu(i) \nabla_a \nu(i) \tag{24}$$

$$\nabla_h J_{\text{ml}} = \frac{1}{\bar{r}} \sum_{i=1}^{N} \nu(i) \nabla_h \nu(i) \tag{25}$$

where

$$\nu(i) = z(i) - h\bar{x}(i)$$

and $\bar{r}$ is given by Eq. (20).

Computing the gradient of the performance measure with respect to the parameters using the smoothing identification algorithm involves the following:

1) Propagating the filter equation:

$$\hat{x}(i+1) = a_f\hat{x}(i) + (1 - hK)u(i) + Kz(i+1) \tag{26}$$

where $[\hat{x}(i)]$ is the sequence of the a posteriori estimates of the state.

2) Propagating the smoothing equation backward:

$$\lambda(i) = a_f\lambda(i+1) + \frac{h^2}{r}\hat{x}(i) - \frac{h}{r}z(i), \qquad \lambda(N+1) = 0 \tag{27}$$

3) Computing the gradients:

$$\nabla_a J_S = \sum_{i=0}^{N-1} \lambda(i+1)x_s(i) \tag{28}$$

$$\nabla_h J_S = -\frac{1}{r}\sum_{i=1}^{N} \nu_s(i)x_s(i) \tag{29}$$

where

$$x_s(i) = \hat{x}(i) - aP\lambda(i+1) \qquad \text{smoothed state}$$

$$\nu_s(i) = z(i) - hx_s(i)$$

The first and third steps in the two algorithms are almost the same. The main difference between the algorithms is in step 2. In the maximum likelihood procedure, two sensitivity functions have to be propagated, Eqs. (22) and (23). If more parameters were to be identified, more sensitivity functions would have to be propagated. In the smoothing identification algorithm, these recursive equations are replaced by one smoothing recursion, Eq. (27), and that would be the only equation for this example even if more parameters were to be identified.

The actual performance comparison for this example with computer simulated data will be presented at the end of this section.

Comparison to Nonlinear Filtering and Nonlinear Smoothing Algorithms

The use of an extended model that includes the parameters as additional states involves a nonlinear filtering or a nonlinear smoothing problem of a dimension much larger than the dimension of the original system. The noniterative, recursive extended Kalman filter (EKF) can be used in this case. However, as noted in Ref. 1, the EKF addresses a different problem than the one posed for off-line identification. It is designed mainly for on-line applications. In general, it leads to estimates that are inferior to the estimates obtained using maximum likelihood or iterative filtering/smoothing techniques. If on-line estimates are required, this degradation in accuracy must be accepted. Since the discussion here is of off-line batch algorithms, the EKF technique will not be discussed any further.

The dimension of the extended nonlinear system usually is much larger than the dimension of the original system. For example, a six-degree-of-freedom rigid-body model that often has eight states can have as many as 50 or more parameters (see Sec. V). In that case, the filtering (or smoothing) will be of a system with 58 states, which is computationally very intensive, since the filtering or smoothing involves matrix equations of the order of the number of states. Since the system is nonlinear, the option of using a simpler time-invariant (steady-state) filter or smoother is not available, even when the original model is linear and time invariant. Also, it should be kept in mind that filtering or smoothing of this nonlinear system is iterative.

This approach is used in SMACK,[23] a nonlinear smoothing program developed at NASA Ames Research Center for aircraft flight and accident data analysis. Some computational savings are introduced by using the fact that the extended states (the parameters) have no dynamics. Overall, though, the increased dimension of the problem introduces a computationally intensive, iterative algorithm.

Using the smoothing identification algorithm, the parameter identification problem is split into two parts. First, holding the parameters constant, a smoothing problem is solved. For linear systems, there are noniterative closed-form solutions for the problem. Then, using the smoothed results, the parameters are updated by a quasi-Newton procedure. The process is repeated until the parameters converge.

The iterative solution of the large nonlinear smoothing problem is replaced in the smoothing identification algorithm by the iterative solution of two smaller subproblems: smoothing and parameter update using a gradient method. More iterations may be required using this technique. However, each iteration is computationally less intensive, especially when the original system model is linear. When the system is both linear and time invariant, additional savings can be introduced by using a time-invariant smoothing algorithm.

To demonstrate the rapid increase of dimension and the introduction of nonlinearities in the extended system, the nonlinear smoothing approach will be applied to the first-order example introduced earlier.

The state of the system given in Eqs. (16) and (17) is expanded to include the two parameters to be identified, i.e.,

$$y(i) = [\,y_1(i)\ \ y_2(i)\ \ y_3(i)\,]^T = [\,x(i)\ \ a\ \ h\,]^T \tag{30}$$

and the state and measurement equations become

$$y_1(i+1) = y_2(i)y_1(i) + u(i) + w(i) \tag{31}$$

$$y_2(i+1) = y_2(i) \tag{32}$$

$$y_3(i+1) = y_3(i) \tag{33}$$

$$z(i) = y_3(i)y_1(i) + v(i) \tag{34}$$

The dimension of the state vector $y(i)$ is increased to three and Eqs. (31) and (34) are nonlinear.

To solve the nonlinear smoothing problem associated with this example, the gradients of the state and measurement equations with respect to the states are required.[14] These gradients, which can be viewed as the system matrices of the linearized nonlinear system, are

$$\Phi(i) = \begin{bmatrix} y_2(i) & y_1(i) & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \Gamma_w(i) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$H(i) = [\,y_3(i)\ \ 0\ \ y_1(i)\,] \tag{35}$$

The matrices $\Phi(i)$ and $H(i)$ depend on the expanded systems states, i.e., these matrices are time dependent. Thus a time-varying iterative algorithm is needed to solve this nonlinear three-dimensional smoothing problem. In contrast, the smoothing identification algorithm uses a simple first-order time-invariant smoother for the same identification task [Eqs. (26) and (27)].

Performance Comparison for a Single-Input, Single-Output Example

The system given by Eqs. (16) and (17) was simulated on a digital computer. The nominal values of the system parameters were taken as $a = 0.9$ and $h = 0.5$. In the simulation, the sequence $[u(i)]$ is a 20-step doublet of magnitude 1 followed by 10 samples with zero input. The process and measurement noises are created using a digital random signal generator with covariances $q = 0.01$ and $r = 0.1$.

The maximum likelihood, the nonlinear smoothing, and the smoothing identification algorithms were used to identify the parameters a and h. The initial guesses of the parameters were taken as $a_0 = 0.7$ and $h_0 = 0.3$ for all three algorithms.

Figure 1 shows the direction of the initial gradients of the performance measures with respect to the parameters in the a vs h plane. It is clearly seen that these gradient directions agree well. However, as described earlier in this section, fewer computations were involved in computing the gradient using the smoothing identification algorithm. The CPU time that was needed to compute the gradients using the smoothing identification technique was 40% of the CPU time used by the maximum likelihood approach and 36% of the CPU time used by the nonlinear smoothing algorithm.

Figure 2a shows the performance measures of the three algorithms as a function of the iteration number. The performance measures are normalized by their initial values. Figure 2b shows the values of the identified parameter a as a function of the iteration number. The maximum likelihood and the smoothing identification algorithms took six iterations to converge, whereas the nonlinear smoothing algorithm converged in four iterations. Convergence was declared when there was no change in four significant digits of the performance measures.

The savings of the smoothing identification algorithm are evident. Compared to maximum likelihood, each iteration is computationally less intensive, and overall, the smoothing identification algorithm used about 40% of computer time to perform the identification. The nonlinear smoother solves a three-dimensional accessory smoothing problem at each iteration. Even though fewer iterations were performed, our algorithm used about 52% of the CPU time compared to the nonlinear smoothing. The savings of our algorithm would be even greater if more parameters were identified.

Table 1 presents the final parameter values and the associated Cramér-Rao bounds (CRB), which are an estimate of the accuracy of the identified parameters. The Cramér-Rao bounds are estimated by

$$\mathrm{CRB} \approx \sqrt{[\mathrm{HESS}^{-1}]_{ii}} \tag{36}$$

where HESS is the Hessian or the second gradient matrix of the performance measure with respect to the parameters, and the ii indicates the diagonal elements of the matrix. In maximum likelihood, the HESS is estimated from the sensitivity functions, Eq. (13). The nonlinear smoother propagates the state covariance matrices and, thus, part of it is the desired Hessian. The smoothing identification algorithm estimates HESS^{-1} using the rank-two update procedure. The results presented in Table 1 demonstrate the comparable accuracy of the three algorithms evaluated here.

The identification was repeated with a wrong value of the process noise covariance $q = 0.1$ instead of 0.01 that was used in the simulation. The identification results are presented in Table 1. The similarity in the results demonstrates the comparable sensitivity of the algorithms to the value of the noise covariance.

Figures 2a and 2b and Table 1 clearly demonstrate the comparable behavior of the smoothing identification algorithm compared to the maximum likelihood and nonlinear smoothing identification algorithms. This means that the significant computational savings introduced by the smoothing identification algorithm do not affect its performance and accuracy.

IV. Multimaneuver Identification Algorithm

To achieve better identification results, flight test maneuvers are often repeated to get more data at a particular point of the flight envelope. Also, most of the maneuvers provide data that can be used to identify only a subset of the parameters of interest by exciting only a few of the aircraft modes (e.g., maneuvers with primarily longitudinal or lateral inputs). Thus, an algorithm that deals with data from several maneuvers is needed. The algorithm presented in Sec. II is extended here to treat such data.

Currently, multimaneuver data are treated in an empirical fashion. In the time domain, the data from different maneuvers are connected, using empirical data segments to obtain a smooth transient between the experimental data.[18,24] The connecting segments are not weighted in the identification. However, since all of the data are processed as one data string, the estimated initial conditions of each maneuver numerically de-

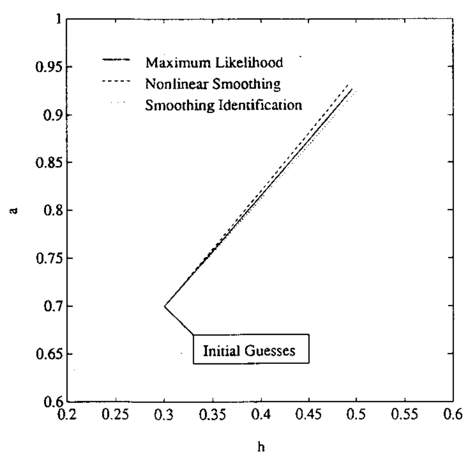

Fig. 1 Single-input, single-output example—the gradients in the parameter space.

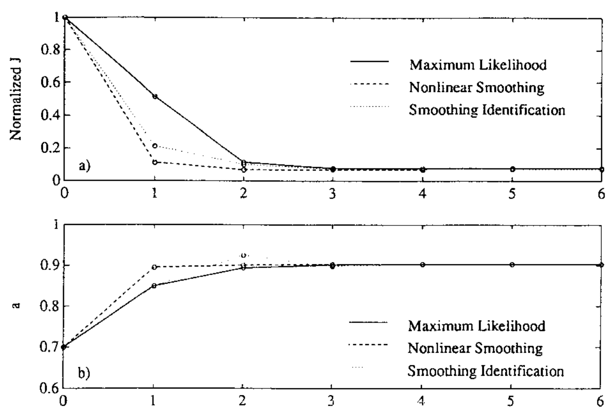

Fig. 2 Single-input, single-output example—results: a) normalized performance measures, b) convergence of the parameter a.

Table 1 Comparison of the identified parameters with correct and wrong q

	Maximum likelihood		Nonlinear smoothing		Smoothing and identification	
Parameters	Correct q	Wrong q	Correct q	Wrong q	Correct q	Wrong q
a	0.9032	0.9042	0.9031	0.9030	0.9031	0.9030
CRB[a]	0.0107	0.0195	0.0097	0.0202	0.0096	0.0203
h	0.5212	0.5127	0.5210	0.5323	0.5210	0.5323
CRB[a]	0.0269	0.0451	0.0242	0.0477	0.0234	0.0492

[a]CRB = Cramér-Rao bounds.

pend on the previous data, which also means that the identification results will depend on the order in which the experimental data are appended. This does not represent actual flight tests, where initial conditions of each maneuver are independent of each other.

In the frequency domain, empirical windowing techniques are used[19] to smooth the transients between the data segments. These techniques do not make use of the often important a priori knowledge of the model describing the identified system.

Using the smoothing and identification technique presented here, the problem of transients between different data segments is avoided. The data from each maneuver are smoothed separately with a fixed set of system parameters. The smoothed data from all of the maneuvers are then used together to compute an overall gradient of the performance index with respect to the parameters. By this approach, at the convergence of the parameters, the initial conditions for each maneuver are estimated independently from each other. To obtain those estimates with the standard maximum likelihood techniques, the initial conditions of all of the maneuvers would have to be treated as additional parameters that would greatly increase the number of parameters to be identified.

For the multimaneuver case, the identification performance measure is defined by

$$J_T = \sum_{j=1}^{M} \alpha_j J_j \tag{37}$$

where J_j is the performance measure defined as in the single maneuver case, Eq. (2), and is evaluated based on the smoothed data (initial conditions, time histories of the states, and process noise vectors) of the jth maneuver only.

The factors α_j allow us to weight the data of different maneuvers by their relative accuracy. For example, data from flight tests performed on a calm day are more reliable and, thus, should be weighted more than data collected on a windy day.

The performance measure J_T is modified to include the constraint equations (1):

$$\bar{J}_T = \sum_{j=1}^{M} \alpha_j \bar{J}_j \tag{38}$$

where $\bar{J}_j$ is the performance measure of each maneuver modified by the constraints similar to the single maneuver case, Eq. (3).

The first variation of $\bar{J}_T$ due to small changes of the unknowns of the problem is given by

$$\delta\bar{J}_T = \sum_{j=1}^{M} \alpha_j \left[\sum_{i=0}^{N} \frac{\partial \bar{J}_j}{\partial x_j(i)} \delta x_j(i) + \sum_{i=0}^{N-1} \frac{\partial \bar{J}_j}{\partial w_j(i)} \delta w_j(i) + \frac{\partial \bar{J}_j}{\partial \theta} \delta\theta \right] \tag{39}$$

At a stationary point of $\bar{J}_T$, the first variation $\delta\bar{J}_T$ vanishes. Since the variations $\delta x_j(t)$, $\delta w_j(i)$, and $\delta\theta$ are independent, each one of the terms in Eq. (39) has to vanish. For a given set of parameters θ, setting the first two terms of Eq. (39) to zero for each maneuver separately and incorporating the constraint equations (1) leads to a smoothing problem for the data of that maneuver, as described in Sec. II.

Once the smoothing problem for each maneuver is solved, it follows from Eq. (39) that

$$\delta\bar{J}_T = \sum_{j=1}^{M} \alpha_j \frac{\partial \bar{J}_j}{\partial \theta} \delta\theta$$

which also has to vanish at the stationary point. In general, that will not be the case and the parameters θ must be changed to decrease the performance measure $\bar{J}_T$.

The gradients of $\bar{J}_j$ with respect to the parameters θ are computed by Eq. (7). Subsequently, the gradients of $\bar{J}_T$ with respect to the parameters are evaluated by

$$\frac{\partial \bar{J}_T}{\partial \theta} = \sum_{j=1}^{M} \alpha_j \frac{\partial \bar{J}_j}{\partial \theta} \tag{40}$$

Based on this gradient information, the parameters θ can be updated using a quasi-Newton procedure to minimize $\bar{J}_T$:

$$\theta_{\text{new}} = \theta_{\text{old}} - (\bar{J}_{T\theta\theta})^{-1} \bar{J}_{T\theta} \tag{41}$$

The inverse of the Hessian matrix $(\bar{J}_{T\theta\theta})^{-1}$ will be again numerically estimated using the rank-two update procedure.

An outline of the multimaneuver identification algorithm is the following:

1) For a set of parameters θ, obtained from the preceding iteration or an initial guess θ_0, for each maneuver separately, solve the associated smoothing problem to compute the time histories of the smoothed states and the forcing functions and evaluate the performance measure J_T given by Eqs. (2) and (37).

2) Evaluate the gradient of $\bar{J}_T$ with respect to the parameters, Eqs. (7) and (40).

3) Update the parameters θ using a quasi-Newton procedure, where the inverse of the Hessian of $\bar{J}_T$ is estimated using the rank-two update algorithm.

4) Repeat until the changes in θ are small and the performance measure is minimized.

V. BO-105 Rigid-Body Model

In this section, the identification results of the linear rigid-body model of the BO-105 helicopter from flight test data are presented. The flight tests were performed at 80 kt forward flight speed at a density altitude of 3000 ft and were conducted by DLR, Institute for Flight Mechanics, Braunschweig, Germany. The goal is to test the suggested smoothing identification algorithm and compare its performance with two similar models that were identified using maximum likelihood and frequency-domain techniques.[18]

The original BO-105 flight test data were taken at 100-Hz sampling rate. For a low-bandwidth rigid-body model identification, a lower sampling rate can be used, reducing the amount of data for processing. The rigid-body modes of a helicopter lie in the frequency range below about 2 Hz, and thus the experimental data bandwidth and sampling rate can be reduced.[19] In this study, the data were filtered to reduce their bandwidth to 2 Hz and the sampling rate was reduced to 20 samples/s using the approach presented by Milne.[8]

The helicopter rigid-body motions are coupled in the longitudinal and lateral axes. Thus, a coupled six-degree-of-freedom (DOF) model is used to identify the rigid-body dynamics. In the flight tests, the helicopter was excited with relatively small inputs, about 4% stick deflection from the trim position, causing only small attitude changes. Thus, a linear 6-DOF model can be used, where the variables of the model are the perturbations of the helicopter states around the trim values of the straight and level flight.[25]

The unmodeled rotor dynamics strongly affect the identification results of the rigid-body model.[18,19] Pure time delays were used to represent these dynamics to improve the identification results. In the time domain, the delays were identified separately before the 6-DOF identification using correlation techniques.[18] In the frequency domain, these delays were identified as part of the overall model from the frequency response data.[19] In the current work, the equivalent rotor delays are modeled by four first-order lag filters, one for each control input.

The state vector of the overall model consists of eight rigid-body states and four rotor delay states. The rigid-body states are the three body axis components of the translational velocity, the angular rates, and the roll and pitch attitude angles. All of the rigid body states are measured. The measurements also include the three body axis components of the translational acceleration.

The flight test data include maneuvers with doublet, multistep "3211," and frequency sweep inputs of the four control axes: longitudinal and lateral stick, collective, and pedals. The inputs were produced by the pilot, who followed a prescribed input time history displayed to him on an onboard monitor. The multistep "3211" input maneuvers were used by DLR for time-domain identification of the BO-105 rigid body model,[18] and the frequency sweep input maneuvers were used by the U.S. Army to identify this model using frequency-domain techniques.[18,19] The maneuvers with the doublet inputs were then used for model validation.

The time-domain smoothing identification algorithm presented in this paper was used to identify the rigid-body model of the BO-105 using the multistep "3211" input maneuvers. Since each maneuver involved excitation of one control axis only, one set of model parameters was identified from the data of four maneuvers, one for each control axis, using the multimaneuver algorithm presented in Sec. IV.

Identification Results

The BO-105 flight tests were carefully monitored to make sure that the data channels were consistent with each other. It was also recently reported that the BO-105 flight test data are of good quality with only a few measurement errors, which include scale factors on the velocity sensors and biases of the roll and pitch rate gyros.[24]

The parameter vector that was identified included the stability and control derivatives, the rotor delay time constants, and the measurement errors mentioned earlier. However, some of the parameters, such as the stability derivatives X_v, X_r, and the control derivative X_{lat}, were not identified because of their negligible effect on helicopter response.

The identified model is compared here to two other identified models.[18,19] The U.S. Army Ames model was identified using frequency-domain techniques, and the DLR model was obtained from a standard maximum likelihood procedure. Tables 2 and 3 present the identification results together with the parameters obtained by NASA Ames and DLR. In Table 4, the natural modes and time constants of the resulting models are compared.

Table 2 BO-105 stability derivatives

Derivative	NASA Ames	DLR	Smoothing
X_u	−0.038	−0.059	−0.055
X_v[a]	0.000	0.000	0.000
X_w	−0.061	0.036	0.041
X_p	0.756	0.000	−0.028
X_q	2.548	0.000	−0.046
X_r[a]	0.000	0.000	0.000
Y_u[a]	0.000	0.000	0.000
Y_v	−0.221	−0.170	−0.195
Y_w	−0.083	0.000	−0.043
Y_p	−2.030	0.000	−0.636
Y_q	4.823	0.000	−0.050
Y_r	0.950	1.332	1.318
Z_u	0.246	0.014	0.161
Z_v[a]	0.000	0.000	0.000
Z_w	−1.187	−0.998	−1.026
Z_p	2.622	0.000	1.084
Z_q	7.011	5.012	4.997
Z_r[a]	0.000	0.000	0.000
L_u	−0.061	−0.081	−0.098
L_v	−0.206	−0.271	−0.279
L_w	0.168	0.116	0.121
L_p	−8.779	−8.501	−8.548
L_q	3.182	3.037	3.009
L_r	0.991	0.410	0.705
M_u	0.000	0.029	0.029
M_v	0.050	0.048	0.059
M_w	0.096	0.053	0.048
M_p	−0.998	−0.419	−0.447
M_q	−4.493	−3.496	−3.496
M_r	−0.438	−0.117	−0.195
N_u[a]	0.000	0.000	0.000
N_v	0.082	0.117	0.117
N_w	−0.119	0.034	0.037
N_p	−0.466	−1.057	−1.038
N_q	5.432	0.809	0.793
N_r	−1.070	−0.858	−1.087

[a]Assumed zero a priori.

Model Validation

The last step of a system model identification is its validation using flight test data other than those used in the identification. Maneuvers with doublet inputs were used to validate the three BO-105 rigid-body models: the model identified in this work and the Ames and DLR models. For that, the models are driven with the actual inputs of these maneuvers and the model output is compared to the experimental data. As part of the validation procedure, the state equation biases and zero shifts in the measurement data, which result from inaccurate estimates of the initial conditions and control trim settings, are estimated by minimizing the weighted-least-squares error between the model and the vehicle responses.

The validation results are demonstrated in Figs. 3 and 4 for the four doublet maneuvers using the three models described here. As an example, the roll and pitch rate time histories are

Table 3 BO-105 control derivatives and equivalent rotor time constants

Derivative	NASA Ames	DLR	Smoothing
X_{lon}	−0.072	−0.028	−0.013
X_{lat}[a]	0.000	0.000	0.000
X_{ped}[a]	0.000	0.000	0.000
X_{col}	−0.046	0.000	0.002
Y_{lon}[a]	0.000	0.000	0.000
Y_{lat}	0.066	0.003	0.033
Y_{ped}	−0.015	−0.011	−0.020
Y_{col}	−0.032	0.000	0.005
Z_{lon}	−0.103	−0.303	−0.275
Z_{lat}[a]	0.000	0.000	0.000
Z_{ped}[a]	0.000	0.000	0.000
Z_{col}	−0.388	−0.349	−0.373
L_{lon}	0.073	0.024	0.037
L_{lat}	0.179	0.185	0.191
L_{ped}	−0.027	−0.028	−0.027
L_{col}	0.058	0.032	0.043
M_{lon}	0.098	0.093	0.104
M_{lat}	0.000	−0.009	−0.011
M_{ped}	0.013	0.000	0.008
M_{col}	0.073	0.057	0.054
N_{lon}	−0.075	0.000	0.004
N_{lat}	0.033	0.026	0.021
N_{ped}	0.057	0.049	0.051
N_{col}	−0.051	0.000	−0.002
τ_{lon}	0.112	0.100	0.099
τ_{lat}	0.062	0.060	0.060
τ_{ped}	0.044	0.040	0.039
τ_{col}	0.168	0.040	0.038

[a]Assumed zero a priori.

Table 4 BO-105 modes and time constants

	NASA Ames		DLR		Smoothing	
	ω[a]	ζ[b]	ω[a]	ζ[b]	ω[a]	ζ[b]
Pitch	2.601	0.217	2.499	0.142	2.520	0.185
Roll	0.300	−0.363	0.328	−0.151	0.323	−0.148
	Real poles 1/s		Real poles, 1/s		Real poles, 1/s	
Roll	8.322		8.490		8.539	
Pitch	6.040		4.362		4.307	
Pitch	0.492		0.599		0.697	
Spiral	0.025		0.021		0.028	

[a]Frequency, (rad/s).
[b]Damping ratio; if negative, unstable poles.

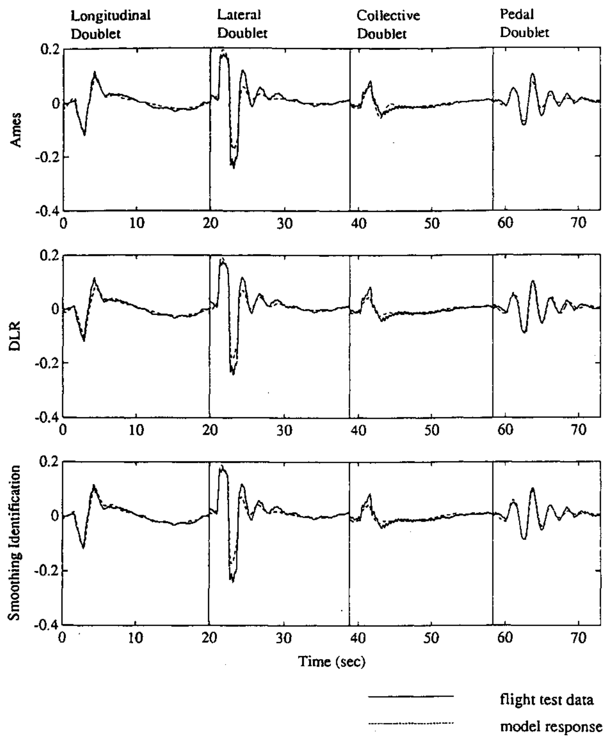

Fig. 3 BO-105 model validation: roll rate, rad/s.

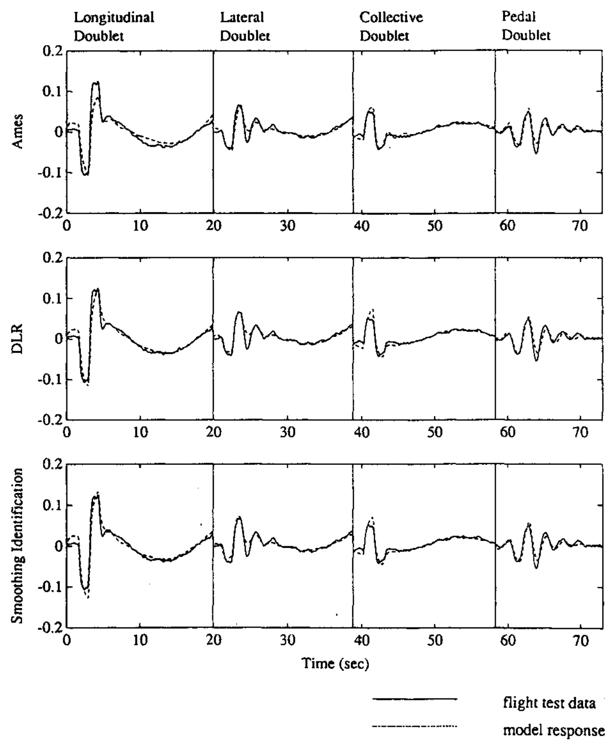

Fig. 4 BO-105 model validation: pitch rate, rad/s.

presented. These plots present the accuracy and the predictive capabilities of the models.

Discussion

The BO-105 rigid-body model identification results are presented in Tables 2–4. The parameter values obtained using the three different identification techniques (frequency domain, maximum likelihood, and the smoothing identification algorithm) agree favorably. The model validation results, presented in Figs. 3 and 4, further support the comparable accuracy of the three models.

These results provide a good check for the performance of the suggested smoothing identification algorithm. The big advantage of this algorithm is computational savings and a rational multimaneuver technique. Compared to the maximum likelihood algorithms, for the current example, and assuming that the initial conditions are not estimated using maximum likelihood, the smoothing algorithm took about 11% CPU time per iteration. If the initial conditions of every maneuver were identified, the parameter vector for maximum likelihood would increase considerably; in this case of 12 states, there would be an additional 48 parameters to identify. The multimaneuver algorithm presented in Sec. IV identifies those initial conditions as a part of the solution of the smoothing problem. Also, the multimaneuver algorithm provides a consistent way to treat flight test data from few maneuvers.

The rigid-body models can describe accurately the helicopter response in the low-frequency range; however, discrepancies can be still seen at higher frequencies or faster motions (see Figs. 3 and 4). These discrepancies are attributed to the unmodeled rotor dynamics. Even though equivalent rotor lags were introduced in the model, they do not correctly represent the coupling of the helicopter body and rotor dynamics. Models that incorporate the coupling between a simplified rotor model and the rigid-body model were recently presented[18,19]; however, identification difficulties using the given BO-105 experimental data were reported due to the lack of rotor state measurements.

VI. Conclusions

In this paper, a parameter identification algorithm for linear systems based on smoothing was presented. The algorithm smoothes the experimental data with different sets of system model parameters. One smoothing pass through the data provides all of the information required to compute the gradients of the performance measure with respect to any number of system parameters. Computing the gradients from the smoothed data provides big computational savings compared to the standard maximum likelihood approach. The gradients are then used to update the parameters. The procedure is repeated until convergence is obtained.

Using this approach, data from several experiments can be used consistently to extract one set of parameters to match all of the data. The initial conditions for each data segment are automatically estimated as part of the smoothing process and do not have to be treated as additional parameters, which provides additional computational savings.

The proposed algorithm was validated by identifying a six-degree-of-freedom linear rigid-body model of the BO-105 helicopter from flight test data. The identification results are compared to two models identified by maximum likelihood and frequency-domain techniques using the same flight test. The good agreement among the three models demonstrates that the computational savings introduced in the smoothing identification algorithm do not affect its accuracy.

Appendix A: Backward Information Filter, Forward Smoother

Backward Information Filter

Using the final conditions $y_{N/N+1} = 0$ and $S_{N/N+1} = 0$, perform the measurement downdates for $i = N, \ldots, 1$ as follows:

$$y_{i/i} = y_{i/i+1} + H^T(i,\theta)R^{-1}[z(i) - D(i,\theta)u(i)]$$

$$S_{i/i} = S_{i/i+1} + H^T(i,\theta)R^{-1}H(i,\theta)$$

Perform the time downdates for $i = N-1, \ldots, 0$ as follows:

$$K_B(i) = [Q^{-1} + \Gamma_w^T(i)S_{i+1/i+1}\Gamma_w(i)]^{-1}\Gamma_w^T(i)S_{i+1/i+1}$$

$$w_B(i) = Q\Gamma_w^T(i)\,[I - \Gamma_w(i)K_B(i)]^T y_{i+1/i+1}$$

$$y_{i/i+1} = \Phi^T(i,\theta)\,[I - \Gamma_w(i)K_B(i)]^T$$

$$\times\ [y_{i+1/i+1} - S_{i+1/i+1}\Gamma_u(i,\theta)u(i)]$$

$$S_{i/i+1} = \Phi^T(i,\theta)\,[I - \Gamma_w(i)K_B(i)]^T S_{i+1/i+1}\Phi(i,\theta)$$

Store $y_{i/i}$, $w_B(i)$, $S_{i+1/i+1}$, and $K_B(i)$.

Forward Smoother

Compute the initial conditions:

$$\hat{x}(0) = [S_{0/1} + P_0^{-1}]^{-1}\,[y_{0/1} + P_0^{-1}x_0]$$

For $i = 0, \ldots, N-1$ compute

$$\hat{w}(i) = w_B(i) - K_B(i)\,[\Phi(i,\theta)\hat{x}(i) + \Gamma_u(i,\theta)u(i)]$$

$$\hat{x}(i+1) = \Phi(i,\theta)\hat{x}(i) + \Gamma_u(i,\theta)u(i) + \Gamma_w(i,\theta)\hat{w}(i)$$

$$\lambda(i) = S_{i/i}\hat{x}(i) - y_{i/i}$$

Appendix B: Rank-Two Update Procedure

The rank-two update procedure is a technique for estimating the second-order gradient matrix, the Hessian, of the performance measure with respect to the parameters. In a broad view, this procedure is a numerical approximation of the Hessian from the values of the gradient evaluated with two sets of parameters at successive iterations. A derivation of the procedure outlined here can be found in Refs. 21 and 22.

Let θ_i and $\nabla_\theta J_i$ denote the parameters and the gradient of the performance measure with respect to those parameters at iteration i. The new values of the parameters are computed using a quasi-Newton procedure:

$$\theta_{i+1} = \theta_i - \alpha_i B_i^{-1}\nabla_\theta J_i \tag{B1}$$

where B_i is the estimate of the Hessian matrix for the parameters θ_i. The scalar α_i is the step size that leads to a minimum of the performance measure along the search direction given by $-B_i^{-1}\nabla_\theta J_i$. The gradient of the performance measure for the new values of the parameters θ_{i+1} is $\nabla_\theta J_{i+1}$.

The change in the parameter and gradient vectors is defined as

$$p_i = \theta_{i+1} - \theta_i$$

$$q_i = \nabla_\theta J_{i+1} - \nabla_\theta J_i$$

The estimate B_i of the Hessian matrix is updated by two rank-one matrices, which are the outer products of the changes p_i and q_i. This provides an at most rank-two update of B_i and is given by

$$B_{i+1} = B_i + \frac{q_i q_i^T}{q_i^T p_i} - \frac{B_i p_i p_i^T B_i^T}{p_i^T B_i^T p_i} \tag{B2}$$

To show the rank-two property of this update better, Eq. (B2) can be arranged as

$$B_{i+1} = B_i + \frac{q_i q_i^T}{q_i^T p_i} + \alpha_i\frac{(\nabla_\theta J_i)(\nabla_\theta J_i)^T}{(\nabla_\theta J_i)^T p_i} \tag{B3}$$

where α_i is a scalar, representing the step size of the quasi-Newton procedure described in Eq. (B1). From Eq. (B3), it can be concluded that, if the change in the gradient vector q_i is parallel to the gradient vector $\nabla_\theta J_i$, this update will be of rank-one only.

The initial value of the Hessian matrix B_0 can be chosen as any symmetric positive definite matrix. The identity matrix is commonly used, causing the first parameter update to be in the steepest descent direction.

In this work, the rank-two procedure is used as a part of the quasi-Newton algorithm to update the system parameters, which uses the inverse of the Hessian matrix. Equation (B2) can be inverted to directly update the estimate of the inverse Hessian matrix H_i. The update is given by

$$H_{i+1} = H_i + \left(1 + \frac{q_i^T H_i q_i}{p_i^T q_i}\right)\frac{p_i p_i^T}{p_i^T q_i} - \frac{1}{p_i^T q_i}\,(p_i q_i^T H_i + H_i q_i p_i^T) \tag{B4}$$

This update equation was incorporated in the identification procedure represented in this paper.

Acknowledgments

This work was performed under Grant 2-106, NASA Ames Research Center, Flight Dynamics and Controls Branch. Financial and technical support was also provided by Advanced Rotorcraft Technology, Inc., Mountain View, California. The BO-105 flight test data were produced by the DLR, Institute for Flight Mechanics, Braunschweig, Germany, for use by the AGARD Flight Mechanics Panel Working Group 18, "Rotorcraft System Identification," December 1987, and were used with permission.

References

[1]Maine, R. E., and Iliff, K. W., "Formulation and Implementation of a Practical Algorithm for Parameter Estimation with Process and Measurement Noise," *Journal of Applied Mathematics,* Vol. 41, No. 3, Dec. 1981, pp. 558–579.

[2]Iliff, K. W., and Maine, R. E., "More than You May Want to Know About Maximum Likelihood Estimation," *Proceedings of the AIAA Atmospheric Flight Mechanics Conference,* AIAA, New York, Aug. 1984, pp. 1–25.

[3]Iliff, K. W., "Parameter Estimation for Flight Vehicles," *Journal of Guidance, Control, and Dynamics,* Vol. 12, No. 5, 1989, pp. 609–622.

[4]Sage, A. P., and Melsa, J. L., *System Identification,* Academic Press, New York, 1971.

[5]Bach, R. E., Jr., and Chen, F. Y., "Filtering and Smoothing Techniques for Parameter Estimation," *Proceedings of the 9th Asilomar Conference on Circuits, Systems and Computers,* Pacific Grove, CA, Nov. 1975, pp. 493–495.

[6]DuVal, R. W., "A Rapidly Converging Adaptive Filter for On-Line Applications," Ph.D. Dissertation, Dept. of Aeronautics and Astronautics, Stanford Univ., Stanford, CA, July 1976.

[7]Jategaonkar, R., and Plaetschke, E., "Maximum Likelihood Parameter Estimation from Flight Test Data for General Nonlinear Systems," German Aerospace Research Establishment, DFVLR-FB 83-14, 1983.

[8]Milne, G. W., "Identification of a Dynamic Model of a Helicopter from Flight Tests," Ph.D. Dissertation, Dept. of Electrical Engineering, Stanford Univ., Stanford, CA, Dec. 1986.

[9]Mehra, R. K., "Identification of Stochastic Linear Dynamic Systems Using Kalman Filter Representation," *AIAA Journal,* Vol. 9, No. 1, 1971, pp. 28–31.

[10]Chen, R. T. N., and Eulrich, B. J., "Parameter and Model Identification of Nonlinear Dynamical Systems Using a Suboptimal Fixed-Point Smoothing Algorithm," Joint Automatic Control Conference, Saint Louis, MO, Aug. 1971.

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

[12]Hansen, R. S., "DEKFIS User's Guide—Discrete Extended Kalman Filter/Smoother Program for Aircraft and Rotorcraft Data Consistency," NASA CR 159081, Nov. 1979.

[13]DeMiroz, M. Y., "Identification of Rotor System Research Aircraft (RSRA) Rigid Body Dynamics in Time and Frequency Domains," Engineer Thesis, Dept. of Aeronautics and Astronautics, Stanford Univ., Stanford, CA, June 1986.

[14]Bach, R. E., Jr., "Variational Algorithms for Nonlinear Smoothing Applications," NASA TM 73,211, June 1977.

[15]Bach, R. E., Jr., "A Variational Technique for Smoothing Flight-Test and Accident Data," *Journal of Aircraft,* Vol. 19, No. 7, 1982, pp. 546–552.

[16]Cox, A. B., and Bryson, A. E., Jr., "Identification by a Combined Smoothing Nonlinear Programming Algorithm," *Automatica,* Vol. 16, No. 6, 1980, pp. 689–694.

[17]Segal, M., and Weinstein, E. "A New Method for Evaluating the Log-Likelihood Gradient (Score) of Linear Dynamic Systems," *IEEE*

Transactions on Automatic Control, Vol. 33, No. 8, 1988, pp. 763–766.

[18]Kaletka, J., von Grünhagen, W., Tischler, M. B., and Fletcher, J. W., "Time and Frequency-Domain Identification and Verification of BO-105 Dynamic Models," Fifteenth European Rotorcraft Forum, Paper No. 66, Amsterdam, Sept. 1989.

[19]Tischler, M. B., and Cauffman, M. G., "Frequency-Response Method for Rotorcraft System Identification with Applications to the BO-105 Helicopter," 46th Annual Forum of the American Helicopter Society, Washington, DC, May 1990.

[20]Kaminski, P. G., "Square Root Filtering and Smoothing for Discrete Processes," Ph.D. Dissertation, Dept. of Aeronautics and Astronautics, Stanford Univ., SUDAAR 427, Stanford, CA, July 1971.

[21]Cuthbert, T. R., Jr., *Optimization Using Personal Computers, with Application to Electrical Networks,* Wiley, New York, 1987.

[22]Luenberger, D. G., *Linear and Nonlinear Programming,* 2nd ed., Addison-Wesley, Reading, MA, 1984.

[23]Bach, R. E., Jr., "State Estimation Applications in Aircraft Flight-Data Analysis: A User's Manual for SMACK," NASA RP 1252, March 1991.

[24]Fletcher, J. W., "Obtaining Consistent Models of Helicopter Flight Data Measurement Errors Using Kinematic Compatibility and State Reconstruction Methods," 46th Annual Forum of the American Helicopter Society, Washington, DC, May 1990.

[25]McRuer, D. T., Ashkenas, I. L., and Graham, D., *Aircraft Dynamics and Control,* Princeton University Press, Princeton, NJ, 1973.