JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 17, No. 4, July–August 1994

Maximum Likelihood Identification and Realization of Stochastic Systems

Yaakov Oshman* and Tal Mendelboim†
Technion—Israel Institute of Technology, Technion City, Haifa 32000, Israel

A new maximum likelihood (ML) realization/identification technique is presented. The method utilizes the recently introduced eigensystem realization algorithm (ERA) in combination with a stochastic adaptive filter/fixed-interval smoother. The resulting algorithm, called ML/ERA, is thus capable of estimating a minimal, internally balanced realization for a stochastic system whose process and/or measurement noise covariances are not necessarily known. Belonging to the ML class of algorithms, the new method is consistent and asymptotically efficient under reasonable conditions. Moreover, by using standard statistical testing techniques, the user is able to assess the quality of the resulting estimates during the iterative estimation process. A numerical investigation of the performance of the new algorithm has shown a vast improvement over the performance of the original, unaugmented ERA. In cases where the ERA could not determine the system order because the data was completely masked by noise, the ML/ERA algorithm was able to identify the order and realize an accurate mathematical model of the system. Numerical examples, demonstrating the performance of the new algorithm, are included in the paper.

I. Introduction

THE problem addressed in this paper is that of constructing a state space mathematical model of a linear, dynamic, Gaussian process driven stochastic system, utilizing Gaussian noise contaminated measurements. Specifically, we focus our attention on the recently presented realization method called eigensystem realization algorithm (ERA), developed by Juang and Pappa.[1] This method is essentially an extension of the Ho and Kalman minimum realization procedure[2] for deterministic linear systems, which uses noise-free data to construct a completely controllable (CC) and completely observable (CO) state space representation (minimal realization) of the system. In contrast to the classical system realization theory, which uses the Hankel matrix, the ERA algorithm operates on a block data matrix which is called the generalized Hankel matrix. This matrix is composed of the system measured (or otherwise estimated) impulse response (the system Markov parameters), and is obtained from the Hankel matrix by deleting some rows and columns. Thus, one is allowed to include in the analysis only data which is believed to be "good" in some sense, e.g., "strongly measured" data, or data pertaining to "strongly excited" modes of the system.

The ERA method consists mainly of an application of the singular value decomposition (SVD) to the generalized Hankel matrix, to determine the system order and, subsequently, the system realization. In a (theoretical) completely deterministic case, the order is determined from the number of nonzero singular values. In a (real) stochastic system, however, nonzero singular values exist, which should have been zero in the noise-free case. A threshold value for the singular values has to be chosen, therefore, below which the singular values are attributed only to noise effects.[1] Assuming that the system order has been successfully determined, the ERA method proceeds to compute a minimal realization of the system. That is, it computes a matrix triple (A, B, C), where A is the system transition matrix, B is the control distribution matrix, and C is the observation matrix, such that (A, B) are CC and (A, C) are CO. The realization produced by the ERA algorithm is internally balanced. Moreover, since the eigenvalues of the system are invariant under a nonsingular transformation, the modal parameters (natural frequencies and damping ratios) of the realized triple are the modal parameters of the system. Hence, in addition to its ability to produce a state space realization for the system, the ERA may also be considered to be a time-domain modal parameter identification method. Note, in that regard, that the modal parameters of the system play an important role in certain applications, e.g., control of large flexible structures.

As described, the original ERA is essentially a deterministic identification method. It is based on the underlying theoretical assumptions that no process noise drives the system and that it operates on noise-free, or at least low noise data. For high noise levels, however, severe problems may arise. In these cases, since the true smallest singular values of the system may be in the order of (or sometimes even smaller than) the singular values which are due to the noise, the choice of the singular value threshold may not be clear.[6] Consequently, the ERA may not determine the system order reliably, which may lead to an inaccurate identification of the system modal parameters. Note, that even if the appropriate system order has been somehow determined, the resulting realization may be inaccurate due to "spillover" of the noise effects into the system-related singular values. On the other hand, it should be noted that the method is very simple to implement, involving only an SVD of a block data matrix. Moreover, it has been shown to perform extremely well with low noise data,[3] mainly by virtue of the excellent numerical robustness of the SVD.[4] Because of these characteristics, the ERA has become very popular in recent years, especially in large space structures applications, where experimental verification of the structure model (especially the modal parameters) is sometimes possible only while in orbit (e.g., huge space antennas).

During the last few years, several methods have been suggested to alleviate the ERAs inherent sensitivity to high noise data. These methods can be categorized as belonging to two classes, which we call passive and active methods. The passive methods[3,5] do not modify the ERA method itself. The basic idea underlying these methods is that the known statistical

Received Dec. 26, 1992; presented as Paper 93-3840 at the AIAA Guidance, Navigation, and Control Conference, Monterey, CA, Aug. 9–11, 1993; revision received Sept. 27, 1993, accepted for publication Sept. 27, 1993.

*Senior Lecturer, Head, Flight Control Laboratory, Department of Aerospace Engineering. Member AIAA.

†Graduate Student, currently with Israeli Air Force, I.D.F.

characteristics of the measurement noise can be utilized to correctly determine an optimum singular values truncation and to assess the quality of the identified modal parameters. Thus, in Ref. 3 a method was developed to compute an optimal cutoff value for the singular values of the block data matrix, based on the measurement noise covariance. Also, the so-called modal amplitude coherence accuracy indicators were presented, which may be used to explicitly evaluate the effects of noise on the identified parameters. In Ref. 5, confidence intervals for the frequencies and damping ratios obtained from the realization are given as a function of the measurement noise covariance. The measurement noise is treated as a perturbation and sensitivities to this perturbation are computed. Again, the ERA is not modified, but the bias and variance information may aid the analyst in determining which data points to use and how much data should be used for best results.

An innovative approach, which we call the active approach, has been recently proposed by Mook and Lew.[6] Their method, called ERA/MME, involves a combination of the original ERA method with the minimum model error (MME) estimator. The MME algorithm[7] is a powerful batch state estimator that presupposes that the available mathematical model of the system is imperfect. Hence, constraining assumptions on the system dynamics and/or the driving process noise are not necessary (compare to the classical state estimation/smoothing theory assumptions). To compensate for the model errors, the method processes the measurements at the same time seeking the least amount of correction of the assumed (albeit erroneous) model, such that the covariance of the "actual measurements minus estimated measurements" process would be as close as possible to the (assumed known) covariance of the measurement noise process (which is the "actual measurements minus noise-free measurements" process). The idea presented in Ref. 6 is that by processing the measurements first, using the MME estimator, then feeding the simulated measurements to the ERA, a better realization could be computed by the ERA. This better realization could, in turn, be used in a second MME run, to produce an even better data, which could then be fed again to the ERA. The resulting ERA/MME thus proceeds in an iterative manner. The numerical example shown in Ref. 6 and the recent application of the method to the identification of the mini-mast truss structure[8] demonstrate the viability of the method.

In this paper we enhance the ERA's capability to deal with inherently stochastic systems by using it within a maximum likelihood (ML) parameter identification/state estimation scheme. Using an active approach somewhat similar to that of the ERA/MME method, the technique presented in this paper augments the ERA, still functioning as a realization algorithm, with a stochastic adaptive filter/fixed-interval smoother that performs the task of optimally filtering the noise from the measurements. Joined based on an underlying maximum likelihood identification approach, the two separate algorithms which constitute the new method iteratively seek the optimal realization and covariance estimates that maximize the log-likelihood function for the identified state space model. The resulting combined realization/identification scheme still enjoys the previously mentioned good characteristics of the ERA but can now cope with the stochastic attributes of the system. Since an integral linear optimal state estimator is used to compute the likelihood function, it is easy to check the performance of the iterative scheme at each stage, by using standard statistical testing techniques to verify the consistency of the state estimator involved. Moreover, belonging to the class of ML identification algorithms, the new method is guaranteed to be consistent and asymptotically efficient under reasonable conditions.

Since the ERA plays a major role in the presented ML scheme, it is briefly summarized in the next section. Then, in Sec. III, the realization/identification problem is mathematically formulated. The maximum likelihood identification algorithm is presented in Sec. IV. In Sec. V we present three numerical examples, which serve to demonstrate the performance of the new algorithm. One of the examples is used for the purpose of a qualitative comparison of the ML/ERA algorithm and the ERA/MME algorithm of Mook and Lew. Concluding remarks and recommendations for further research are offered in the last section.

II. Eigensystem Realization Algorithm

For the purpose of the ensuing development, we present in this section a brief description of the eigensystem realization algorithm (ERA).

As previously mentioned, the ERA is essentially a deterministic realization method which was developed by Juang and Pappa[1] in 1985 using a system-theoretic approach, based on the minimal realization procedure developed by Ho and Kalman.[2] The realization task is to estimate, based on the system impulse response measured (or otherwise estimated) data, the matrix triple (A, B, C) which appears in the state space canonical representation of the linear, time-invariant (LTIV) system of order n_0

$$\boldsymbol{x}_{k+1} = A\boldsymbol{x}_k + B\boldsymbol{u}_k \tag{1a}$$

$$\boldsymbol{y}_k = C\boldsymbol{x}_k \tag{1b}$$

Here $\boldsymbol{x} \in \mathcal{R}^{n_0}$ is the state vector, $\boldsymbol{u} \in \mathcal{R}^p$ is the deterministic input to the system (the control action), and $\boldsymbol{y} \in \mathcal{R}^m$ is the output vector (the measurement). The state transition matrix is $A \in \mathcal{R}^{n_0, n_0}$ and the matrices $B \in \mathcal{R}^{n_0, p}$ and $C \in \mathcal{R}^{m, n_0}$ are the control distribution and observation matrices, respectively. The impulse response of the system (1) is given by the following set of Markov parameters as

$$Y(k) = CA^kB, \qquad k = 0, 1, 2, \ldots \tag{2}$$

The ERA operates on the following (generalized Hankel) block data matrix $H(k) \in \mathcal{R}^{rm, sp}$,

$$H(k) = \begin{bmatrix} Y(k) & Y(k+j_1) & Y(k+j_2) & \cdots & Y(k+j_{s-1}) \\ Y(k+i_1) & Y(k+i_1+j_1) & Y(k+i_1+j_2) & \cdots & Y(k+i_1+j_{s-1}) \\ \vdots & \vdots & \vdots & \ddots & \\ Y(k+i_{r-1}) & Y(k+i_{r-1}+j_1) & Y(k+i_{r-1}+j_2) & \cdots & Y(k+i_{r-1}+j_{s-1}) \end{bmatrix} \tag{3}$$

where r and s are arbitrary integers, chosen such that

$$\min(rm, sp) \geq n_0 \tag{4}$$

and the integer sequences $\{i_l\}_{l=1}^{r-1}$ and $\{j_l\}_{l=1}^{s-1}$ are arbitrary. To obtain a realized triple (A, B, C) from the impulse response data, the ERA uses the matrices $H(0)$ and $H(1)$. Perform a singular value decomposition of $H(0)$

$$H(0) = U\Sigma V^T \tag{5}$$

where $U \in \mathcal{R}^{rm, rm}$ and $V \in \mathcal{R}^{sp, sp}$ are the orthogonal left and right singular vector matrices, respectively, and $\Sigma \in \mathcal{R}^{rm, sp}$ is the diagonal singular value matrix, which can be written as

$$\Sigma = \begin{bmatrix} \Sigma_n & 0 \\ 0 & 0 \end{bmatrix} \tag{6}$$

Here n, the number of strictly positive singular values of the matrix $H(0)$ (the entries of the diagonal matrix Σ_n), is the rank of $H(0)$. Note that for a completely (theoretical) deterministic system (i.e., no measurement or process noise) the rank of H(0) is equal to the true system order n_0. That is, in the singular value decomposition (5) we should obtain exactly n_0 nonzero singular values. However, due to measurement noise, computer roundoff errors, and process noise (which can also be interpreted as system nonlinearity and model uncertainty effects), H(0) is usually of full mathematical rank. In this case, one has to decide which singular values actually represent the information portion of the measured data and which can be attributed to noise effects, and thus should be disregarded (i.e., those singular values that should have been zero in the deterministic case). For sufficiently low noise data, the task of determining the singular values threshold should not be too difficult, as has been observed in several applications of the ERA. This merit of the ERA can be attributed to the usage of the SVD, which is the most reliable method for cases of rank deficient matrices.[4] However, when the noise level is not very low, this task may become quite delicate, and several methods have been suggested for that purpose.[3]

Assuming that the singular values threshold has been determined, resulting in the n most "significant" singular values, the rest of the realization procedure is straightforward. Define the rectangular matrices U_n and V_n to be the matrices formed by the first n left and right singular vectors, respectively. Then, the realization of order n is determined according to the following equations:

$$\hat{A} = \Sigma_n^{-1/2}\ U_n^T H(1) V_n\ \Sigma_n^{-1/2} \tag{7}$$

$$\hat{B} = \Sigma_n^{-1/2}\ V_n^T E_p \tag{8}$$

$$\hat{C} = E_m^T U_n\ \Sigma_n^{1/2} \tag{9}$$

In Eqs. (8) and (9), the selection matrix E_j is defined as

$$E_j = \begin{bmatrix} I_j \\ 0 \end{bmatrix} \tag{10}$$

where $I_j \in \mathcal{R}^{j,j}$ is a unit matrix and the row dimension of E_p and E_m is determined in accordance with Eqs. (8) and (9), respectively. It can be shown[1] that the realization [Eqs. (7–9)] is an internally balanced minimal realization, which is another attractive merit of the ERA.

In the next section we define mathematically the realization/identification problem to be addressed in the rest of this paper.

III. Problem Statement

The algorithm to be presented in the sequel is a stochastic estimator, which explicitly takes into account the stochastic inputs of the system under consideration. For that purpose, the following stochastic mathematical model, which is an extension of the model in Eqs. (1a) and (1b), is assumed:

$$\boldsymbol{x}_{k+1} = A\boldsymbol{x}_k + B\boldsymbol{u}_k + G\boldsymbol{w}_k \tag{11a}$$

$$\boldsymbol{y}_k = C\boldsymbol{x}_k + \boldsymbol{v}_k \tag{11b}$$

Here, $\boldsymbol{w} \in \mathcal{R}^q$, the process noise, is a zero mean, Gaussian distributed white sequence with covariance

$$E\boldsymbol{w}_k\boldsymbol{w}_j^T = Q\delta_{kj} \tag{12}$$

and the measurement noise $\boldsymbol{v}$ is a zero mean, Gaussian distributed white sequence with covariance

$$E\boldsymbol{v}_k\boldsymbol{v}_j^T = R\delta_{kj} \tag{13}$$

As previously discussed, the objective of the ERA is to identify a matrix triple (A, B, C) from a given set of (impulse response) measurements. Since we are dealing with a stochastic system where it cannot be assumed, in general, that the noise covariance matrices are known, the identification objective of the new algorithm is set as an extension of the ERA's objective. In the most general way, this identification objective is as follows.

Find a maximum likelihood estimate for the parameter set

$$\theta \equiv \{(A, B, C), (Q, R)\} \tag{14}$$

Remarks:

1) As is well known, the realization (A, B, C) that has a given (measured) impulse response is not unique. Therefore, searching for a realization which is optimal in the maximum likelihood sense in the space of all possible realizations cannot yield a unique estimate. However, the uniqueness problem is settled by noting that the ERA is the tool that will be used to compute the realization. Hence, the subspace of admissible realizations in which the estimator is searching for a maximum likelihood estimate is the subspace of ERA-producible realizations, i.e., the space of internally balanced minimal realizations.

2) For simplicity it has been assumed that the process noise and the measurement noise are stationary; however, as will be clear from the ensuing development, this assumption is not a necessary requirement and the resulting algorithm can be adapted very easily to deal with the case of nonstationary noise processes, i.e., when the covariance matrices Q and R are replaced in the parameter set θ by the sequences $\{Q_k\}_{k=1}^N$ and $\{R_k\}_{k=1}^N$.

3) As a worst-case type of assumption, it has been assumed that neither the process noise covariance nor the measurement covariance are known and, therefore, they should be estimated from the measurements along with the system's realization. However, should either of these covariances be known, this assumption can be relaxed by using the known value instead of estimating it. We will return to this point later.

4) As previously discussed, in the deterministic case or when the measurement noise is sufficiently low, the ERA estimates the system order from the number of positive singular values. In the high noise case, since the ERA cannot reliably estimate the system order, we invoke the maximum likelihood principle. The algorithm to be described does not assume a precise knowledge of the system order. Note, however, that in general the system order is at least known to lie within a certain range, which can be estimated using pre-experiment analyses. Such knowledge can be utilized by the estimator to alleviate the computations involved.

IV. Maximum Likelihood Algorithm

As is well known, the log-likelihood function for the model (11) can be expressed as[9]

$$\mathcal{L}(\theta|\mathcal{Y}^N) = -\frac{1}{2}\sum_{i=1}^{N} \{\tilde{\boldsymbol{y}}_{i/i-1}^T \mathcal{N}_{i/i-1}^{-1} \tilde{\boldsymbol{y}}_{i/i-1} + \log\det \mathcal{N}_{i/i-1}\} - \frac{1}{2} mN \log(2\pi) \tag{15}$$

where $\mathcal{Y}^N$ represents the total measurement history (the data block), $\tilde{\boldsymbol{y}}_{i/i-1}$ is the innovation process, $\mathcal{N}_{i/i-1}$ is the innovation vector covariance, m is the dimension of the measurement, and N is the number of measurements. The innovation vector and its covariance matrix can be computed from the Kalman filter variables as

$$\tilde{y}_{i/i-1} = y_i - C_i \hat{x}_{i/i-1} \tag{16}$$

and

$$\mathcal{N}_{i/i-1} = C_i P_{i/i-1} C_i^T + R_i \tag{17}$$

where $\hat{x}_{i/i-1}$ is the a priori state estimate (prediction) at time i, and $P_{i/i-1}$ is the a priori estimation error covariance matrix. In accordance with the maximum likelihood estimation principle,[15] the goal of the algorithm which will be described in the sequel is to find the parameter set $\hat{\theta}_{ML}$ that maximizes the log-likelihood function (15). The algorithm presented, called ML/ERA, is a two-layer algorithm, as described next.

Outer Layer

The algorithm starts by assuming a certain system order (within the range predicted by the preliminary analysis) and initial conditions for the optimization process, i.e., initial guesses for the optimization parameters, which, usually, are the process and measurement noise covariance matrices. For the ensuing presentation, we need the following definitions.

Let Ω_1 be the set of all possible system orders, and Ω_2 the space of initial conditions for the optimization control variables. Define Ω to be the Cartesian product

$$\Omega = \Omega_1 \times \Omega_2 \tag{18}$$

Then, every element $\omega \in \Omega$ is the pair

$$\omega := \{\text{system order; optimization parameters initial conditions}\} \tag{19}$$

For each $\omega \in \Omega$, the inner layer of the algorithm computes, in the manner to be described next, estimates of admissible (i.e., minimal and internally balanced) realizations and noise covariance matrices, i.e.,

$$\hat{\theta}(\omega) \equiv \{(\hat{A}(\omega), \hat{B}(\omega), \hat{C}(\omega)), (\hat{Q}(\omega), \hat{R}(\omega))\} \tag{20}$$

Denote the set of all such estimates by Θ, i.e.,

$$\Theta = \{\hat{\theta}(\omega) | \omega \in \Omega\} \tag{21}$$

Then, the outer layer searches for the maximum likelihood optimal parameter estimate, denoted by $\hat{\theta}_{ML}$

$$\hat{\theta}_{ML} = \arg \max_{\Theta} \mathcal{L}[\hat{\theta}(\omega) | \mathcal{Y}^N] \tag{22}$$

where $\mathcal{L}[\hat{\theta}(\omega) | \mathcal{Y}^N]$ is the log-likelihood function (15) pertaining to the model (11).

Inner Layer

The inner layer is based on an idea similar to that of the extended least squares algorithm,[10] namely, a boot-strap combination of a parameter identifier which operates on filtered data, and a state estimator that filters measurements given an identified system. The layer is comprised of the ERA algorithm, which provides a realization $(\hat{A}, \hat{B}, \hat{C})$ per given smoothed measurements, and a stochastic adaptive filter/fixed-interval state smoother, which estimates the process and measurement noise covariance matrices and smooths the measurements per given realization. The adaptive filter is a maximum likelihood identification scheme, whereas the fixed-interval state smoother may be taken as any of the known smoothing algorithms.[11-13] The smoother used in this study was the square-root algorithm recently introduced in Ref. 14. For a given realization, the ML scheme finds the covariance matrices $\hat{Q}$ and $\hat{R}$ which maximize the log-likelihood function. Then, for the optimal $\hat{Q}$ and $\hat{R}$, the fixed-interval smoother smooths the measurements using the realization $(\hat{A}, \hat{B}, \hat{C})$ and the covariances $\hat{Q}$ and $\hat{R}$. The inner layer works iteratively in the following manner.

1) Let i be an index indicating the iteration number. Applying the ERA to the original, unsmoothed measurements, yields an initial triple $(\hat{A}, \hat{B}, \hat{C})_i$ for i = 0. Also, assume some initial value for the optimization control variables, the process and measurement covariance matrices $(\hat{Q}, \hat{R})_i$, $i = 0$. If either one of these matrices is known, then its known value should be used. The initial values thus selected are denoted by

$$\hat{\theta}_0 = \{(\hat{A}, \hat{B}, \hat{C})_0, (\hat{Q}, \hat{R})_0\} \tag{23}$$

2) For the current triple $(\hat{A}, \hat{B}, \hat{C})_i$, numerically search for the optimal values of $(\hat{Q}, \hat{R})$ that maximize the log-likelihood function. This function is computed by implementing a Kalman filter which is based on the triple $(\hat{A}, \hat{B}, \hat{C})_i$. The optimal values of the covariances thus found are denoted by $(\hat{Q}, \hat{R})_{i+1}$.

3) For the realized triple $(\hat{A}, \hat{B}, \hat{C})_i$ and the covariances $(\hat{Q}, \hat{R})_{i+1}$, smooth the measurements via a fixed-interval state smoother. Denote the smoothed measurements by $\hat{\mathcal{Y}}^N_{i+1}$ (note that $\hat{\mathcal{Y}}^N_0 := \mathcal{Y}^N$).

4) Feeding the smoothed measurements $\hat{\mathcal{Y}}^N_{i+1}$ to the ERA yields a new realization, denoted by $(\hat{A}, \hat{B}, \hat{C})_{i+1}$. We have thus obtained a new estimate for the parameter vector θ

$$\hat{\theta}_{i+1} = \{(\hat{A}, \hat{B}, \hat{C})_{i+1}, (\hat{Q}, \hat{R})_{i+1}\} \tag{24}$$

5) Compute the current value of the log-likelihood function $\mathcal{L}(\hat{\theta}_{i+1} | \mathcal{Y}^N)$, by using the current value for the triple $(\hat{A}, \hat{B}, \hat{C})_{i+1}$ and the covariances $(\hat{Q}, \hat{R})_{i+1}$.

6) Assuming that $\mathcal{L}(\hat{\theta}_{i+1} | \mathcal{Y}^N) > \mathcal{L}(\hat{\theta}_i | \mathcal{Y}^N)$, check for convergence, by testing the criterion

$$\mathcal{L}(\hat{\theta}_{i+1} | \mathcal{Y}^N) - \mathcal{L}(\hat{\theta}_i | \mathcal{Y}^N) \leq \epsilon \tag{25}$$

where ϵ is a prespecified convergence criterion. If convergence has been achieved, stop the inner layer process and yield the resulting estimate, $\hat{\theta}(\omega) := \hat{\theta}_{i+1}$ to the outer layer. Otherwise, return to stage 2 and continue until convergence has been reached. In case $\mathcal{L}(\hat{\theta}_{i+1} | \mathcal{Y}^N) < \mathcal{L}(\hat{\theta}_i | \mathcal{Y}^N)$, this usually means that the optimization process ran into numerical problems. In this case choose $\hat{\theta}(\omega) := \hat{\theta}_i$.

For the reader's convenience, the two-layer algorithm's flow chart is shown in Fig. 1.

Remarks:

1) Selection of optimization parameters. As mentioned, the optimization carried out in the inner layer searches, in principle, for the optimal covariance matrices Q and R. In practice, it was found that using diagonal matrices (i.e., using a relatively small number of optimization parameters) suffices for obtaining satisfactory results, at the same time maintaining a reasonable computational load. However, the number and type of optimization parameters should be decided by the user according to the specific characteristics of the case at hand and the available computational power (e.g., the user may wish to optimize over only some of the covariance matrix elements). Also, note that in case the measurement noise covariance matrix is considered known, this matrix can be kept "frozen" during the optimization process, or it can be left as an optimization parameter during the initial stages of the optimization (when the estimated observation matrix C cannot be considered very accurate).

2) Sensitivity to choice of initial conditions. As is well known, the results of every complex, nonlinear optimization process may depend, to some extent, on the choice of initial conditions. In the presented algorithm, this is taken care of in the outer layer by maximizing the log-likelihood function over any chosen set of initial conditions. However, it should be noted that in numerous computer simulations which were run, it was observed that the algorithm is very insensitive to the choice of initial conditions.

3) Statistical properties of the ML estimator. The ML/ERA has good asymptotic statistical properties, which follow from

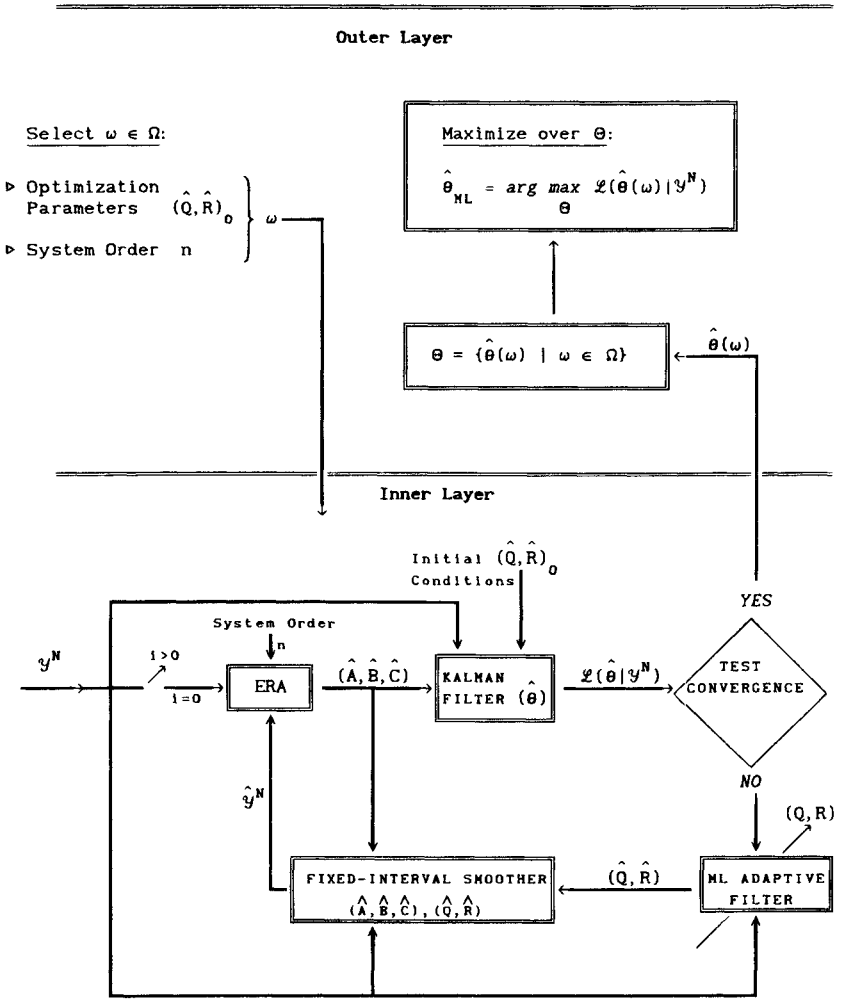

Fig. 1 Flow chart of the ML/ERA two-layer scheme.

its being a maximum likelihood algorithm. Thus, the new identification scheme is guaranteed to be consistent and asymptotically efficient, assuming that the data blocks are independent and identically-distributed (iid).[15]

4) Statistical evaluation of the estimated parameters. One of the merits of the new algorithm is the ability to assess the quality of the results by performing statistical tests on the consistency of the state estimator which is based on the identified realization. Define the following normalized innovation statistic:

$$\epsilon_i = \tilde{\mathbf{y}}^T_{i/i-1} \mathcal{N}^{-1}_{i/i-1} \tilde{\mathbf{y}}_{i/i-1} \tag{26}$$

Here, $\tilde{\mathbf{y}}_{i/i-1}$ is the innovation process and $\mathcal{N}_{i/i-1}$ is the innovation covariance. These variables are computed via the Kalman filter, which is run as a part of the inner layer computational scheme. Assuming that the process and measurement noises are Gaussian and the Kalmanfilter is statistically consistent, the innovation process is white, zero mean, and Gaussian. Hence, the statistic ϵ_i is a χ^2 random variable with m degrees of freedom ($\tilde{\mathbf{y}}_{i/i-1} \in \mathcal{R}^m$). This yields a 1-$\alpha$ confidence interval as a function of m and α, i.e.,

$$\text{Prob}\{\epsilon_i \in [r_{\alpha/2}, r_{1-\alpha/2}]\} = 1 - \alpha \tag{27}$$

This confidence interval may be used at each stage of the identification process to test the validity of the results by checking the consistency of the Kalman filter which is involved in the computation of the log-likelihood function. This procedure is demonstrated in example 3, in the sequel.

5) Estimation of the system order. Note that the system order is estimated via the outer layer of the algorithm by maximizing the log-likelihood function over the estimated range of system orders. Moreover, although an a priori knowledge of the system order is not necessary, such a knowledge can be easily utilized to reduce the work load of the estimator, if available.

V. Numerical Examples

In this section we present three numerical examples that demonstrate the performance of the new ML/ERA technique.

Example 1

The system considered in the first example is taken from Ref. 16. The system is a lumped-mass, damped, beam-like structure that has three modes (see Fig. 2). The input to the system is the force acting on the lower mass, and the displacements of the other two masses are measured. The true system model is given by Eq. (11), with

$$A = \text{diag}\left\{\begin{bmatrix} 0.9856 & 0.1628 \\ -0.1628 & 0.9856 \end{bmatrix}, \begin{bmatrix} 0.8976 & 0.4305 \\ -.4305 & 0.8976 \end{bmatrix}, \begin{bmatrix} 0.8127 & 0.5690 \\ -.5690 & 0.8127 \end{bmatrix}\right\} \tag{28a}$$

$$B = [0.0011 \quad 0.0134 \quad -0.0016 \quad -0.0072 \quad 0.0011 \quad 0.0034]^T \tag{28b}$$

$$C = \begin{bmatrix} 1.5119 & 0.0 & 2.0 & 0.0 & 1.5119 & 0.0 \\ 1.3093 & 0.0 & 0.0 & 0.0 & -1.3093 & 0.0 \end{bmatrix} \tag{28c}$$

$$G = I_6 \tag{28d}$$

The system was simulated with the following noise covariance matrices:

$$Q = 1.0E-6 \times I_6; \qquad R = 4.0E-2 \times I_2 \tag{29}$$

For these covariances, the ratio of the measurement noise to the measured signal is about 3%. In a typical run, the singular values of the 60 × 60 generalized Hankel matrix were

155.30 145.20 65.11 62.11 24.45 23.97 3.85 3.66

whereas without any noises (i.e., $Q = R = 0$), the following singular values are obtained:

154.98 144.67 63.64 60.50 23.76 23.23 1.0E − 15 1.0E − 15

Note, that the (wrong) interpretation of the last two additional singular values as another, fourth-system mode cannot be easily rejected.

When the new algorithm was run, a varying system order was assumed in the outer layer, in the range of $n \in (2, 4, 6, 8, 10)$. For each system order 10 data blocks were processed. The average, over these 10 blocks, of the maximal value of the log-likelihood function, denoted by $\max \mathcal{L}(\theta|\mathcal{Y}^N)$, is shown in Table 1. From this table, it is clear that the order of the minimal realization is 6, since the assumption of a higher system order does not yield a larger log-likelihood value. Note that using the log-likelihood function it became much easier to estimate the true system order (which, in the maximum likelihood sense, is the order of the system that explains best the given measurements).

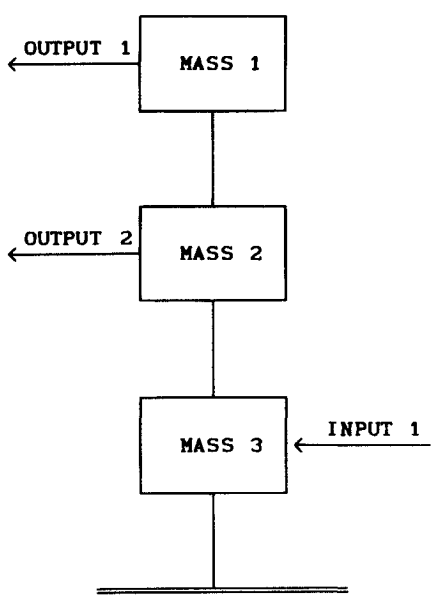

Fig. 2 Structural arrangement, example 1.

In Table 2 we compare the results of 27 Monte Carlo runs of the ML/ERA algorithm with similar results obtained via the ERA algorithm. The table shows the average values of the estimated damping ratios and the standard deviation of the estimates. As can be observed, using the new algorithm results in a noticeable improvement in the quality of the estimates in terms of both the average values and the standard deviations.

Example 2

In this example we consider the three-mode system analyzed in Ref. 6. The system natural frequencies are 1 rad/s, 2.76 rad/s, and 5.4 rad/s, and the damping of each mode is zero. The impulse response of the system is given by

$$z(t) = 1.0 \sin(t) + 0.05 \sin(2.76t) + 0.001 \sin(5.4t) \tag{30}$$

Using a discretization interval of 0.05 s the discretized measurement, contaminated by a measurement noise, is

$$y(k) = z(t_k) + v(k) \tag{31}$$

The measurement noise v is zero mean, Gaussian distributed, and with known variance.

The singular values of the 50 × 50 generalized Hankel matrix for the certainty equivalent system ($Q = R = 0.0$) are compared with those of the real system with $Q = 0.0$ and $R = 0.01$ in Table 3. Note that whereas the first mode is highly controllable/observable, the other two modes are very weakly controllable/observable. Using the singular values as an indication, it is very difficult, if not impossible, to determine the true system order.

The ML/ERA algorithm was run for system orders $n = 2$, 4, 6, and 8 in the outer layer. For each assumed order the maximal value of the log-likelihood function was computed and averaged over 10 independent data blocks. The average maximal log-likelihood, $\max \mathcal{L}(\Theta|\mathcal{Y}^N)$, is shown in Table 4 as am function of the assumed system order.

From Table 4 it is clear that the true system order is 6 (assuming that the order is 8 actually decreases the log-likelihood value). Moreover, observe that the second and third mode hardly contribute to the log-likelihood value, which is a strong indication that these modes are very weakly controllable and observable (and, therefore, could potentially be deleted if a model order reduction should be needed).

Table 1 Average maximal log-likelihood function values

System order	$\max \mathcal{L}(\Theta\|\mathcal{Y}^N)$
2	−5380
4	−5308
6	1645
8	1641
10	1638

Table 2 Estimated damping ratios

	Average damping		Standard deviation	
True value	ERA	ML/ERA	ERA	ML/ERA
0.638E-02	0.77E-02	0.638E-02	0.168E-02	0.038E-02
1.008E-02	1.30E-02	1.025E-02	0.018E-01	0.004E-01
1.300E-02	2.89E-02	1.40E-02	0.070E-01	0.012E-01

Table 3 Generalized Hankel matrix singular values

$Q = 0.0$ $R = 0.0$	$Q = 0.0$ $R = 0.01$
29.29	29.63
20.15	20.14
1.27	1.63
0.847	1.60
0.0257	1.31
0.0214	1.16
-	1.14
-	1.11

Table 4 Average maximal log-likelihood function values

System order	max $\mathcal{L}(\theta\|\mathcal{Y}^N)$
2	780
4	787
6	789
8	767

Table 5 Estimated modal frequencies

True value	Average Frequency		Standard Deviation	
	ERA/MME	ML/ERA	ERA/MME	ML/ERA
1.00	0.9972	1.0002	0.0046	0.0021
2.76	2.7284	2.762	0.1098	0.068
5.40	5.4254	5.12	0.3570	0.65

Table 6 Modal damping ratios

Mode no.	True value	Estimates	
		Average	St. dev.
1	0.00	8.66*E*-04	3.93*E*-03
2	0.00	1.78*E*-03	3.44*E*-02
3	0.00	3.07*E*-02	2.91*E*-02

The estimated frequencies obtained via the new algorithm are compared to those obtained via the ERA/MME algorithm in Table 5. The table also shows the standard deviations of the estimates. The ML/ERA results shown are based on 18 Monte Carlo runs. The ERA/MME results are taken from Ref. 6. The size and structure of the generalized Hankel matrix used by both algorithms were identical. The covariance matrices were $Q = 0.00$ and $R = 0.01$, which represent highly contaminated measurements (note that the standard deviation of the measurement noise is 100 times larger than the third-mode amplitude). Since the initial realization obtained via the ERA from the measurements could identify only the first mode, the MME algorithm used in Ref. 6 was run in the first ERA/MME iteration assuming a single-mode system only; then, a three-mode realization was obtained from the ERA and the second MME iteration used a three-mode dynamics. For the sake of comparison, the same strategy was adopted using the ML/ERA algorithm. It must be noted that since the results shown were computed in different computer environments, only qualitative conclusions may be drawn from this comparison.

As can be observed from Table 5, the first two frequencies were identified somewhat better by the ML/ERA algorithm, in terms of both the average frequency estimates and the standard deviations, whereas the ERA/MME performed remarkably in identifying the third modal frequency. In Table 6 we show the average identified damping ratios and the standard deviations of these estimates. The ML/ERA algorithm performed well in identifying the modal damping ratios. Note that Ref. 6 reports much worse estimates for the third modal damping.

Example 3

The system considered in this example is the short period mathematical model of the F-89 aircraft.[17] For the simplified short period model, the state variables are the pitch angle and the pitch rate. The aircraft is controlled by the elevator, and the pitch rate is measured. The true values of the frequency and damping of the short period mode are 4.228 rad/s and 0.383, respectively. The discrete-time state space model, obtained for a discretization step of 0.005 s, is

$$A = \begin{bmatrix} 0.99285 & 1.006 \\ -.000385 & 0.9904 \end{bmatrix}; \quad B = \begin{bmatrix} 0.10635 \\ 0.1305 \end{bmatrix} \tag{32}$$

$$C = [0.0 \quad 0.2]; \quad G = I_2$$

Table 7 compares the identification results obtained via the ERA method and the ML/ERA algorithm. The study is based on the results of 60 Monte Carlo simulations. In all runs, the process noise and measurement noise covariance matrices were $Q = 1.0E - 06 \times I_2$ and $R = 0.1$, respectively. These values represent a measurement noise amplitude of about 20% of the maximum amplitude of the measured signal (which, in reality, would be considered a nonacceptable measurement). In only 29 of the cases was the ERA successful in identifying a transition matrix with a complex pair of eigenvalues. In all other cases, two real eigenvalues were identified. The comparison is based, therefore, on the results of the 29 successful ERA runs. On the other hand, the ML/ERA algorithm was able to identify the short period mode in all 60 runs. The average values of the frequencies and damping ratios for the ERA and ML/ERA are shown in Table 7, along with the standard deviations of the estimates. The improvement achieved by the ML/ERA method is substantial, both in terms of the estimated values and the smaller standard deviations.

To assess the quality of the estimated parameters, the statistical testing technique (outlined in Sec. IV) was used. In the example presented, 500 samples of the normalized innovation

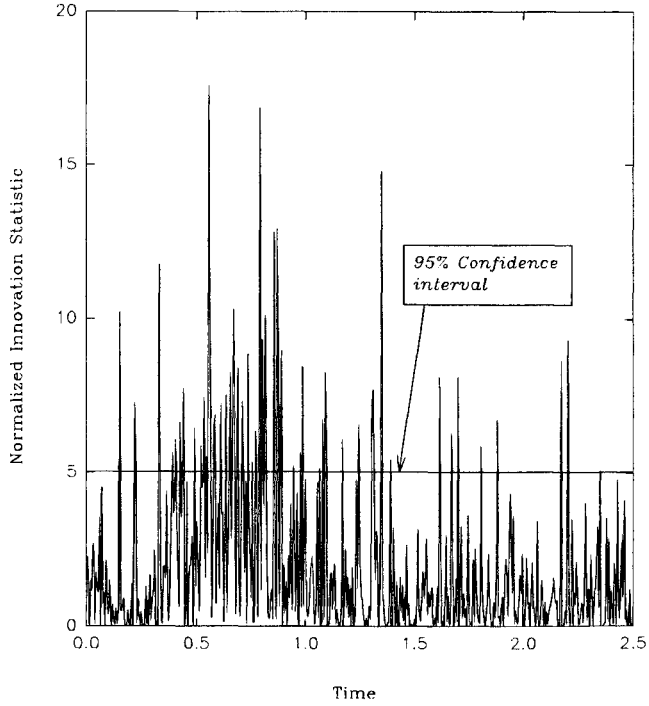

Fig. 3 Consistency test for a Kalman filter based on the system identified via the ERA algorithm.

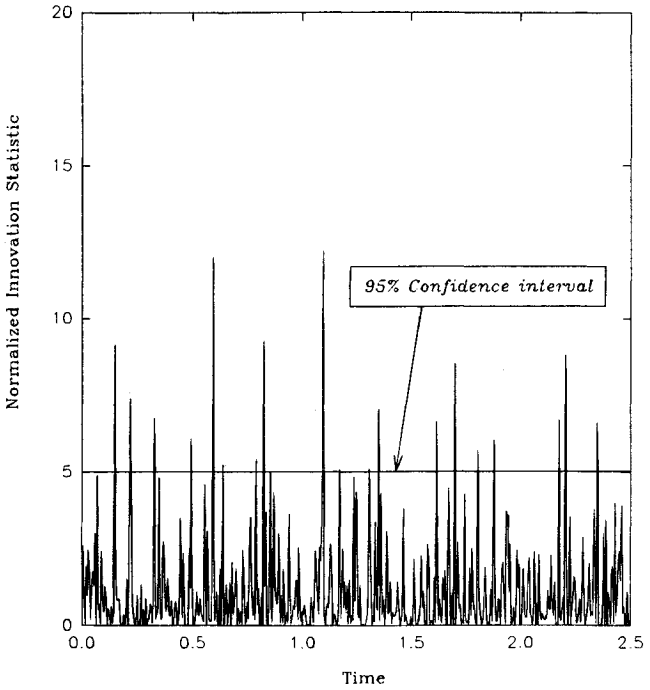

Fig. 4 Consistency test for a Kalman filter based on the system identified via the ML/ERA algorithm.

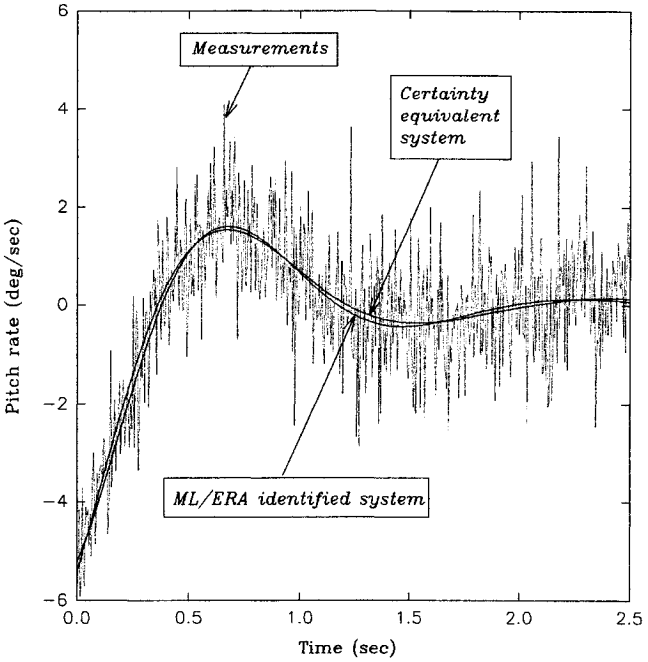

Fig. 5 F-89 short-period pitch rate impulse response.

Table 7 F-89 Short period frequency and damping ratio

		Average		Standard Deviation	
Parameter	True	ERA[a]	ML/ERA[b]	ERA[a]	ML/ERA[b]
Frequency	4.228	8.922	4.196	2.445	0.320
Damping ratio	0.383	0.821	0.410	0.135	0.052

[a]Based on 29 out of 60 Monte Carlo runs.
[b]Based on all 60 Monte Carlo runs.

statistic (26) were tested for $\alpha = 0.05$ (i.e., a 95% confidence interval was chosen). Figure 3 graphically shows the χ^2 test performed for the Kalman filter based on the matrix triple realized by the ERA, whereas Fig. 4 shows corresponding results for the system realized by the ML/ERA algorithm (note that although the statistical test performed was a two-sided test, only the upper bound of the confidence interval is shown in these figures, since the lower bound is too close to zero). Of the samples in Fig. 3, 14% are outside the confidence interval, which means that the Kalman filter which is based on the ERA-realized triple is significantly statistically inconsistent. This result indicates that, for this case, the estimation quality of the ERA-realized triple is poor. On the other hand, the ML/ERA based Kalman filter is much more consistent, since only 6.4% of the points in Fig. 4 are outside the 95% confidence interval. We conclude, therefore, that the ML/ERA algorithm yielded higher quality estimates.

Finally, Fig. 5 shows the time histories of the measured (noisy) impulse response of the system, the impulse response of the corresponding deterministic system (the certainty-equivalent system), and the impulse response of the ML/ERA-realized system. It is easy to observe that the ML/ERA method successfully filtered out most of the noise in the measurements, thus yielding an accurate identification of the true (i.e., certainty equivalent) system.

IV. Conclusions

A new maximum likelihood (ML) realization/identification technique has been introduced, that is based on the eigensystem realization algorithm (ERA) of Juang and Pappa and significantly extends its capability to deal with stochastic systems. The ERA algorithm serves for the purpose of producing the realization, whereas a combination of an adaptive ML filter and a fixed-interval state smoother is used for estimating the noise statistics and smoothing the measurements. The quality of the resulting identified system model can be evaluated via standard statistical consistency tests which can be performed on a Kalman filter based on these parameters. Being a maximum likelihood estimator, the new algorithm is consistent and asymptotically efficient, assuming that the data blocks are independent and identically distributed.

The performance of the new method was demonstrated via three numerical examples. It was shown that the algorithm significantly improves the capability of the ERA to deal with high noise stochastic systems. However, it is felt that a more thorough investigation is needed, that will include an application of the algorithm to a "real life" problem.

Acknowledgments

This research was supported by the Technion V.P.R.-J. and J. Gringorten Aeronautical Research Fund and by the Fund for the Promotion of Research at the Technion.

References

[1]Juang, J.-N., and Pappa, R. S.,"An Eigensystem Realization Algorithm (ERA) for Modal Parameter Identification and Model Reduction," *Journal of Guidance, Control, and Dynamics,* Vol. 8, No. 5, 1985, pp. 620–627.

[2]Ho, B. L., and Kalman, R. E., "Effective Construction of Linear

State Space Models from Input/Output Data," *Proceedings of 3rd Allerton Conference on Circuits and System Theory,* 1965, pp. 449–459.

[3] Juang, J.-N., and Pappa, R. S., "Effects of Noise on Modal Parameters Identified by the Eigensystem Realization Theory," *Journal of Guidance, Control, and Dynamics,* Vol. 9, No. 3, 1986, pp. 294–303.

[4] Klema, V. C., and Laub, A. J., "The Singular Value Decomposition: Its Computation and Some Applications," *IEEE Transactions on Automatic Control,* Vol. AC-25, No. 2, 1980, pp. 164–176.

[5] Longman, R. W., Bergmann, M., and Juang, J.-N., "Variance and Bias Confidence Criteria for ERA Modal Parameter Identification," *Proceedings of the AIAA/AAS Astrodynamics Conference* (Minneapolis, MN), AIAA, Washington, DC, 1988, pp. 729–739 (AIAA Paper 88-4312).

[6] Mook, D. J., and Lew, J.-S., "A Robust Algorithm for System Realization/Identification," *Journal of the Astronautical Sciences,* Vol. 38, No. 2, 1990, pp. 229–243.

[7] Mook, D. J., and Junkins, J. L., "Minimum Model Error Estimation for Poorly Modeled Dynamic Systems," *Journal of Guidance, Control, and Dynamics,* Vol. 11, No. 3, 1988, pp. 256–261.

[8] Roemer, M. J., and Mook, D. J., "Robust Modal Identification/Estimation of the Mini-Mast Testbed," *Journal of Guidance, Control, and Dynamics,* Vol. 15, No. 3, 1992, pp. 642–647.

[9] Mendel, J. M., *Lessons in Digital Estimation Theory,* Prentice-Hall, Englewood-Cliffs, NJ, 1987, pp. 258–260.

[10] Goodwin, G. C., and Sin, K. S., *Adaptive Filtering Prediction and Control,* Prentice-Hall, Englewood Cliffs, NJ, 1984, pp. 319–323.

[11] Rauch, H. E., Tung, F., and Striebel, C. T., "Maximum Likelihood Estimates of Linear Dynamic Systems," *AIAA Journal,* Vol. 3, No. 8, 1965, pp. 1445–1450.

[12] Bierman, G. J., "A New Computationally Efficient Fixed-interval, Discrete-time Smoother," *Automatica,* Vol. 19, No. 5, 1983, pp. 503–511.

[13] McReynolds, S. R., "Covariance Factorization Algorithms for Fixed-Interval Smoothing of linear Discrete Dynamic Systems," *IEEE Transactions on Automatic Control,* Vol. 35, No. 10, 1990, pp. 1181–1183.

[14] Oshman, Y., "A New, Factorized, Fixed-Interval Smoother," *Proceedings of the AIAA Guidance, Navigation and Control Conference* (Hilton Head, SC), AIAA, Washington, DC, 1992, pp. 1255–1263 (AIAA Paper 92-4595).

[15] Sorenson, H. W., *Parameter Estimation: Principles and Problems,* Marcel Dekker, New York, 1980, pp. 187–193.

[16] Chen, C. W., Huang, J. K., Phan, M., and Juang, J.-N., "Integrated System Identification and State Estimation for Control of Flexible Space Structures," *Journal of Guidance, Control, and Dynamics,* Vol. 15, No. 1, 1992, pp. 88–95.

[17] McRuer, D., Ashkenas, I., and Graham, D., *Aircraft Dynamics and Automatic Control,* Princeton Univ. Press, Princeton, NJ, 1973, pp. 689–692.