Identification of Observer/Kalman Filter Markov Parameters: Theory and Experiments

Jer-Nan Juang,* Minh Phan,† Lucas G. Horta,‡ and Richard W. Longman§
NASA Langley Research Center, Hampton, Virginia 23665

This paper discusses an algorithm to compute the Markov parameters of an observer or Kalman filter from experimental input and output data. The Markov parameters can then be used for identification of a state-space representation, with associated Kalman or observer gain, for the purpose of controller design. The algorithm is a nonrecursive matrix version of two recursive algorithms developed in previous works for different purposes, and the relationship between these other algorithms is developed. The new matrix formulation here gives insight into the existence and uniqueness of solutions of certain equations and offers bounds on the proper choice of observer order. It is shown that if one uses data containing noise and seeks the fastest possible deterministic observer, the deadbeat observer, one instead obtains the Kalman filter, which is the fastest possible observer in the stochastic environment. The results of the paper are demonstrated in numerical studies and experiments on the Hubble space telescope.

Introduction

MANY future spacecraft such as the space station will be large and flexible and require control of the vibrational motion induced by internal and external disturbances for fine pointing and shape control. One can classify controllers for flexible structures into two types: model-independent and model-dependent controllers. The model-independent controller[1] is attractive because it provides stability without precise knowledge of the system. However, it generally produces low-authority control that may not meet the performance requirements. On the other hand, model-dependent controllers require an accurate model to achieve high performance, but inaccuracies in the model may result in instability of the controlled system.[2] Current results indicate that an accurate model is necessary to design controllers with the needed performance level.

In the past decade, many system identification techniques were developed and/or applied to identify a state-space model for modal parameter identification of large flexible space structures. The modal parameters include frequencies, damping, and mode shapes. The identified state-space model is also used in controller design. Many satisfactory results were reported in the literature.[3,4] Most techniques are based on sampled pulse or impulse system response histories that are known as Markov parameters. The usual practice uses the fast Fourier transforms (FFT) of the inputs and measured outputs to compute the sampled pulse response histories. The discrete nature of the FFT causes one to obtain pulse response rather than an impulse response, and a somewhat rich input is required to prevent numerical ill-conditioning in the computation. An-

other approach is to solve directly in the time domain for the Markov parameters from the input and output data. The drawbacks of this method include the need to invert an input matrix that necessarily becomes particularly large for lightly damped systems.[5]

Recently, an approach was developed[6-8] to address the problem of inverting a large-dimensional input matrix. Reference 8 gives a more detailed presentation of the developments in Refs. 6 and 7 including additional examples. Rather than identifying the system Markov parameters that may exhibit very slow decay, it uses an asymptotically stable observer to form a stable state-space discrete model for the system to be identified. The primary purpose of introducing an observer in Ref. 8 is as an artifice to compress the data and improve system identification results in practice. The system identification engineeer can assign any poles desired and hence specify the decay rate of the observer Markov parameters to be determined from the data and simultaneously the number of parameters needed before they have decayed to a negligible level. The desired poles can be real, complex, or deadbeat. The deadbeat means that all the poles are zero in the complex plane for a discrete model.

The treatment in Ref. 8 is purely deterministic. When stochastic models are considered, it would be desirable to identify not only the system matrices of a realization, but also the noise or uncertainty characteristics of the model directly from the experimental data. This presumes that the same sensors and actuators used in the identification tests will also be employed in the control system that is to be designed from the system identification results. There are basically two ways to characterize system uncertainties, including plant and measurement noises. One method is to describe the input and output uncertainties directly in terms of their covariances. Another way is to specify the Kalman filter equation with its steady-state Kalman gain that is a function of the input and output covariances. Recently, a recursive identification method was presented in Ref. 9 to identify Markov parameters for the identification of not only the system matrices, but also the Kalman filter gain. Note that the work in Ref. 9 was motivated by the unsolved problem in Ref. 10 in which single-mode projection filters were developed for modal parameter identification. There exist many unsolved issues, such as the relationship between the order of the Kalman filter and that of the system using the approach of Ref. 9. Furthermore, the computation time and data length are too long to become attractive in practice. Examination of the mathematics involved in Ref. 8 for accelerated identification using observers, and that in Ref. 9 for direct identification of the Kalman filter

*Principal Scientist, Spacecraft Dynamics Branch. Fellow AIAA.
†National Research Council Research Associate, Spacecraft Dynamics Branch; currently Senior Engineer, Lockheed Engineering and Sciences Company, Hampton, VA.
‡Aerospace Engineer, Spacecraft Dynamics Branch.
§National Research Council Senior Research Associate, Spacecraft Dynamics Branch. Also Professor of Mechanical Engineering, Columbia University, New York, NY. Fellow AIAA.

Markov parameters, shows striking parallels that are investigated here.

One objective of this paper is to present an algorithm to directly compute the Markov parameters of a steady-state Kalman filter from experimental data. From these parameters, one can use various methods to obtain the Kalman filter state-space realization. The approach used in this paper is to reformulate in matrix form the equations presented in Ref. 8 for deadbeat observers and in Ref. 9 for Kalman filters. This matrix form gives added insight into the uniqueness of the transformation from observer or filter Markov parameters to the system Markov parameters and also allows the development of upper and lower bounds on the choice of observer order. Also, the recursive least-squares solution in Refs. 8 and 9 is replaced by a nonrecursive least-squares solution. This results in an improved rate of convergence for the Kalman filter identification process by comparison to Ref. 9.

Underlying this work is a second objective, to establish the relationship between the observer identification equations in Ref. 8 and the Kalman filter identification equations in Ref. 9. When the observer poles in Ref. 8 are all placed at the origin in the z plane in order to obtain a deadbeat observer of a sufficiently high order and then data containing both plant and measurement noise are used to develop the desired Markov parameters, the result is the Kalman filter Markov parameters. Stated in different words, if one uses data containing noise and seeks the Markov parameters for the fastest possible deterministic discrete time observer, one instead obtains the Markov parameters of the slower Kalman filter, which is the fastest possible observer in the stochastic environment.

This paper starts by explaining the relationship between the input and output histories in terms of system Markov parameters without any observer. An observer is then introduced into the input and output matrix relation, which is solved by a nonrecursive least-squares approach to compute the observer Markov parameters. Formulations are derived to compute the system Markov parameters and the observer gain from the observer Markov parameters. The relationship between the identified deadbeat observer and a Kalman filter is then established through the use of the ergodic property of stationary random processes. The optimal nature of the identified observer is also discussed. Numerical and experimental results are given to illustrate the validity of the algorithm presented in this paper. The experimental results are obtained from the Hubble space telescope having six gyros and four torque wheels.

Basic Formulation

Consider a discrete multivariable linear system described by

$$x(i+1) = Ax(i) + Bu(i)$$

$$y(i) = Cx(i) + Du(i) \tag{1}$$

where $x(i) \in R^n$, $y(i) \in R^q$, $u(i) \in R^m$. If we assume zero initial conditions $x(0) = 0$, the set of this equations for a sequence of i can be written as

$$\overset{q \times \ell}{y} = \underset{q \times m\ell}{Y} \overset{m\ell \times \ell}{U} \tag{2}$$

where

$$y = \begin{bmatrix} y(0) & y(1) & y(2) & \cdots & y(\ell-1) \end{bmatrix}$$

$$Y = \begin{bmatrix} D & CB & CAB & \cdots & CA^{\ell-2}B \end{bmatrix}$$

and

$$U = \begin{bmatrix} u(0) & u(1) & u(2) & \cdots & u(\ell-1) \\ & u(0) & u(1) & \cdots & u(\ell-2) \\ & & u(0) & \cdots & u(\ell-3) \\ & & & \ddots & \vdots \\ & & & & u(0) \end{bmatrix}$$

Equation (2) is a matrix representation of the relationship between input and output histories. The matrix y is a $q \times \ell$ output data matrix, where q is the number of outputs and ℓ the number of data samples. The matrix Y, of dimension $q \times m\ell$ with m the number of inputs, contains all of the Markov parameters D, CB, CAB, ..., $CA^{\ell-2}B$ to be determined. The matrix U is an $m\ell \times \ell$ upper-block triangular input matrix. It is square in the case of a single-input system and otherwise has more rows than columns.

Inspection of Eq. (2) indicates that there are $q \times m\ell$ unknowns in the Markov parameter matrix, but only $q \times \ell$ equations. For the case where $m > 1$, the solution for Y is not unique. However, it is known that, for a finite-dimensional linear system, Y must be unique. The matrix Y can only be uniquely determined from this set of equations for $m = 1$. Even in this case, if the input has zero initial value, i.e., $u(0) = 0$, or the input signals are not rich enough such as the case with sinusoidal input signals, the matrix U becomes ill-conditioned and thus the matrix $Y = yU^{-1}$ cannot be accurately computed.

Consider the case where A is asymptotically stable so that for some sufficiently large p, $A^i \approx 0$ for all time steps $i \geq p$. Equation (2) can then be approximated by

$$\overset{q \times \ell}{y} \approx \underset{q \times mp}{Y} \overset{mp \times \ell}{U} \tag{3}$$

where

$$y = \begin{bmatrix} y(0) & y(1) & y(2) & \cdots & y(p) & \cdots & y(\ell-1) \end{bmatrix}$$

$$Y = \begin{bmatrix} D & CB & CAB & \cdots & CA^{p-1}B \end{bmatrix}$$

$$U = \begin{bmatrix} u(0) & u(1) & u(2) & \cdots & u(p) & \cdots & u(\ell-1) \\ & u(0) & u(1) & \cdots & u(p-1) & \cdots & u(\ell-2) \\ & & u(0) & \cdots & u(p-2) & \cdots & u(\ell-3) \\ & & & \ddots & \vdots & \cdots & \vdots \\ & & & & u(0) & \cdots & u(\ell-p-1) \end{bmatrix}$$

Note that $U [m(p+1) \times \ell]$ and $Y [q \times m(p+1)]$ refer to truncated versions of U and Y in Eq. (2). Choose the data length ℓ greater than $m(p+1)$, where again m is the number of inputs and p an integer such that $CA^iB \approx 0$ for $i \geq p$. Equation (3) indicates that there are more equations ($q \times \ell$) than unknowns $[q \times m(p+1)]$ because $\ell \geq m(p+1)$. We conclude that if the data have a realization in the form of Eq. (1), then the first p Markov parameters approximately satisfy $Y = yU^+$, where U^+ is the pseudoinverse of the matrix U, and the approximation error decreases as p increases.

Unfortunately, for lightly damped space structures, the integer p and thus the ℓ required to make the approximation in Eq. (3) valid become impractically large in the sense that the size of the matrix U is too large to solve for its pseudoinverse U^+ numerically. The question arises, is there any way to artificially increase the damping of the system in order to allow the solution of Eq. (3) for the Markov parameters? A control engineer will immediately suggest that a feedback loop can be added to make the system as stable as desired. The same effect can be achieved by considering the following algebraic manipulation as presented in Ref. 8.

Add and subtract the term $My(i)$ to the right-hand side of the state equation in Eq. (1) to yield

$$x(i+1) = Ax(i) + Bu(i) + My(i) - My(i)$$

$$= (A + MC)x(i) + (B + MD)u(i) - My(i)$$

or

$$x(i+1) = \bar{A}x(i) + \bar{B}v(i)$$

$$y(i) = Cx(i) + Du(i) \tag{4}$$

where

$$\bar{A} = A + MC$$

$$\bar{B} = [B + MD, \ -M]$$

$$v(i) = \begin{bmatrix} u(i) \\ y(i) \end{bmatrix} \tag{5}$$

and M is an $n \times q$ arbitrary matrix chosen to make the matrix $\bar{A}$ as stable as desired. Although Eq. (4) is mathematically identical to Eq. (1), it is expressed using different system matrices and has a different input. In fact, Eq. (4) is an *observer* equation if the state $x(i)$ is considered an observer state (see Ref. 7 or 8). Therefore, the Markov parameters of the system in Eq. (4) will be referred to as the *observer Markov parameters*. The input-output description in matrix form for Eq. (4) becomes

$$\underset{q \times [(q+m)(\ell-1)+m]}{\overset{q \times \ell}{y}} = \bar{Y} \overset{[(q+m)(\ell-1)+m] \times \ell}{V} \tag{6}$$

where

$$y = \begin{bmatrix} y(0) & y(1) & y(2) & \cdots & y(p) & \cdots & y(\ell-1) \end{bmatrix}$$

$$\bar{Y} = \begin{bmatrix} D & C\bar{B} & C\bar{A}\bar{B} & \cdots & C\bar{A}^{p-1}\bar{B} & \cdots & C\bar{A}^{\ell-2}\bar{B} \end{bmatrix}$$

$$V = \begin{bmatrix} u(0) & u(1) & u(2) & \cdots & u(p) & \cdots & u(\ell-1) \\ & v(0) & v(1) & \cdots & v(p-1) & \cdots & v(\ell-2) \\ & & v(0) & \cdots & v(p-2) & \cdots & v(\ell-3) \\ & & & \ddots & \vdots & \cdots & \vdots \\ & & & & v(0) & \cdots & v(\ell-p-1) \\ & & & & & \ddots & \vdots \\ & & & & & & v(0) \end{bmatrix}$$

Equation (6) is obtained from Eq. (2) by replacing A by $\bar{A}$, B by $\bar{B}$, and u by v except for the first row partition. Because the $n \times q$ matrix M can be arbitrarily chosen, the eigenvalues of $\bar{A}$ may be arbitrarily assigned for an observable system. Reference 8 considers the identification of observer Markov parameters for any chosen observer pole locations for $\bar{A} = A + MC$. The mathematical development here can be interpreted from the point of view of Ref. 8 as attempting to place all the eigenvalues of $\bar{A}$ at the origin, i.e., a deadbeat observer. This provides that $C\bar{A}^i\bar{B} = 0$ for $i \geq p$. When using real data including noise, the eigenvalues of $\bar{A}$ are, in fact, placed such that $C\bar{A}^i\bar{B} \approx 0$ for $i \geq p$, where p is a sufficiently large integer. Alternatively, if $\bar{A}$ represents the state matrix of the Kalman filter including the steady-state Kalman filter gain, the same property is satisfied as used in Ref. 7, which will be discussed in a later section.

When $C\bar{A}^i\bar{B} \approx 0$ for $i \geq p$, one can solve for the observer Markov parameters from real data, using the same approach as in Eq. (3)

$$\underset{q \times [(m+q)p+m]}{\overset{q \times \ell}{y}} = \bar{Y} \overset{[(m+q)p+m] \times \ell}{V} \tag{7}$$

where

$$y = \begin{bmatrix} y(0) & y(1) & y(2) & \cdots & y(p) & \cdots & y(\ell-1) \end{bmatrix}$$

$$\bar{Y} = \begin{bmatrix} D & C\bar{B} & C\bar{A}\bar{B} & \cdots & C\bar{A}^{p-1}\bar{B} \end{bmatrix}$$

$$V = \begin{bmatrix} u(0) & u(1) & u(2) & \cdots & u(p) & \cdots & u(\ell-1) \\ & v(0) & v(1) & \cdots & v(p-1) & \cdots & v(\ell-2) \\ & & v(0) & \cdots & v(p-2) & \cdots & v(\ell-3) \\ & & & \ddots & \vdots & \cdots & \vdots \\ & & & & v(0) & \cdots & v(\ell-p-1) \end{bmatrix}$$

Note that V and $\bar{Y}$ refer to truncated versions of V and $\bar{Y}$ in Eq. (6). Similar to Eq. (3), if the data have a realization in the form of Eq. (1) or its equivalent, Eq. (4), then the first p Markov parameters approximately satisfy $\bar{Y} = yV^+$, where V^+ is the pseudoinverse of the matrix V and the approximation error decreases as p increases. Note that the observer Markov parameters thus identified may not necessarily appear to be asymptotically decaying during the first $p - 1$ steps, although they produce $C\bar{A}^i\bar{B} = 0$ for $i \geq p$ and noise-free data. Reference 8 allows one to place the observer poles to produce more typical asymptotic decay of the observer Markov parameters. To solve for $\bar{Y}$ uniquely, all the rows of V must be linearly independent. Furthermore, to minimize any numerical error due to the computation of the pseudoinverse, the rows of V should be chosen as independent as possible. As a result, the maximum p is the number that maximizes the number $(m + q)p + m$ of independent rows of V. The maximum p means the upper bound of the order of the deadbeat observer. The lower bound of the order of the observer will be addressed in the next section.

There are many ways of producing the least-squares solution to equations such as Eq. (7) for $\bar{Y}$. Reference 5 presents three different approaches to solving equations similar to Eq. (7), including a bootstrapping procedure, the singular-value decomposition, and a recursive algorithm. However, the recursive least-squares algorithm presented in Ref. 8 includes substitution of the desired eigenvalues into Eq. (7) to minimize the number of unknown parameters in $\bar{Y}$. This is probably a very efficient computational procedure. However, it is not obtaining the least-squares solution of Eq. (7), but rather a somewhat modified problem that can be interpreted as a weighted least-squares solution. Disadvantages of the recursive formula include a problem-dependent choice of error sensitivity that requires experiences in least-squares methods.

All of the preceding equations assume zero initial conditions, $x(0) = 0$. For nonzero initial conditions, a somewhat different formula should be used. Rewrite Eq. (4) in another matrix form as

$$\bar{y} = C\bar{A}^p x + \bar{Y}\bar{V} \tag{8}$$

where

$$\bar{y} = \begin{bmatrix} y(p) & y(p+1) & \cdots & y(\ell-1) \end{bmatrix}$$

$$x = \begin{bmatrix} x(0) & x(1) & \cdots & x(\ell-p-1) \end{bmatrix}$$

$$\bar{Y} = \begin{bmatrix} D & C\bar{B} & C\bar{A}\bar{B} & \cdots & C\bar{A}^{p-1}\bar{B} \end{bmatrix}$$

$$\bar{V} = \begin{bmatrix} u(p) & u(p+1) & \cdots & u(\ell-1) \\ v(p-1) & v(p) & \cdots & v(\ell-2) \\ v(p-2) & v(p-1) & \cdots & v(\ell-3) \\ \vdots & \vdots & \vdots & \ddots \\ v(0) & v(1) & \cdots & v(\ell-p-1) \end{bmatrix}$$

For the case where $\bar{A}^p$ is sufficiently small and all the states in x are bounded, Eq. (8) can be approximated by neglecting the first term on the right-hand side

$$\bar{y} = \bar{Y}\bar{V} \tag{9}$$

which has the following least-squares solution

$$\bar{Y} = \bar{y}\bar{V}^T[\bar{V}\bar{V}^T]^{-1} \quad \text{or} \quad \bar{Y} = \bar{y}\bar{V}^T \tag{10}$$

provided that $[\bar{V}\bar{V}^T]^{-1}$ exists; otherwise $\bar{V}^T[\bar{V}\bar{V}^T]^{-1}$ should be replaced by $\bar{V}^T$. Equation (9) is identical to Eq. (7), except that the y in Eq. (7) is replaced by $\bar{y}$ and V by $\bar{V}$. The matrices $\bar{y}$ and $\bar{V}$ are subsets of y and V, respectively, produced by deleting the first p columns. For nonzero unknown initial con-

ditions, Eq. (9) must be used in order to eliminate the effect of initial conditions, because the initial conditions become negligible when they are multiplied by $\bar{A}^p$. In other words, the initial conditions have negligible influence on the measured data after p time steps. When there is both system and measurement noise present, the elimination of initial condition dependence makes the system response become stationary, a fact that is used later to obtain the steady-state Kalman filter gain.

Computation of Actual System Markov Parameters and Observer Gain

To recover the system Markov parameters in Y from the observer Markov parameters in $\bar{Y}$, partition $\bar{Y}$ such that

$$\bar{Y} = \begin{bmatrix} \bar{Y}_{-1} & \bar{Y}_0 & \bar{Y}_1 & \cdots & \bar{Y}_{p-1} \end{bmatrix} \tag{11}$$

where

$$\bar{Y}_k = C\bar{A}^k\bar{B}$$

$$= \begin{bmatrix} C(A + MC)^k(B + MD), & -C(A + MC)^kM \end{bmatrix}$$

$$\equiv \begin{bmatrix} \bar{Y}_k^{(1)}, & \bar{Y}_k^{(2)} \end{bmatrix}; \qquad k = 0,1,2,\ldots$$

$$\bar{Y}_{-1} = D$$

Note that the Markov parameter $\bar{Y}_{-1}$ has a smaller dimension than the remaining Markov parameters. From the second equation in Eq. (11), the Markov parameter CB of the system is simply

$$Y_0 = CB = C(B + MD) - (CM)D$$

$$= \bar{Y}_0^{(1)} + \bar{Y}_0^{(2)}D \tag{12}$$

To obtain the Markov parameter CAB, first consider the product $\bar{Y}_1^{(1)}$

$$\bar{Y}_1^{(1)} = C(A + MC)(B + MD)$$

$$= CAB + CMCB + C(A + MC)MD$$

Hence,

$$Y_1 = CAB$$

$$= \bar{Y}_1^{(1)} + \bar{Y}_0^{(1)}Y_0 + \bar{Y}_1^{(1)}D \tag{13}$$

Similarly, to obtain the Markov parameter CA^2B, consider the product $\bar{Y}_2^{(1)}$

$$\bar{Y}_2^{(1)} = C(A + MC)^2(B + MD)$$

$$= C(A^2 + MCA + AMC + MCMC)(B + MD)$$

$$= CA^2B + CMCAB + C(A + MC)MCB$$

$$+ C(A + MC)^2MD$$

Therefore,

$$Y_2 = CA^2B$$

$$= \bar{Y}_2^{(1)} - CMCAB - C(A + MC)MCB$$

$$- C(A + MC)^2MD$$

$$= \bar{Y}_2^{(1)} + \bar{Y}_0^{(2)}Y_1 + \bar{Y}_1^{(2)}Y_0 + \bar{Y}_2^{(2)}D \tag{14}$$

As established in Ref. 8, the general relationship between the actual system Markov parameters and the observer Markov parameters is

$$Y_k = \bar{Y}_k^{(1)} + \sum_{i=0}^{k-1} \bar{Y}_i^{(2)}Y_{k-i-1} + \bar{Y}_k^{(2)}D \tag{15}$$

Knowledge of the actual system Markov parameters allows one to obtain a state-space realization of the system of interest. Modal parameters including natural frequencies, damping ratios, and mode shapes can then be found. Note that there are only $p + 1$ observer Markov parameters computed as a least-squares solution from Eq. (7). By the choice of p, $\bar{Y}_k^{(1)}$ and $\bar{Y}_k^{(2)}$ are considered to be zero for $k > p$.

The relationship between observer Markov parameters and system Markov parameters can be further developed as follows. Let the matrices H, $\bar{Y}^{(2)}$, and $\underline{Y}$ be defined as

$$\bar{Y}^{(2)} = \begin{bmatrix} -\bar{Y}_{p-1}^{(2)} & -\bar{Y}_{p-2}^{(2)} & -\bar{Y}_{p-3}^{(2)} & \cdots & -\bar{Y}_0^{(2)} \end{bmatrix}$$

$$H = \begin{bmatrix} Y_1 & Y_2 & Y_3 & \cdots & Y_N \\ Y_2 & Y_3 & Y_4 & \cdots & Y_{N+1} \\ Y_3 & Y_4 & Y_5 & \cdots & Y_{N+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_p & Y_{p+1} & Y_{p+2} & \cdots & Y_{N+p-1} \end{bmatrix} \tag{16}$$

and

$$\underline{Y} = \begin{bmatrix} Y_{p+1} & Y_{p+2} & Y_{p+3} & \cdots & Y_{p+N} \end{bmatrix}$$

where N is a sufficiently large arbitrary integer and H is obviously a generalized Hankel matrix consisting of a number of system Markov parameters. From Eqs. (16), one obtains

$$\bar{Y}^{(2)}H = \underline{Y} \tag{17}$$

By using the definition of system Markov parameters, the Hankel matrix can be expressed by

$$H = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{P-1} \end{bmatrix} A \begin{bmatrix} B & AB & A^2B & \cdots & A^{N-1}B \end{bmatrix} = PAQ \tag{18}$$

where A is the system state matrix for a discrete model representation, P the observability matrix, and Q the controllability matrix. Then Eq. (17) becomes

$$\bar{Y}^{(2)}H = \bar{Y}^{(2)}[PAQ] = \underline{Y} \tag{19}$$

It is known that the rank of a sufficiently large H is the order of the controllable and observable part of the system. From an experimental point of view, the identified state matrix A represents only the controllable and observable part of the system. The size of the matrix H is $qp \times Nm$, where N is an arbitrary integer. If we assume that $Nm > qp$, the maximum rank of H is thus qp. If p is chosen such that $qp > n$ (the order of the matrix A) and $\bar{Y}^{(2)}$ is obtained uniquely, then a realized state matrix A with order n should exist.

Therefore, we conclude that the number of observer Markov parameters computed p must be chosen such that $qp > n$, where q is the number of outputs and n the order of the system and obviously p can be smaller than the true order of the system for a multiple-output system. For a single-output system, the number p must be greater than or equal to the true order of the system. The number p determines the maximum number of independent system Markov parameters as seen from Eq. (15). Therefore, p represents the upper bound on the identified system model. When a Hankel matrix is formed for the purpose of system identification, there is no benefit to include additional system Markov parameters beyond the necessary number to create a full-rank Hankel matrix.

Equation (15) can be written in the following matrix form:

$$
\begin{bmatrix}
I & & & & \\
-\overline{Y}_0^{(2)} & I & & & \\
-\overline{Y}_1^{(2)} & -\overline{Y}_0^{(2)} & I & & \\
\vdots & \vdots & \vdots & \ddots & \\
-\overline{Y}_{k-1}^{(2)} & -\overline{Y}_{k-2}^{(2)} & -\overline{Y}_{k-3}^{(2)} & \cdots & I
\end{bmatrix}
\begin{bmatrix}
Y_0 \\ Y_1 \\ Y_2 \\ \vdots \\ Y_k
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
\overline{Y}_0^{(1)} + \overline{Y}_0^{(2)}D \\
\overline{Y}_1^{(1)} + \overline{Y}_1^{(2)}D \\
\overline{Y}_2^{(1)} + \overline{Y}_2^{(2)}D \\
\vdots \\
\overline{Y}_k^{(1)} + \overline{Y}_k^{(2)}D
\end{bmatrix}
\tag{20}
$$

Note that I and all $\overline{Y}_i^{(2)}$ ($i = 0, 1, \ldots, k$) are $q \times q$ square matrices. It is immediately seen that back substitution for $Y_0, Y_1, \ldots, Y_k$ from Eq. (20) yields Eq. (15). It is known that recursive back substitution without pivoting may result in a significant error accumulation in the solution. For numerical accuracy, it is better to use some type of pivoting procedure to minimize the error accumulation, unless the diagonal terms are dominant. However, the recursive back substitution is superior in computational efficiency relative to other methods, particularly when real-time computation is required.

To identify the observer gain M, first recover the sequence of parameters

$$
Y_k^o = CA^kM; \qquad k = 0, 1, 2, \ldots \tag{21}
$$

in terms of the observer Markov parameters. In fact, the first parameter in the sequence is simply

$$
Y_0^o = CM = -\overline{Y}_0^{(2)} \tag{22}
$$

The next parameter in the sequence is obtained by considering $\overline{Y}_1^{(2)}$

$$
\overline{Y}_1^{(2)} = -C\overline{A}M = -(CAM + CMCM) = -Y_1^o + \overline{Y}_0^{(2)}Y_0^o
$$

which yields

$$
Y_1^o = -\overline{Y}_1^{(2)} + \overline{Y}_0^{(2)}Y_0^o \tag{23}
$$

Similarly,

$$
\overline{Y}_2^{(2)} = -C\overline{A}^2M = -(CA^2M + CMCAM + C\overline{A}MCM)
$$

$$
= -Y_2^o + \overline{Y}_0^{(2)}Y_1^o + \overline{Y}_1^{(2)}Y_0^o
$$

which yields

$$
Y_2^o = -\overline{Y}_2^{(2)} + \overline{Y}_0^{(2)}Y_1^o + \overline{Y}_1^{(2)}Y_0^o \tag{24}
$$

By induction, the general relationship is

$$
Y_k^o = -\overline{Y}_k^{(2)} + \sum_{i=0}^{k-1} \overline{Y}_i^{(2)}Y_{k-i-1}^o \tag{25}
$$

Having obtained the sequence $Y_k^o = CA^kM$, $k = 0, 1, 2, \ldots$, where C and A can be realized by an identification method[11,12] from the Markov parameter sequence $Y_k = CA^kB$, $k = 0, 1, 2, \ldots$ obtained from Eq. (20), we can compute the observer gain M from

$$
M = (O^TO)^{-1}O^TY^o \tag{26}
$$

where

$$
O =
\begin{bmatrix}
C \\ CA \\ CA^2 \\ \vdots \\ CA^k
\end{bmatrix}, \qquad
Y^o =
\begin{bmatrix}
Y_0^o \\ Y_1^o \\ Y_2^o \\ \vdots \\ Y_k^o
\end{bmatrix}
=
\begin{bmatrix}
CM \\ CAM \\ CA^2M \\ \vdots \\ CA^kM
\end{bmatrix}
\tag{27}
$$

Equation (25) can be written in matrix form as

$$
\begin{bmatrix}
I & & & & \\
-\overline{Y}_0^{(2)} & I & & & \\
-\overline{Y}_1^{(2)} & -\overline{Y}_0^{(2)} & I & & \\
\vdots & \vdots & \vdots & \ddots & \\
-\overline{Y}_{k-1}^{(2)} & -\overline{Y}_{k-2}^{(2)} & -\overline{Y}_{k-3}^{(2)} & \cdots & I
\end{bmatrix}
\begin{bmatrix}
Y_0^o \\ Y_1^o \\ Y_2^o \\ \vdots \\ Y_k^o
\end{bmatrix}
=
\begin{bmatrix}
-\overline{Y}_0^{(2)} \\
-\overline{Y}_1^{(2)} \\
-\overline{Y}_2^{(2)} \\
\vdots \\
-\overline{Y}_k^{(2)}
\end{bmatrix}
\tag{28}
$$

The foregoing statement about Eq. (15) regarding the number of independent system Markov parameters also applies to the observer gain Markov parameters Y_k^o in Eq. (25) or (28). Note that I and all $\overline{Y}_i^{(2)}$ ($i = 0, 1, \ldots$) are $q \times q$ square matrices. Therefore, the leftmost matrix in Eq. (28) is square and full-rank and identical to that in Eq. (20). Hence, Y^o is determined uniquely from an identified set of observer Markov parameters. Equation (26) implies that the observer gain M computed from Eq. (26) is automatically the same coordinates as those for a set of A, C (and B) resulting from any realization. Recall that the set of system Markov parameters used for the system realization is also uniquely determined from the same identified set of observer Markov parameters, Eqs. (15) or (20). Computationally, Eqs. (15) and (25) or Eqs. (20) and (28) can be combined as a single matrix equation to solve for Y_k and Y_k^o simultaneously, i.e.,

$$
P_k = [Y_k \quad Y_k^o] = [CA^kB \quad CA^kM] = CA^k[B \quad M]
$$

$$
= \left[\overline{Y}_k^{(1)} + \overline{Y}_k^{(2)}D \quad -\overline{Y}_k^{(2)} \right] + \sum_{i=0}^{k-1} \overline{Y}_i^{(2)} \left[Y_{k-i-1} \quad Y_{k-i-1}^o \right]
\tag{29}
$$

Conventional system identification methods would use only the impulse response history Y_k to determine A, B, C, and D. Here, the combined system and observer gain Markov parameters P_k are used in a Hankel matrix to identify A, $[B \quad M]$, C, and D by some time domain method such as ERA[11] or ERA/DC.[12] There are several advantages to this approach. First, the observer gain M is obtained directly, which will be shown to be related to the Kalman filter gain in the next section. Second, the number of independent Markov parameters has been compressed by using the observer. This allows one to use a smaller Hankel matrix and thus reduce the computational effort in the identification algorithm. Third, one can identify the number of independent system Markov parameters from a single set of data for lightly damped systems with multiple inputs and outputs. This is a result of the increased stability produced by adding an observer gain that allows one to use a smaller p in Eq. (7) than in Eq. (3).

Relationship Between the Identified Observer and a Kalman Filter

Let Eq. (1) be extended to include process and measurement noise described as

$$
x(i+1) = Ax(i) + Bu(i) + w_1(i)
$$

$$
y(i) = Cx(i) + Du(i) + w_2(k) \tag{30}
$$

where $w_1(k)$ is the process noise assumed to be Gaussian, zero mean, and white with the covariance matrix Q, and $w_2(i)$ is the measurement noise with the same assumption as $w_1(i)$ but a different covariance matrix R. The sequence $w_1(i)$ and $w_2(i)$ are assumed to be statistically independent of each other.

A typical Kalman filter for Eq. (30) can then be written as

$$\hat{x}^+(i) = \hat{x}^-(i) + K\left[y(i) - \hat{y}(i)\right] \triangleq \hat{x}^-(i) + K\epsilon_r(i)$$

$$\hat{x}^-(i) = A\hat{x}^+(i-1) + Bu(i-1)$$

$$\hat{y}(i) = C\hat{x}^-(i) + Du(i) \tag{31}$$

where $\hat{x}^+(i)$ is the estimated state. The term $\epsilon_r(i)$ is called the residual and is defined as the difference between the real measurement $y(i)$ and predicted measurement $\hat{y}(i)$. Combination of the first two equations in Eq. (31) yields

$$\hat{x}^-(i+1) = A\,[I - KC]\hat{x}^-(i) + [B - AKD]u(i) + AKy(i)$$

or

$$\hat{x}^-(i+1) = \tilde{A}\hat{x}^-(i) + \tilde{B}v(i)$$

$$y(i) = C\hat{x}^-(i) + Du(i) + \epsilon_r(i) \tag{32}$$

where

$$\tilde{A} = A\,[I - KC]$$

$$\tilde{B} = [B - AKD, AK]$$

$$v(i) = \begin{bmatrix} u(i) \\ y(i) \end{bmatrix}$$

A comparison of Eqs. (4) and (32) reveals that they are identical if $M = -AK$ and $\epsilon_r(i) = 0$, and so are their Markov parameters. A question immediately arises as to whether $K = -A^{-1}M$ if M is computed using the computational procedure just developed. It is known that the Kalman filter gain K depends on the process covariance Q and measurement covariance R. Some condition must exist such that the equation $M = -AK$ is valid because the same equations are used to solve for the Kalman filter gain K and observer gain M. The key is the error term. The conditions will be derived in the following discussion.

Equation (32) can be written in the following matrix form:

$$\bar{y} = \tilde{Y}\bar{V} + \epsilon + C\tilde{A}^p\hat{x}^- \tag{33}$$

where $\bar{y}$ and $\bar{V}$ are defined in Eq. (8) and

$$\tilde{Y} = [D \quad C\tilde{B} \quad C\tilde{A}\tilde{B} \quad \cdots \quad C\tilde{A}^{p-1}\tilde{B}]$$

$$\hat{x}^- = \left[\hat{x}^-(0) \quad \hat{x}^-(1) \quad \hat{x}^-(2) \quad \cdots \quad \hat{x}^-(\ell-p-1)\right]$$

$$\epsilon = \left[\epsilon_r(p) \quad \epsilon_r(p+1) \quad \epsilon_r(p+2) \quad \cdots \quad \epsilon_r(\ell-1)\right]$$

and ϵ is the residual error as defined in Eq. (31) and ℓ the data length. Equation (33) applies to any equation with the same observer structure as Eq. (31). If the observer happens to be a Kalman filter, then the residual is white, zero-mean, and Gaussian.

Postmultiplying Eq. (33) by $\bar{V}^T$ yields

$$\bar{y}\bar{V}^T = \tilde{Y}\bar{V}\bar{V}^T + \epsilon\bar{V}^T + C\tilde{A}^p\hat{x}^-\bar{V}^T \tag{34}$$

Let $\bar{V}$ be partitioned in rows as

$$\bar{V}^T = \begin{bmatrix} v_p^T & v_{p-1}^T & \cdots & v_0^T \end{bmatrix}$$

Equation (34) can then be rewritten as

$$\begin{bmatrix} \bar{y}v_p^T & \bar{y}v_{p-1}^T & \cdots & \bar{y}v_0^T \end{bmatrix}$$

$$-\tilde{Y}\begin{bmatrix} v_p v_p^T & v_p v_{p-1}^T & \cdots & v_p v_0^T \\ v_{p-1}v_p^T & v_{p-1}v_{p-1}^T & \cdots & v_{p-1}v_0^T \\ \vdots & \vdots & \ddots & \vdots \\ v_0 v_p^T & v_0 v_{p-1}^T & \cdots & v_0 v_0^T \end{bmatrix}$$

$$= \begin{bmatrix} \epsilon v_p^T & \epsilon v_{p-1}^T & \cdots & \epsilon v_0^T \end{bmatrix}$$

$$+ C\tilde{A}^p\begin{bmatrix} \hat{x}^- v_p^T & \hat{x}^- v_{p-1}^T & \cdots & \hat{x}^- v_0^T \end{bmatrix}$$

Let us examine a term from $\epsilon\bar{V}^T$

$$\epsilon v_i^T = \sum_{j=0}^{\ell-p-1} \epsilon_r(p+j)v^T(i+j); \qquad i = 0,\ldots,p \tag{35}$$

If this term is divided by $\ell-p$, it represents the time average of the product $\epsilon(k)v^T(k-i)$ from $k=p$ to $\ell-1$. By the ergodic property,[13] if the product is a sample function of a stationary random process, it can be replaced by its ensemble average, provided that ℓ goes to infinity, $\ell\to\infty$,

$$E\left[\epsilon_r(k)v^T(k-i)\right] = \lim_{\ell\to\infty}\frac{1}{\ell-p}\sum_{j=p}^{\ell-1}\epsilon_r(j)v^T(j-i); \quad k > p \tag{36}$$

Physically, ergodicity implies that a sufficiently long record of a stationary random process contains all the statistical information about the random phenomenon. In practical applications, the ergodic property makes it possible to obtain the noise-related moment functions of a stationary random process from a single long record. The conversion from a time average to an expected value is performed similarly for the other terms, including $\bar{y}v_j^T$, $\hat{x}^- v_j^T$, $v_i v_j^T$ $(i,j = 0,\ldots,p)$.

The concept of stationarity in a random process is analogous to the steady-state behavior of a deterministic process. In practice, no random process can be truly stationary. However, a long segment of a random process exhibiting uniform characteristics can be treated as stationary. One must allow sufficient time for the system transients to decay before the data sequence starts. In addition, the choice of p in Eq. (33) has to be sufficiently large that the transients of the Kalman filter are negligible.

Equation (34) can now be written as

$$\lim_{\ell\to\infty}\frac{1}{\ell-p}\left[\bar{y}\bar{V}^T - \tilde{Y}\bar{V}\bar{V}^T\right] = E\left[\epsilon_r(k)v^T(k) \quad \epsilon_r(k)v^T(k-1) \quad \cdots \quad \epsilon_r(k)v^T(k-p)\right]$$

$$+ C\tilde{A}^p E\left[\hat{x}^-(k)v^T(k+p) \quad \hat{x}^-(k)v^T(k+p-1) \quad \cdots \quad \hat{x}^-(k)v^T(k)\right] \tag{37}$$

for all $k > p$. If we choose the observer such that

$$\tilde{Y} = \bar{y}\bar{V}^T\left[\bar{V}\bar{V}^T\right]^{-1} \tag{38}$$

in the limit $\ell\to\infty$, then

$$E\left[\epsilon_r(k)v^T(k) \quad \epsilon_r(k)v^T(k-1) \quad \cdots \quad \epsilon_r(k)v^T(k-p)\right] = -C\tilde{A}^p E\left[\hat{x}^-(k)v^T(k+p) \quad \hat{x}^-(k)v^T(k+p-1) \quad \cdots \quad \hat{x}^-(k)v^T(k)\right] \tag{39}$$

Because $\bar{A}$ for an observer is asymptotically stable, let p be chosen sufficiently large that the right-hand side of the aforementioned equation is negligible, i.e.,

$$E\left[\epsilon_r(k)v^T(k-i)\right] = 0 \tag{40}$$

for $i = 0,1,\ldots,p$ and $k > p$. Substitution of the definition for $v(k-i)$ from Eq. (32) into Eq. (40) yields

$$E\left[\epsilon_r(k)u^T(k-i)\right] = 0; \qquad i = 0,\ldots,p$$

$$E\left[\epsilon_r(k)y^T(k-j)\right] = 0; \qquad j = 1,\ldots,p \tag{41}$$

for $k > p$, which implies that the residual error $\epsilon_r(k)$ at any time k is orthogonal to the input function $u(k-i)$ with the time delay i from 1 up to p, and the output function $y(k-j)$ with the time delay j from 0 up to p. In other words, if we choose the observer with the observer Markov parameters that satisfy the least-squares equation (38), the residual describing the difference between the estimated output measurement and real measurement is orthogonal to the given input and the measured output with time delay. This has application to model reduction based on the orthogonality of the output measurements and the residuals, representing the output errors between the full and the reduced model.[14]

Now given a set of data from a finite-dimensional system of Eq. (30), a Kalman filter exists with the property that the residual is white, zero-mean, and Gaussian, i.e.,

$$E\left[\epsilon_r(k)\right] = 0, \qquad E\left[\epsilon_r(j)\epsilon_r^T(k)\right] = 0; \qquad j \neq k \tag{42}$$

and satisfies the principle of orthogonality

$$E\left[\epsilon_r(k)y^T(k-i)\right] = 0; \qquad i = 1,\ldots,k \tag{43}$$

If the experimental process is stationary and random, the Kalman filter gain is a constant that produces the Kalman filter Markov parameters in the limit $\ell \to \infty$ satisfying the least-squares equation (36), provided that the inverse $[\bar{V}\bar{V}^T]^{-1}$ exists. For a sufficiently rich input, the inverse always exists.

We conclude that any observer satisfying Eq. (10), or its equivalent, Eq. (38), produces the same input-output map as a Kalman filter if the data length is sufficiently long and the order of the observer sufficiently large so that the truncation error is negligible. Therefore, when reduced to the system order, the identified observer has to be a Kalman filter and thus the M computed from the combined Markov parameters of Eq. (29) gives the steady-state Kalman filter gain
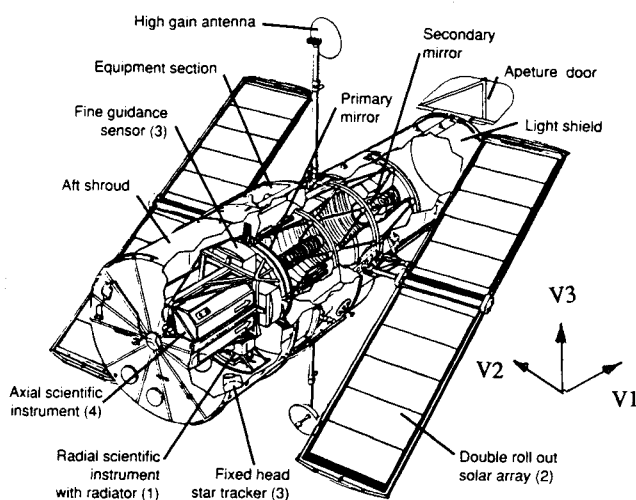
$$K = -A^{-1}M \tag{44}$$

Fig. 1 Hubble space telescope.

Table 1 Comparison of identified modal parameters

Case no.	Mode 1		Mode 2		Mode 3	
	Freq., Hz	Damp., %	Freq., Hz	Damp., %	Freq., Hz	Damp., %
0	0.261	0.63	0.712	1.01	0.972	1.30
1	0.261	0.55	0.712	0.96	0.970	1.65
2	0.261	0.56	0.712	0.95	0.970	1.67
3	0.261	0.59	0.712	0.99	0.971	1.52
4	0.261	0.59	0.712	1.00	0.971	1.51
5	0.261	0.65	0.712	0.99	0.971	1.52
6	0.261	0.65	0.712	0.98	0.971	1.53

Case 0: true values; case 1: 1000 data points, $p = 40$; case 2: 1000 data points, $p = 50$; case 3: 2000 data points, $p = 40$; case 4: 2000 data points, $p = 50$; case 5: 4000 data points, $p = 40$; and case 6: 4000 data points, $p = 50$.

Computational Algorithm

Given a set of experimental input and output data, the identification algorithm proceeds as follows.

Step 1: Choose a value of p [see Eq. (7)] that determines the number of observer Markov parameters to be identified from the given set of input and output data. In general, p is required to be sufficiently larger (at least four or five times) than the effective order of the system for identification of the Kalman filter gain with accuracy.

Step 2: Form the two data matrices y and V as shown in Eq. (7) for zero initial conditions, or $\bar{y}$ and $\bar{V}$ as shown in Eq. (9) for nonzero initial conditions, and compute the least-squares solution of the observer Markov parameter matrix $\bar{Y}$.

Step 3: Recover the combined system and Kalman filter Markov parameters P_k from the identified observer Markov parameters using Eq. (29). To solve for more Markov parameters than the number of identified observer Markov parameters, simply set the extra observer Markov parameters to zero.

Step 4: Realize a state-space model of the system and the corresponding Kalman filter gain from the recovered sequence P_k using ERA or ERA/DC.

Simulation Results

As an example, a spring/mass three-degree-of-freedom system is used to simulate data with known noise properties. The simulated system, used in Ref. 9, has one input and two outputs. The continuous system is discretized at a sampling frequency of 10 Hz. The discrete-time model for this system with process and measurement noise covariances used in this example is given by Eqs. (A1) and (A2) in the Appendix.

These covariances were chosen by the following procedure. First, a simulation was performed using random $u(k)$ with a standard deviation of 20 to determine the noise-free sequences $Bu(k)$ and $y(k)$. The standard deviation of the process noise was computed to be 5% that of the sequence $Bu(k)$. Similarly, the standard deviation of the measurement noise was chosen as 5% that of the sequence $y(k)$. To examine the stochastic properties of the system, one must assume that the sample histories are infinitely long, but in practice they are not. Therefore, the effect of short time records must be examined. Also in the theoretical development, the observer order p is specified a priori. In the simulations, two different values for the observer order parameter p are used and the results compared.

The computational algorithm is applied to identify the system and corresponding Kalman filter gain in the presence of the prescribed noise levels. Examination of Table 1 shows that the frequencies are accurately identified in all cases to within 0.2%. Damping estimates, however, vary up to 28%, with improved results obtained when the number of data samples is increased. Computation of the frequencies and damping values is based on a realization of the system matrix A from the Markov parameter sequence. The realization algorithm presented in Ref. 12 is used, and a minimum-order realization is obtained from the Hankel correlation matrix HH^T, where H

Table 2 Comparison of Kalman filter gains

Case no.	Kalman filter gain matrix
0	$K = \begin{bmatrix} 0.0293 & -0.0012 & 0.0025 & 0.0000 & 0.0241 & 0.0047 \\ 0.0295 & 0.0012 & 0.0000 & 0.0005 & -0.0251 & 0.0042 \end{bmatrix}^T$
2	$K = \begin{bmatrix} 0.0261 & -0.0166 & 0.0198 & 0.0046 & 0.0236 & 0.0086 \\ 0.0253 & -0.0020 & -0.0010 & 0.0116 & -0.0328 & 0.0241 \end{bmatrix}^T$
6	$K = \begin{bmatrix} 0.0255 & -0.0066 & 0.0196 & -0.0047 & 0.0246 & 0.0071 \\ 0.0265 & -0.0052 & 0.0008 & 0.0037 & -0.0298 & 0.0062 \end{bmatrix}^T$

Table 3 Identified modal parameters for the Hubble space telescope

Mode no.	Frequency, Hz	Damping, %	Mode SV
1	0.147	55.6	0.76
2	0.155	58.4	0.98
3	0.169	67.4	1.00
4	0.633	5.73	0.68
5	1.273	4.06	0.37
6	2.433	5.23	0.02
7	2.822	6.33	0.01
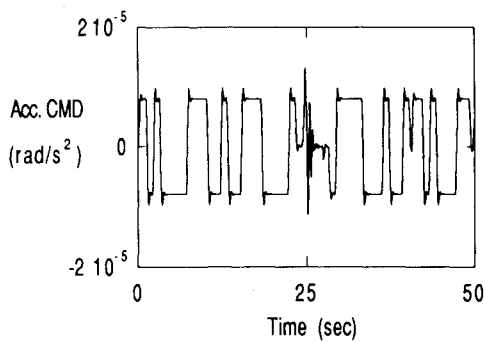
Fig. 2a Excitation input signal.
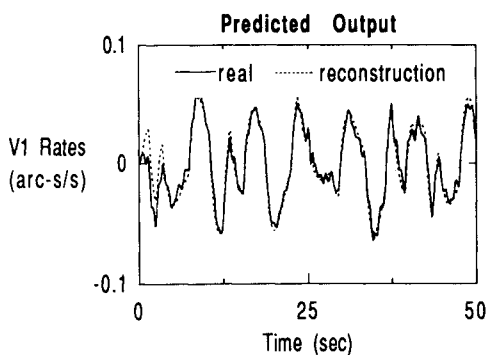
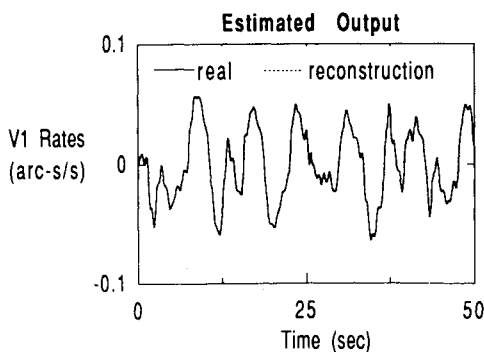Fig. 2b Comparison of test data with predicted output.

Fig. 2c Comparison of test data with estimated output.

is as in Eq. (16). For deterministic systems, the rank of the correlation matrix is equal to the system order. For stochastic systems, the problem of rank determination is not as clear and the method of singular value decomposition is used to determine the system order. Retaining only those singular values with a significant contribution to the correlation matrix renders a model of the same order as the number of retained singular values. The value of p is chosen to be either 40 or 50 and only the first six singular values are retained.

Kalman filter gains are shown in Table 2. Although the numerical comparison in terms of frequencies and damping values is good, the estimated Kalman filter gains for the different cases could be quite different from the true value because of the finiteness of the data lengths. Table 2 shows that as the number of data points used in the identification is increased, the identified Kalman filter gains approach the true value.

Comparing the A, B, and C matrices in Eq. (A3) in the Appendix with the true ones in Eq. (A1) shows excellent agreement, but a nonzero direct transmission term is picked up by the identification. The reconstruction (not shown) of the system response using the identified observer parameters in Eq. (A3) and the identified Kalman filter gain in Table 2, when compared to the actual response, shows excellent agreement.

Experimental Results

To demonstrate the identification procedure using real experimental data, the Hubble space telescope shown in Fig. 1 is employed. There are six gyros located on the optical telescope assembly (OTA) and four torque wheels on the spacecraft subsystem module (SSM). The OTA is fixed inside the SSM. The gyros are used mainly to measure the motion of the primary mirror. Data from four out of the six gyros are recorded at a time. The measurement resolution is 0.005 arcsec/s, which implies that the gyro data are not adequate because the requirement is 0.007 arcsec pointing. The angular rates, which are measured along the four gyro directions, are combined and transformed using least-squares to recover the three rates in vehicle coordinates. Least-squares is used to smooth the poor resolution of the data. The input commands are given in terms of angular acceleration in the three rotational vehicle coordinates and then projected on the four torque wheel axes to excite the telescope mirror and spacecraft. The data were sampled at 40 Hz. Pulses combined with sine-sweeping in the middle of an excitation period (50.975 s) were used as input commands to the torque wheels. The excitation period was repeated six times for a total of approximately 12,000 samples taken for each experiment. The experiment was repeated three times for the other two vehicle coordinates. As a result, there were three inputs and four outputs for a total of 3 sets of 12,000 input samples and 12 sets of 12,000 output samples to be used for the identification of vibration parameters.

A linear model and observer were identified for the Hubble space telescope. The system order was chosen to be 30 for realization of the system matrices. Seven dominant modes were identified as shown in Table 3. The mode SV in the table describes the singular value contribution of each individual

mode to the pulse responses.[15] It has been normalized relative to the maximum singular value. The first three modes are attitude rigid modes. The 0.65-Hz mode is believed to be an in-plane bending mode of the solar array, the 1.29-Hz mode is a coupled solar and membrane mode, and the 2.45-Hz mode is the first mode of the primary deployment mechanism with the solar array housing attached. The identified dampings are higher than expected because there is an attitude control for maneuvering during testing and mechanical friction of solar array mechanism.

Figure 2a illustrates the excitation input signal, including pulse combined with a sine-sweeping signal in the middle of an excitation period for the first vehicle axis. Figures 2b and 2c show overlapping 50 s of the reconstruction from the identified system models and the test data for the first vehicle axis. Figure 2b depicts the predicted output in comparison with the real output data and Fig. 2c the estimated output in comparison with the real output data. The predicted output is the output reconstructed from the identified model only, whereas the estimated output is the output reconstructed from the identified observer. There are visible differences in the predicted and estimated outputs. Comparison of the observer output with the measured response shows extremely good agreement, indicating that the observer is correcting for the system uncertainties including nonlinearities. The covariance of the estimated output residuals is about three orders in magnitude less than the predicted output residuals. Similar results of the pre-

dicted and estimated outputs were obtained for the second and third vehicle axes and thus are not shown here.

Concluding Remarks

An algorithm for the direct computation of observer/Kalman filter Markov parameters, and from them the observer/Kalman filter matrices, has been presented. The matrix formulation developed here allows one to establish the uniqueness and invertibility of the transformation from observer/Kalman filter Markov parameters to the system Markov parameters. The matrix formulation also establishes bounds on the choice of the observer/Kalman filter order for the data. The algorithm is a nonrecursive matrix version of two previous algorithms: One is a recursive algorithm for Kalman filter identification, and the other is an algorithm for direct identification of observers with chosen pole locations, specialized to have all poles at the origin (the deadbeat observer). The nonrecursive form of the least-squares solution used here results in a substantial improvement in the convergence rate to the true Kalman gain that was reported before. The relationship between the deadbeat observer and the Kalman filter Markov parameter identification problems is established here. It is shown that using the equations for deterministic deadbeat observer parameters on noisy data results in obtaining the Kalman filter parameters, in the limit as the amount of data used tends to infinity. When a finite set of data is used, the resulting filter satisfies an optimality condition, indicating it is the best filter that can be obtained with the data length available.

Appendix: Simulation Results

$$A = \text{diag}\left\{ \begin{bmatrix} 0.9856 & 0.1628 \\ -0.1628 & 0.9856 \end{bmatrix}, \begin{bmatrix} 0.8976 & 0.4305 \\ -0.4305 & 0.8976 \end{bmatrix}, \begin{bmatrix} 0.8127 & 0.5690 \\ -0.5690 & 0.8127 \end{bmatrix} \right\}$$

$$B = \begin{bmatrix} 0.0011 & 0.0134 & -0.0016 & -0.0072 & 0.0011 & 0.0034 \end{bmatrix}^T$$

$$C = \begin{bmatrix} 1.5119 & 0.0000 & 2.0000 & 0.0000 & 1.5119 & 0.0000 \\ 1.3093 & 0.0000 & 0.0000 & 0.0000 & -1.3093 & 0.0000 \end{bmatrix}$$

$$D = \begin{bmatrix} 0.0000 & 0.0000 \end{bmatrix}^T \tag{A1}$$

The matrix A is given in block diagonal form for later comparison with the identification results. The process noise and measurement noise covariances are specified, respectively, to be

$$Q = \text{diag}\begin{bmatrix} 0.0242 & 3.5920 & 0.0534 & 1.034 & 0.0226 & 0.2279 \end{bmatrix} \times 10^{-4}$$

$$R = \text{diag}\begin{bmatrix} 2.785 & 2.785 \end{bmatrix} \times 10^{-2} \tag{A2}$$

The corresponding realized system matrices for case no. 6 are

$$A = \text{diag}\left\{ \begin{bmatrix} 0.9856 & 0.1629 \\ -0.1629 & 0.9856 \end{bmatrix}, \begin{bmatrix} 0.8977 & 0.4306 \\ -0.4306 & 0.8977 \end{bmatrix}, \begin{bmatrix} 0.8120 & 0.5676 \\ -0.5676 & 0.8120 \end{bmatrix} \right\}$$

$$B = \begin{bmatrix} 0.0011 & 0.0132 & -0.0016 & -0.0068 & 0.0011 & 0.0034 \end{bmatrix}^T$$

$$C = \begin{bmatrix} 1.5107 & 0.0000 & 2.0000 & 0.0000 & 1.5133 & 0.0000 \\ 1.3107 & 0.0011 & -0.0087 & -0.0104 & -1.3073 & 0.0288 \end{bmatrix}$$

$$D = \begin{bmatrix} 0.0007 & 0.0012 \end{bmatrix}^T \tag{A3}$$

Acknowledgment

References

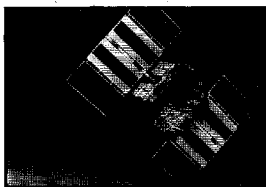[1] Juang, J.-N., and Phan, M., "Robust Controller Designs for Second-Order Dynamic Systems: A Virtual Passive Approach," NASA TM-102666, May 1990.

[2] Juang, J.-N., and Lew, J.-S., "Integration of System Identification and Robust Controller Designs for Flexible Structures in Space," *Proceedings of the AIAA Guidance, Navigation, and Control Conference* (Portland, OR), AIAA, Washington, DC, 1990, pp. 1361-1375 (AIAA Paper 90-3467).

[3] Juang, J.-N., and Pappa, R. S., "A Comparative Overview of Modal Testing and System Identification for Control of Structures," *Shock and Vibration Digest,* Vol. 20, No. 5, 1988, pp. 4-15.

[4] Pappa, R. S., and Juang, J.-N., "Some Experiences with the Eigensystem Realization Algorithm," *Sound and Vibration,* Vol. 22, No. 1, 1988, pp. 30-35.

[5] Juang, J.-N., Horta, L. G., and Longman, R. W., "Input/Output System Identification: Learning from Repeated Experiments," *Mechanics and Control of Large Space Structures,* edited by John L. Junkins, Vol. 129, Progress in Astronautics and Aeronautics, AIAA, Washington, DC, 1990, pp. 87-99.

[6] Phan, M., Juang, J.-N., and Longman, R. W., "Identification of Linear Systems by Identification of Observers with Assigned Real Eigenvalues," *Journal of the Astronautical Sciences,* Vol. 40, No. 2, April-June 1992, pp. 261-279.

[7] Phan, M., Horta, L. G., Juang, J.-N., and Longman, R. W., "Linear System Identification via an Asymptotically Stable Observer," *Proceedings of the AIAA Guidance, Navigation and Control Conference* (New Orleans, LA), AIAA, Washington, DC, 1991, pp. 1180-1194.

[8] Phan, M., Horta, L. G., Juang, J.-N., and Longman, R. W., "Identification of Linear Systems via an Asymptotically Stable Observer," NASA Tech. Paper 3164, June 1992.

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

[10] Huang, J.-K., Juang, J.-N., and Chen, C.-W., "Single-Mode Projection Filters for Modal Parameter Identification for Flexible Structures," *Journal of Guidance, Control, and Dynamics,* Vol. 12, No. 4, 1989, pp. 568-576.

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

[12] Juang, J.-N., Cooper, J. E., and Wright, J. R., "An Eigensystem Realization Algorithm Using Data Correlations (ERA/DC) for Modal Parameter Identification," *Control Theory and Advanced Technology,* Vol. 4, No. 1, 1988, pp. 5-14.

[13] Nigam, N. C., *Introduction to Random Vibration,* MIT Press, Cambridge, MA, 1983.

[14] Gawronski, W., and Juang, J.-N., "Model Reduction for Flexible Structures," *Advances in Large Scale Systems Dynamics,* edited by C. T. Leondes, Academic, New York, 1990, pp. 143-222.

[15] Juang, J.-N., Horta, L. G., and Phan, M., "System/Observer/Controller Identification Toolbox," NASA TM 107566, Feb. 1992.