JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 19, No. 6, November–December 1996

Removing Bias from Discrete Time Models Based on Finite Interval Response Data

Myles L. Baker* and D. Lewis Mingori†
University of California, Los Angeles, Los Angeles, California 90095

Modern off-line system identification methods necessarily form a system realization using a finite amount of data. The models constructed are usually of an order smaller than the actual order of the physical system because it is not practical to model the relatively unimportant states. A method is presented to improve the selection of the important states and the model order in system identification methods based on generalized Hankel matrices. When forming a reduced-order model, it is critical to be able to accurately identify the important states and to select the appropriate model order. Some currently popular methods implicitly contain a frequency weighting that comes from the use of a finite time interval of data in the identification. A method is presented to remove this implicit bias, and excellent results are obtained, even when the response data used span only a fraction of a cycle of the dominant frequency.

Introduction

THIS paper is concerned with the problem of constructing a discrete time state-space model of a system of unknown order based on knowledge of the unit pulse response (i.e., Markov parameters) over a finite time interval. The focus is on reducing the errors introduced by using a finite time interval of pulse response data to construct a state-space model. The results are also of interest whenever a low-order (reduced) model is constructed using a small amount of data. A procedure is proposed that has much in common with the Eigensystem Realization Algorithm (ERA)[1,2] but differs in the steps in which the model order is determined and the state-space matrices are constructed. The new procedure produces a model that is internally balanced on the semi-infinite interval $0 \leq t \leq \infty$. It has been used very successfully even when the available response data span only a fraction of an oscillation of the dominant frequency.

Consider a discrete time dynamic system described by the state-space equations

$$\boldsymbol{x}(k+1) = \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{B}\boldsymbol{u}(k) \tag{1}$$

$$\boldsymbol{y}(k) = \boldsymbol{C}\boldsymbol{x}(k) \tag{2}$$

where $\boldsymbol{x}$ is a state vector of length n, $\boldsymbol{u}$ is an input vector of length m, and $\boldsymbol{y}$ is an output vector of length p. For zero initial conditions, the response of the system can be expressed as

$$\boldsymbol{y}(i) = \sum_{k=1}^{i} \boldsymbol{Y}(k)\boldsymbol{u}(i-k) \tag{3}$$

$$\boldsymbol{Y}(k) = \boldsymbol{C}\boldsymbol{A}^{k-1}\boldsymbol{B} \tag{4}$$

The quantities $\boldsymbol{Y}(k)$ are known as the Markov parameters, and they describe the unit pulse response of the system. [Consider Eq. (3) and let $u(k) = 1$ for $k = 0$ and 0 for $k > 0$.] The Markov parameters could in principle be determined by conducting an experiment in which the actual input to the system was a unit pulse. In practice, however, it is typical to use any of several methods by which Markov parameters can be estimated from experimental data even if the input is not a unit pulse. It is not our objective here to address the question of how the unit pulse response data are obtained. This is considered, for example, in Refs. 3–5. Instead, we consider what happens if accurate (noise-free) values for the $\boldsymbol{Y}(k)$ are known for some finite number of k, and we wish to use these data to generate a state-space model of the system.

In Refs. 1 and 2, the authors describe a procedure called the ERA for constructing a state-space model from known Markov parameters. To illustrate the effects of using a finite amount of data, first consider the implementation of the ERA. The first step in this procedure is to introduce a generalized Hankel matrix defined as

$$\boldsymbol{H}_{rs}(k) = \begin{bmatrix} \boldsymbol{C}\boldsymbol{A}^{k+j_0+t_0}\boldsymbol{B} & \boldsymbol{C}\boldsymbol{A}^{k+j_0+t_1}\boldsymbol{B} & \cdots & \boldsymbol{C}\boldsymbol{A}^{k+j_0+t_{s-1}}\boldsymbol{B} \\ \boldsymbol{C}\boldsymbol{A}^{k+j_1+t_0}\boldsymbol{B} & \boldsymbol{C}\boldsymbol{A}^{k+j_1+t_1}\boldsymbol{B} & \cdots & \boldsymbol{C}\boldsymbol{A}^{k+j_1+t_{s-1}}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{C}\boldsymbol{A}^{k+j_{r-1}+t_0}\boldsymbol{B} & \boldsymbol{C}\boldsymbol{A}^{k+j_{r-1}+t_1}\boldsymbol{B} & \cdots & \boldsymbol{C}\boldsymbol{A}^{k+j_{r-1}+t_{s-1}}\boldsymbol{B} \end{bmatrix} \tag{5}$$

where r, s, t_i $(i = 0 \cdots s-1)$, and j_i $(i = 0 \cdots r-1)$ are arbitrary integers. $\boldsymbol{H}_{rs}(k)$ can be expressed in the form

$$\boldsymbol{H}_{rs}(k) = \boldsymbol{V}_r \boldsymbol{A}^k \boldsymbol{W}_s \tag{6}$$

where

$$\boldsymbol{V}_r = \begin{bmatrix} \boldsymbol{C}\boldsymbol{A}^{j_0} \\ \boldsymbol{C}\boldsymbol{A}^{j_1} \\ \vdots \\ \boldsymbol{C}\boldsymbol{A}^{j_{r-1}} \end{bmatrix} \qquad \boldsymbol{W}_s = [\boldsymbol{A}^{t_0}\boldsymbol{B} \quad \boldsymbol{A}^{t_1}\boldsymbol{B} \quad \cdots \quad \boldsymbol{A}^{t_{s-1}}\boldsymbol{B}] \tag{7}$$

The Hankel matrix $\boldsymbol{H}_{rs}(0)$ may be constructed using the Markov parameters that are presumed known from the pulse response. This step involves choosing values for $(j_0, \ldots, j_{r-1}, t_0, \ldots, t_{s-1})$ and making a decision about the size of $\boldsymbol{H}_{rs}(0)$. $\boldsymbol{H}_{rs}(0)$ must be small enough so that it can be constructed using only the known Markov parameters and also so that the computational cost associated with the subsequent singular-value decomposition and matrix operations is not prohibitive. Once $\boldsymbol{H}_{rs}(0)$ is constructed, a singular-value decomposition[6,7] leads to the representation of $\boldsymbol{H}_{rs}(0)$ as

$$\boldsymbol{H}_{rs}(0) = \boldsymbol{P}\boldsymbol{D}\boldsymbol{Q}^T \tag{8}$$

where $\boldsymbol{P}$ is $rp \times n$ and unitary, $\boldsymbol{Q}$ is $ms \times n$ and unitary, $\boldsymbol{D}$ is $n \times n$ and diagonal, and $n = \text{rank}[\boldsymbol{H}_{rs}(0)]$. In this context, the integer n

Presented as Paper 94-3640 at the AIAA Guidance, Navigation, and Control Conference, Scottsdale, AZ, Aug. 1–3, 1994; received Aug. 11, 1994; revision received Jan. 17, 1996; accepted for publication March 18, 1996.

*Graduate Student, Mechanical, Aerospace, and Nuclear Engineering Department. Student Member AIAA.

†Professor, Mechanical, Aerospace, and Nuclear Engineering Department. Associate Fellow AIAA.

represents the maximum possible dimension of the state-space system identified rather than the true order of the system. The singular values of $\boldsymbol{H}_{rs}(0)$ are arranged in descending order on the diagonal of $\boldsymbol{D}$. The matrices $\boldsymbol{P}$, $\boldsymbol{D}$, and $\boldsymbol{Q}$ may be partitioned such that

$$\begin{aligned}\boldsymbol{H}_{rs}(0) &= [\boldsymbol{P}_\ell \quad \boldsymbol{P}_{n-\ell}]\begin{bmatrix}\boldsymbol{D}_\ell & 0\\ 0 & \boldsymbol{D}_{n-\ell}\end{bmatrix}\begin{bmatrix}\boldsymbol{Q}_\ell^T\\ \boldsymbol{Q}_{n-\ell}^T\end{bmatrix}\\ &= \boldsymbol{P}_\ell\boldsymbol{D}_\ell\boldsymbol{Q}_\ell^T + \boldsymbol{P}_{n-\ell}\boldsymbol{D}_{n-\ell}\boldsymbol{Q}_{n-\ell}^T\end{aligned} \tag{9}$$

This partitioning is guided by the sizes of the singular values of $\boldsymbol{D}_{n-\ell}$ relative to those of $\boldsymbol{D}_\ell$. If the diagonal entries in $\boldsymbol{D}_{n-\ell}$ are sufficiently small compared with those of $\boldsymbol{D}_\ell$, it can be argued that the second term in (9) may be neglected, leading to

$$\boldsymbol{H}_{rs}(0) \cong \boldsymbol{P}_\ell\boldsymbol{D}_\ell\boldsymbol{Q}_\ell^T \tag{10}$$

An ℓth-order system representation $(\boldsymbol{A}_\ell, \boldsymbol{B}_\ell, \boldsymbol{C}_\ell)$ that satisfies Eq. (5) within the truncation approximation can be constructed using

$$\begin{aligned}\boldsymbol{A}_\ell &= \boldsymbol{D}_\ell^{-\frac{1}{2}}\boldsymbol{P}_\ell^T\boldsymbol{H}_{rs}(1)\boldsymbol{Q}_\ell\boldsymbol{D}_\ell^{-\frac{1}{2}}\\ \boldsymbol{B}_\ell &= \boldsymbol{D}_\ell^{\frac{1}{2}}\boldsymbol{Q}_\ell^T\boldsymbol{E}_m \qquad \boldsymbol{C}_\ell = \boldsymbol{E}_p^T\boldsymbol{P}_\ell\boldsymbol{D}_\ell^{\frac{1}{2}}\end{aligned} \tag{11}$$

where $\boldsymbol{E}_m$ consists of the first m rows of an ms-order identity matrix and $\boldsymbol{E}_p$ consists of the first p rows of an rp-order identity matrix. This is equivalent to forming an nth-order model $(\boldsymbol{A}_n, \boldsymbol{B}_n, \boldsymbol{C}_n)$, partitioning the matrices such that

$$\boldsymbol{A}_n = \begin{bmatrix}\boldsymbol{A}_\ell & \boldsymbol{A}_{\ell,n-\ell}\\ \boldsymbol{A}_{n-\ell,\ell} & \boldsymbol{A}_{n-\ell}\end{bmatrix} \qquad \boldsymbol{B}_n = \begin{bmatrix}\boldsymbol{B}_\ell\\ \boldsymbol{B}_{n-\ell}\end{bmatrix}$$

$$\boldsymbol{C}_n = [\boldsymbol{C}_\ell \quad \boldsymbol{C}_{n-\ell}] \tag{12}$$

and using $(\boldsymbol{A}_\ell, \boldsymbol{B}_\ell, \boldsymbol{C}_\ell)$ as a truncated model. Note that the dimension of $\boldsymbol{D}_\ell$ directly determines the dimension of $\boldsymbol{A}_\ell$, and hence, the order of the state-space model [see Eqs. (11)].

When a continuous (infinite-order) system such as a physical structure is being identified, it is necessary to make an engineering decision on which (and how many) states to retain. Typically, many states are weakly represented in the data and, if chosen carefully, can be truncated without having a large effect on the identified model's response. If the truncated states are chosen inappropriately, however, the performance of the reduced model can be degraded significantly. The method of model truncation in the ERA is to truncate states with small singular values $(\boldsymbol{D}_{n-\ell})$, as described above. In the examples that follow, a state will be considered negligible if its singular value is less than 1/100 the size of the largest singular value. It will be shown that if the dimensions of $\boldsymbol{H}_{rs}(0)$ are small, a better truncation criterion can sometimes be found.

The ERA has been used with both simulated and actual experimental data, and it has been shown to be an effective approach.[1,8,9] Three important factors affect the accuracy of the state-space model produced by the ERA: 1) noise in the measurement data that leads to errors in the Markov parameters; 2) the size and content of $\boldsymbol{H}_{rs}(0)$, i.e., the choice of r, s, t_i, and j_i; and 3) the dimension of $\boldsymbol{D}_\ell$ and $\boldsymbol{D}_{n-\ell}$, the retained and truncated portions of $\boldsymbol{D}$. In this paper it is assumed that the Markov parameters are accurately known so that no significant errors are introduced by item 1 in this list. It is then shown how the accuracy of the truncated model can be improved significantly without increasing the size of $\boldsymbol{H}_{rs}(0)$ or $\boldsymbol{D}_\ell$.

Example of the Standard ERA

Before presenting the modified procedure that is the subject of this paper, we consider a simple example that illustrates how the choice of r, s, and the size of $\boldsymbol{D}_\ell$ influences the accuracy of the resulting state-space model. For illustration, the simple discretized model of a bar in axial motion shown in Fig. 1 is used. In general, the time domain motion of a physical structure can be described by the second-order differential equation

$$\boldsymbol{M}\ddot{\boldsymbol{x}} + \boldsymbol{F}\dot{\boldsymbol{x}} + \boldsymbol{K}\boldsymbol{x} = \boldsymbol{f} \tag{13}$$

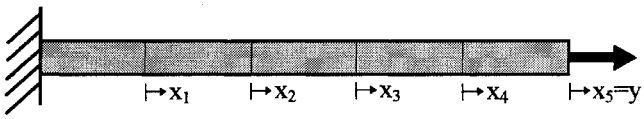

Fig. 1 Discretized bar model.

where $\boldsymbol{x}$ is the deformation vector, $\boldsymbol{M}$ is the mass matrix, $\boldsymbol{F}$ is a viscous damping matrix, $\boldsymbol{K}$ is the stiffness matrix, and $\boldsymbol{f}$ is the applied force vector. Equivalently, the dynamics can be represented by the first-order equation

$$\begin{Bmatrix}\dot{\boldsymbol{x}}\\ \ddot{\boldsymbol{x}}\end{Bmatrix} = \begin{bmatrix}0 & \boldsymbol{I}\\ -\boldsymbol{M}^{-1}\boldsymbol{K} & -\boldsymbol{M}^{-1}\boldsymbol{F}\end{bmatrix}\begin{Bmatrix}\boldsymbol{x}\\ \dot{\boldsymbol{x}}\end{Bmatrix} + \begin{Bmatrix}0\\ \boldsymbol{f}\end{Bmatrix} \tag{14}$$

The bar is modeled using five linear finite elements with unit nondimensional geometric and material properties ($A = \Delta L = E = \rho = 1.0$), leading to the following mass and stiffness matrices:

$$\boldsymbol{M} = \begin{bmatrix}1 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0.5\end{bmatrix}$$

$$\boldsymbol{K} = \begin{bmatrix}2 & -1 & 0 & 0 & 0\\ -1 & 2 & -1 & 0 & 0\\ 0 & -1 & 2 & -1 & 0\\ 0 & 0 & -1 & 2 & -1\\ 0 & 0 & 0 & -1 & 1\end{bmatrix}$$

The damping matrix is assumed to be proportional to the stiffness matrix: $\boldsymbol{F} = 0.005 \times \boldsymbol{K}$. The time step for discretization is $\Delta t = 0.15$. The discrete-time system matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are given by

$$\begin{aligned}\boldsymbol{A} &= \exp\left(\Delta t\begin{bmatrix}0 & \boldsymbol{I}\\ -\boldsymbol{M}^{-1}\boldsymbol{K} & -\boldsymbol{M}^{-1}\boldsymbol{F}\end{bmatrix}\right)\\ \boldsymbol{B} &= \boldsymbol{A}[0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1]^T\\ \boldsymbol{C} &= [0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]\end{aligned} \tag{15}$$

This system has five natural modes, with discrete-time eigenvalues of $0.9998 \pm 0.0469i$, $0.9904 \pm 0.1357i$, $0.9769 \pm 0.2104i$, $0.9633 \pm 0.2638i$, and $0.9550 \pm 0.2916i$, corresponding to periods of 134, 46, 30, 24, and 21 time steps, respectively.

Consider first the case in which a very limited amount of data is available and a state-space realization is to be constructed using $r = s = 20$. Using $t_i = j_i = i$, the first 40 Markov parameters, representing approximately $\frac{1}{3}$ of a cycle of the dominant mode and almost a full cycle of the second mode of the system, are required in the construction of $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$. In this case, the first few singular values of the Hankel matrix are

$$\operatorname{diag}(\boldsymbol{D}) = \begin{Bmatrix}1.5055\\ 0.2593\\ 0.2584\\ 0.0245\\ 0.0014\\ 0.0000\\ 0.0000\\ 0.0000\\ \vdots\end{Bmatrix}$$

If the Hankel matrix singular values are the only indicators used to determine the required order of the realized model using the criterion described above, a fourth-order model containing only the first

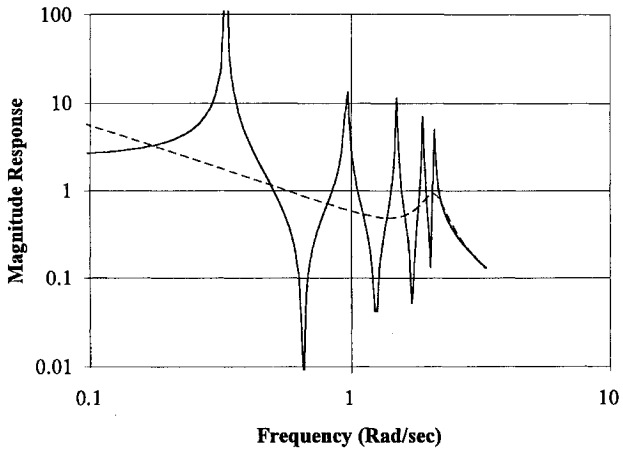

Fig. 2 Magnitude frequency response: ——, model $M_{10,\infty}$ (true model) and – – –, model $M_{4,40}$.

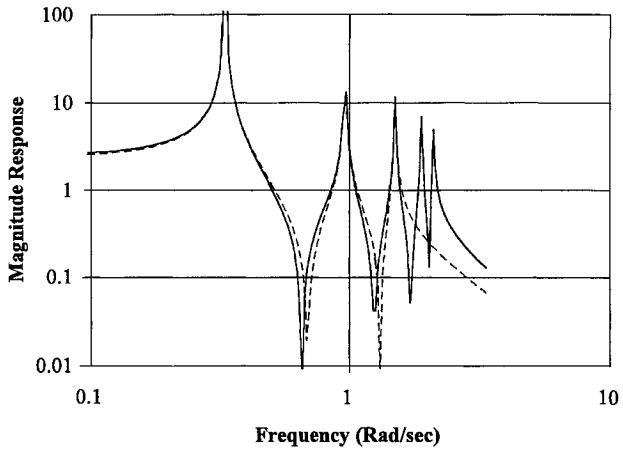

Fig. 3 Magnitude frequency response: ——, model $M_{10,\infty}$ (true model) and – – –, model $M_{6,2000}$.

four states would be expected to perform very well. This model will be referred to as $M_{4,40}$. The first subscript indicates the order of the model, and the second subscript indicates the number of Markov parameters used to construct the model. The frequency response of model $M_{4,40}$ is plotted in Fig. 2, along with the exact 10th-order system response (model $M_{10,\infty}$). Indeed, model $M_{4,40}$ is an extremely good approximation at high frequencies (the frequencies most visible in the limited data window), but the low-frequency performance misses the largest peaks in the frequency response. The eigenvalues of this model are 0.9970, 0.7582, and $0.9292 \pm 0.2828i$. Note that model $M_{4,40}$ has nonoscillatory eigenvalues, which do not exist in the physical system.

Consider now a case in which a large amount of data is available, $r = s = 1000$. The construction of $\boldsymbol{A}, \boldsymbol{B}$, and $\boldsymbol{C}$ requires the first 2000 Markov parameters, representing many cycles of each mode of the system. In this case, the first few singular values of $\boldsymbol{H}_{rs}(0)$ are

$$\operatorname{diag}(\boldsymbol{D}) = \begin{Bmatrix} 46.3003 \\ 46.1442 \\ 12.3867 \\ 12.2322 \\ 5.5083 \\ 5.4491 \\ 3.1968 \\ 3.1664 \\ 2.4427 \\ 2.4185 \\ 0 \\ \vdots \end{Bmatrix}$$

Here, there are measurable contributions from the first 10 states (as expected for a 10th-order system), but only the first six appear to be important. Therefore, we decide that a sixth-order model (model $M_{6,2000}$) will be constructed. The frequency response of model $M_{6,2000}$ is plotted in Fig. 3, along with the frequency response of model $M_{10,\infty}$. The eigenvalues of this model are $0.9989 \pm 0.0469i$, $0.9904 \pm 0.1357i$, and $0.9769 \pm 0.2103i$, which are essentially identical to the three lowest-frequency eigenvalues of the physical system. Model $M_{6,2000}$ provides a much better approximation than model $M_{4,40}$.

Because a sixth-order model (model $M_{6,2000}$) has been shown to work well, the reader's first reaction might be to go back to $r = s = 20$ and construct a sixth-order model based on the first 40 Markov parameters (model $M_{6,40}$). The frequency response of model $M_{6,40}$ is plotted against model $M_{10,\infty}$ in Fig. 4, and the eigenvalues of the model are 0.9963, 0.8841, $0.9477 \pm 0.2855i$, and $0.9064 \pm 0.2000i$. The performance is only slightly better than that of the fourth-order model.

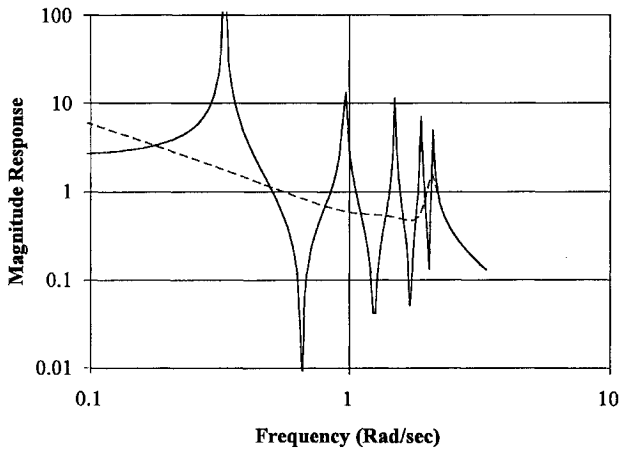

Fig. 4 Magnitude frequency response: ——, model $M_{10,\infty}$ (true model) and – – –, model $M_{6,40}$.

Rebalancing

To shed some light on why the singular values of the generalized Hankel matrix can be unreliable indicators of the relative importance of states when short data records are used and to motivate the development of a procedure with better performance, we define the finite-time generalized controllability and observability gramians as

$$\begin{aligned} \boldsymbol{W}^C_{\ell,s} &= \sum_{k=0}^{s-1} \boldsymbol{A}_\ell^{j_k} \boldsymbol{B}_\ell \boldsymbol{B}_\ell^T \left(\boldsymbol{A}_\ell^T\right)^{j_k} \\ \boldsymbol{W}^O_{\ell,r} &= \sum_{k=0}^{r-1} \left(\boldsymbol{A}_\ell^T\right)^{t_k} \boldsymbol{C}_\ell^T \boldsymbol{C}_\ell \boldsymbol{A}_\ell^{t_k} \end{aligned} \tag{16}$$

The definitions of $\boldsymbol{W}^C_{\ell,s}$ and $\boldsymbol{W}^O_{\ell,r}$ are more general than the usual definition of controllability and observability gramians. $\boldsymbol{W}^C_{\ell,s}$ and $\boldsymbol{W}^O_{\ell,r}$ reduce to the more conventional controllability and observability gramians when $j_i = i$ and $t_i = i$. If an untruncated triple $(\boldsymbol{A}_n, \boldsymbol{B}_n, \boldsymbol{C}_n)$ obtained using the ERA is used to formulate $\boldsymbol{W}^C_{n,s}$ and $\boldsymbol{W}^O_{n,r}$, then it can be shown that

$$\boldsymbol{W}^C_{n,s} = \boldsymbol{W}^O_{n,r} = \boldsymbol{D} \tag{17}$$

where $\boldsymbol{D}$ is the diagonal matrix of singular values in Eq. (8).

When the generalized controllability and observability gramians are equal and diagonal, the realization is said to be internally balanced. Internally balanced models have been studied by a number of researchers, and they have been shown to possess some useful properties.[10] In most cases, however, these studies focus on stable

models that are balanced with respect to infinite-time controllability and observability gramians defined as

$$W^C_{n,\infty} = \sum_{k=0}^{\infty} A_n^k B_n B_n^T \left(A_n^T\right)^k \qquad W^O_{n,\infty} = \sum_{k=0}^{\infty} \left(A_n^T\right)^k C_n^T C_n A_n^k \tag{18}$$

Models that are balanced with respect to $W^C_{n,s}$ and $W^O_{n,r}$ may not possess the useful properties of those balanced with respect to $W^C_{n,\infty}$ and $W^O_{n,\infty}$. One way to approximate infinite-time balancing is to let $j_i = i$ and $t_i = i$ and let r and s get very large, as in creating model $M_{4,2000}$. The difficulty with this is that the Hankel matrices $H_{rs}(0)$ and $H_{rs}(1)$ get very large and they require a large number of Markov parameters for their construction and a large computational effort for their decomposition. Fortunately, there is an alternative, which is to perform a transformation on the state-space model before the truncation is performed. The modified procedure is as follows.

First, $H_{rs}(0)$ is formed and decomposed as described earlier [Eq. (8)]. The minimum size of $H_{rs}(0)$ depends on the uncertainty (noise) in the Markov parameters and the order of the model. For the noise-free single-input, single-output case, the minimum required number of Markov parameters is twice the order of the system. Using P, D, and Q, form the largest possible state matrices (A_n, B_n, C_n), where $n = \text{rank}[H_{rs}(0)]$. The system described by this model is internally balanced with respect to the finite-time controllability and observability gramians. With a change of variables, a model can be produced that is internally balanced with respect to the infinite-time gramians and that offers improved performance when truncated. Consider a transformation of the form

$$\tilde{x} = T^{-1}x \tag{19}$$

The state matrices in the transformed coordinates can be expressed as

$$\tilde{A} = T^{-1}AT \qquad \tilde{B} = T^{-1}B \qquad \tilde{C} = CT \tag{20}$$

Substituting the new state matrices into Eqs. (18), it can easily be seen that the infinite-time controllability and observability gramians for the transformed coordinates are related to those in the original coordinates by the relationships

$$\tilde{W}^C_{n,\infty} = T^{-1}W^C_{n,\infty}T^{-T} \qquad \tilde{W}^O_{n,\infty} = T^T W^O_{n,\infty} T \tag{21}$$

The transformation matrix T that produces a system that is balanced with respect to the infinite-time controllability and observability gramians can be efficiently computed from the Cholesky factors of the infinite-time gramians.[11,12] The infinite-time gramians in the rebalanced coordinates are

$$\tilde{W}^C_{n,\infty} = \tilde{W}^O_{n,\infty} = \tilde{W} = \begin{bmatrix} \ddots & & \\ & w_i & \\ & & \ddots \end{bmatrix} \tag{22}$$

The w_i can then be used as measures of the relative importance of the various states but are free of the bias toward the data sample interval reflected in the singular values of $H_{rs}(0)$. The gramians and the state matrices can be partitioned as

$$\tilde{W} = \begin{bmatrix} \tilde{W}_\ell & \\ & \tilde{W}_{n-\ell} \end{bmatrix} \qquad \tilde{A} = \begin{bmatrix} \tilde{A}_\ell & \tilde{A}_{\ell,n-\ell} \\ \tilde{A}_{n-\ell,\ell} & \tilde{A}_{n-\ell} \end{bmatrix}$$
$$\tilde{B} = \begin{bmatrix} \tilde{B}_\ell \\ \tilde{B}_{n-\ell} \end{bmatrix} \qquad \tilde{C} = [\tilde{C}_\ell \quad \tilde{C}_{n-\ell}] \tag{23}$$

If the elements of $\tilde{W}_\ell$ are large compared to the elements of $\tilde{W}_{n-\ell}$, the system can be approximated by the triple $(\tilde{A}_\ell, \tilde{B}_\ell, \tilde{C}_\ell)$. Note the similarity to Eqs. (12).

Within the context of the ERA, the new procedure can be summarized as follows:

1) Choose appropriate values for r, s, t_i, and j_i. The choices of r and s must be large enough to accurately identify untruncated system matrices of sufficient order. This required size is much smaller than that required to produce accurate reduced-order models without rebalancing.

2) Form the generalized Hankel matrices $H_{rs}(0)$ and $H_{rs}(1)$ [Eq. (5)].

3) Using the singular-value decomposition, decompose $H_{rs}(0)$ [Eq. (8)].

4) Form the largest possible system matrices A_n, B_n, and C_n [Eqs. (11)].

5) Compute the infinite-time controllability and observability gramians $W^C_{n,\infty}$ and $W^O_{n,\infty}$ [Eqs. (18)].

6) Compute the transformation T and the diagonal infinite-time gramian $\tilde{W}$, and transform A_n, B_n, and C_n into their balanced counterparts $\tilde{A}_n$, $\tilde{B}_n$, and $\tilde{C}_n$ [Eqs. (20)].

7) Truncate the rebalanced system matrices based on the diagonal elements of $\tilde{W}$.

The following example illustrates the performance improvement that can be realized in the identified model using the new procedure.

Example of Rebalancing

To illustrate the improvement provided by this procedure, consider again the problem of modeling the discretized bar model using the first 40 Markov parameters ($r = s = 20$). After forming a 10th-order model and performing the rebalancing transformation, the first few diagonal elements of $\tilde{W}_\infty$ are

$$\text{diag}(\tilde{W}_\infty) = \begin{Bmatrix} 653.5073 \\ 652.4857 \\ 26.7607 \\ 26.6392 \\ 7.0810 \\ 7.0306 \\ 3.5279 \\ 3.4948 \\ 2.5843 \\ 2.5609 \\ 0 \\ 0 \end{Bmatrix}$$

Because the system is 10th order and there is no noise, rank $H_{rs}(0)$ is 10, even though $H_{rs}(0)$ is 20×20. The first six states appear to be important (as expected from the results of the case in which $r = s = 1000$). Retaining the most important states, a sixth-order model (model $\tilde{M}_{6,40}$) was constructed, and the frequency response is plotted against model $M_{10,\infty}$ in Fig. 5. The eigenvalues of this model are $0.9989 \pm 0.0469i$, $0.9904 \pm 0.1357i$, and $0.9769 \pm 0.2103i$, which are essentially identical to the three lowest-frequency eigenvalues of the physical system and to the eigenvalues of model $M_{6,2000}$. Model $\tilde{M}_{6,40}$ proves to be very accurate, surpassing even model $M_{6,2000}$.

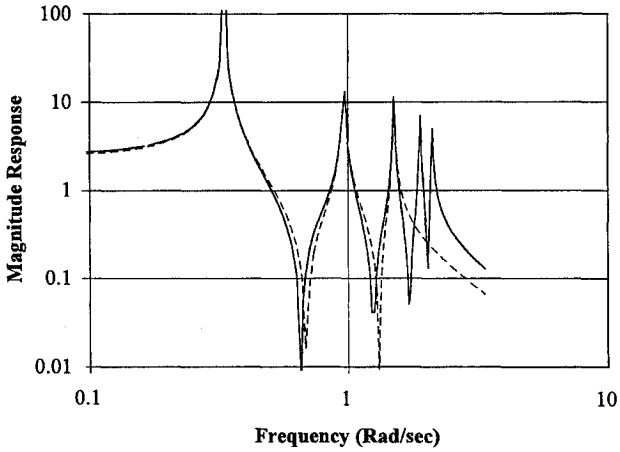

Fig. 5 Magnitude frequency response: ——, model $M_{10,\infty}$ (true model) and – – –, model $\tilde{M}_{6,40}$.

Conclusion

In off-line system identification algorithms, one of the most important problems is determining how many states are important to the input–output relationship of the system (the order of the identified model). Some current system identification techniques, such as the ERA, base the selection of important states on the relative magnitudes of the singular values of a generalized Hankel matrix. This results in state-space models that are internally balanced with respect to finite-time controllability and observability gramians, which are implicitly defined by the Markov parameters used to form the generalized Hankel matrix. If the order of the identified model is lower than the order of the physical system, unimportant states (states corresponding to small singular values of the generalized Hankel matrix) are neglected. If the amount of pulse-response data used in the procedure is insufficient, the criterion used to determine the model order and select the unimportant states can become unreliable and the truncation step can introduce significant errors into the identified model. It has been shown that truncating models that are balanced with respect to infinite-time gramians can produce results superior to those of truncating models that are balanced with respect to the implicitly defined finite-time gramians. This observation can be applied directly to the ERA, resulting in changes to the criterion for selecting model order and a modified definition of the identified system matrices. This removes an implicit bias toward the time interval represented by the pulse response data and can dramatically improve the fidelity of a reduced-order identified model when a limited amount of response data is available.

References

[1]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.

[2]Juang, J. N., *Applied System Identification*, Prentice–Hall, Englewood Cliffs, NJ, 1994, Chap. 5.

[3]Ljung, L., *System Identification: Theory for the User*, Prentice–Hall, Englewood Cliffs, NJ, 1987.

[4]Ho, B. L., and Kalman, R. E., "Effective Construction of Linear State Variable Models from Input/Output Data," *Proceedings of the 3rd Annual Allerton Conference on Circuit and System Theory*, 1965, pp. 449–459; also "Effective Construction of Linear State-Variable Models from Input/Output Functions," *Regelungstechnik*, Vol. 14, 1966, pp. 545–548.

[5]Juang, J.-N., Phan, M., Horta, L. G., and Longman, R. W., "Identification of Observer/Kalman Filter Markov Parameters: Theory and Experiments," *Journal of Guidance, Control, and Dynamics*, Vol. 16, No. 2, 1993, pp. 320–329.

[6]Press, W. H., et al., *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge Univ. Press, Cambridge, England, UK, 1988.

[7]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.

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

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

[10]Moore, B. C., "Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction," *IEEE Transactions on Automatic Control*, Vol. AC-26, No. 1, 1981, pp. 17–31.

[11]Horta, L. G., Juang, J. N., and Longman, R. W., "Discrete-Time Model Reduction in Limited Frequency Ranges," *Journal of Guidance, Control, and Dynamics*, Vol. 16, No. 6, 1993, pp. 1125–1130.

[12]Laub, A. J., Heath, M. T., Paige, C. C., and Ward, R. C., "Computation of System Balancing Transformations and Other Applications of Simultaneous Diagonalization Algorithms," *IEEE Transactions on Automatic Control*, Vol. AC-32, No. 2, 1987, pp. 115–122.