AIAA 2001-2611

BALANCED MODEL REDUCTION VIA THE PROPER ORTHOGONAL DECOMPOSITION

K. Willcox and J. Peraire
Fluid Dynamics Research Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139

15th AIAA Computational Fluid Dynamics Conference
11-14 June 2001
Anaheim, CA

For permission to copy or republish, contact the American Institute of Aeronautics and Astronautics
1801 Alexander Bell Drive, Suite 500, Reston, VA 20191–4344

BALANCED MODEL REDUCTION VIA THE PROPER ORTHOGONAL DECOMPOSITION

K. Willcox*and J. Peraire†
Fluid Dynamics Research Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

A new method for performing a balanced reduction of a high-order linear system is presented. The technique combines the proper orthogonal decomposition and concepts from balanced realization theory. The method of snapshots is used to obtain low-rank, reduced-range approximations to the system controllability and observability grammians in either the time or frequency domain. The approximations are then used to obtain a balanced reduced-order model. The method is demonstrated for a linearized high-order system which models unsteady motion of a two-dimensional airfoil. Computation of the exact grammians would be impractical for such a large system. For this problem, very accurate reduced-order models are obtained which capture the required dynamics with just three states. The new models exhibit far superior performance than those derived using a conventional proper orthogonal decomposition. Although further development is necessary, the concept also extends to nonlinear systems.

Introduction

Model reduction is a powerful tool which has been applied throughout many different disciplines, including controls, fluid dynamics and structural dynamics. In many situations, high-order, complicated numerical models accurately represent the problem at hand, but are unsuitable for the desired application, for instance for optimization or for control design. Ideally, we would like to develop a model with a low number of states, but which captures the system dynamics accurately over a range of frequencies and forcing inputs. This can be achieved via reduced-order modeling in which a high-order, high-fidelity model is projected onto a reduced-space basis. If the basis is chosen appropriately, the relevant high-fidelity system dynamics can be captured with a greatly reduced number of states. The range of validity of the reduced-order model is determined by the specifics of the reduction procedure.

Many methods have been suggested for determining an appropriate basis. Much of the work has been derived in a controls context. In particular, the idea of a balanced truncation has been shown to provide accurate low-order representations of state-space systems.[1] In order to determine the balanced realization, it is necessary to compute the grammians of the system, which use information pertaining to both system inputs and outputs. While it is relatively straightforward to compute these matrices in a controls setting where the system order is moderate, the technique does not extend easily to high-order systems, where state orders exceed 10^4. For this reason, many of the control-based reduction concepts have not been transferred to other disciplines where model order is typically much higher, such as computational fluid dynamics (CFD). Several methods have been developed for computing approximations to the grammians for large systems, including the approximate subspace iteration,[2] least squares approximation[3] and Krylov subspace methods,[4,5] however these algorithms are complicated, computationally intensive and restricted to linear systems. As an alternative means of performing the reduction, Padé approximations have also been used to approximate the system transfer function,[6] however the resulting reduced-order models often suffer from instability.

The challenge has therefore been to develop effective reduction procedures suitable for very high-order systems. One possibility for a basis is to compute the eigenmodes of the system.[7–9] Along with the use of static corrections,[10] this approach can lead to efficient models and the eigenmodes themselves often lend physical insight to the problem. However, for these high-order systems, solution of such a large non-symmetric eigen-problem is in itself a very difficult task, and in many cases not a viable option. The proper orthogonal decomposition technique (POD), also known as Karhunen-Loéve expansions,[11] has been developed as an alternate method of deriving basis vectors for high-order systems, and in particular has

*kwillcox@mit.edu, Member AIAA
†peraire@mit.edu, Associate Fellow AIAA

Copyright © 2001 by Massachusetts Institute of Technology. Published by the American Institute of Aeronautics and Astronautics, Inc. with permission.

been widely applied to fluid dynamic problems.[12–14] Frequency domain POD methods have also been developed and applied to a variety of flow problems.[15–17] However, in existing applications, only information pertaining to system inputs has been considered.

Here a method will be presented which allows both inputs and outputs to be taken into account to obtain a balanced reduced-order model. Lall et al.[18] noted the connection between the system grammians and the POD, and used a Kahunen-Loéve decomposition to obtain an approximate balanced truncation. Here, we make use of a similar concept which does not require the construction of the approximate grammians, which in our case would be computationally prohibitive. Instead, the POD method of snapshots will be used to approximate the grammians of the system in a very efficient way which does not require large computations or complicated algorithms. The method can also be implemented in the frequency domain, making it even more computationally efficient.

In this paper, the existing concepts of POD and balanced realization will be outlined. The new method which combines the two approaches will then be presented. Results will be shown for reduction of two high-order systems. The first case analyzed is a randomly generated state-space system whose exact balanced realization can be computed, allowing some insight to the performance of the method. The second example is the reduction of a CFD model which describes the unsteady linearized motion of a two-dimensional airfoil. In this case, reduction results will be compared to a full simulation of the CFD model and also to a conventional POD reduction approach. We then briefly discuss extension of the methodology to nonlinear problems, and finally we present some conclusions.

Model Order Reduction

Consider an n^{th}-order linear system

$$\begin{aligned} \dot{\mathbf{x}} &= A\mathbf{x} + B\mathbf{u} && (1)\\ \mathbf{y} &= C\mathbf{x}, && (2) \end{aligned}$$

where $\mathbf{x}$ is the state vector, the vectors $\mathbf{u}$ and $\mathbf{y}$ contain the system inputs and outputs respectively, and the order of the system, n, is high. The objective of the reduction procedure is to determine an n_r^{th}-order reduced-space basis onto which the state vector can be projected, that is $\mathbf{x} = V\mathbf{x_r}$, and an orthonormal set $\tilde{V}$, so that $\tilde{V}V = I$. This basis is chosen appropriately so that the reduced-order system

$$\begin{aligned} \dot{\mathbf{x}}_r &= \tilde{V}AV\mathbf{x}_r + \tilde{V}B\mathbf{u} && (3)\\ \mathbf{y}_r &= CV\mathbf{x}_r && (4) \end{aligned}$$

accurately reproduces the desired dynamics of the original system (1,2) with many fewer states ($n_r << n$).

Balanced Truncation

The concept of a balanced truncation of a system was first introduced by Moore.[1] The underlying idea is to take account of both the inputs and outputs of the system when determining which states to retain in the reduced-state representation, but to do so with appropriate internal scaling. This scaling is important, since a particular representation of the system is not unique: any non-singular linear transformation can be applied to the system (1,2). For example, if we choose the transformation $\mathbf{x} = T\mathbf{x_t}$, we obtain the scaled system

$$\begin{aligned} \dot{\mathbf{x}}_t &= T^{-1}AT\mathbf{x}_t + T^{-1}B\mathbf{u} && (5)\\ \mathbf{y} &= CT\mathbf{x}_t, && (6) \end{aligned}$$

which is fully equivalent to (1,2). In a balanced truncation, we therefore seek a reduction method which is independent of the particular system scaling.

Reduction of the system will be achieved by retaining only certain states in the representation. This is equivalent to defining a certain subspace within the state space. Two important subspaces are the controllable and observable subspaces. The controllable subspace is that set of states which can be obtained with zero initial state and a given input $\mathbf{u}(t)$, while the observable subspace comprises those states which as initial conditions could produce a non-zero output $\mathbf{y}(t)$ with no external input. The controllability and observability grammians are each an $n \times n$ matrix whose eigenvectors span the controllable and observable subspaces respectively. These matrices are defined for the linear system (1,2) as

$$W_c = \int_0^\infty e^{At}BB^*e^{A^*t}dt \qquad (7)$$

and

$$W_o = \int_0^\infty e^{A^*t}C^*Ce^{At}dt. \qquad (8)$$

By noting that for a single-input, single-output (SISO) system the quantity $\mathbf{x}_\delta(t) = e^{At}B$ is simply the impulse response of the system (set $u(t) = \delta(t)$ in (1)), the controllability grammian can also be written

$$W_c = \int_0^\infty \mathbf{x}_\delta(t)\mathbf{x}_\delta^*(t)dt. \qquad (9)$$

For the observability grammian, we need to consider the dual system of (1,2):

$$\begin{aligned} \dot{\mathbf{z}} &= A^*\mathbf{z} + C^*\mathbf{u}_d && (10)\\ \mathbf{y}_d &= B^*\mathbf{z}. && (11) \end{aligned}$$

Here, $\mathbf{z}$ is the dual state vector. Analogously to (9), we can write

$$W_o = \int_0^\infty \mathbf{z}_\delta(t)\mathbf{z}_\delta^*(t)dt, \qquad (12)$$

where $\mathbf{z}_\delta(t) = e^{A^* t} C^*$ is the impulse response of the dual SISO system.

To obtain a balanced realization of the system (1,2), a state transformation T is chosen so that the controllability and observability grammians are diagonal and equal. This so-called balancing transformation can be computed by first calculating the matrix $W_{co} = W_c W_o$, and then determining its eigenmodes:

$$W_{co} = T^{-1} \Lambda T. \tag{13}$$

The eigenvectors of W_{co}, $\mathbf{T}_i$, are the basis vectors which describe the balancing transformation. The eigenvalues, λ_i, contained in the diagonal matrix Λ, are positive, real numbers,[19] and $\sigma_i = \sqrt{\lambda_i}$ are known as the Hankel singular values of the system. The eigenmodes of W_{co} correspond to states through which the input is transmitted to the output. The magnitudes of the Hankel singular values describe the relative importance of these states, and are independent of the particular realization of the system. In a balanced truncation, only those states are retained which correspond to large Hankel singular values.

Assuming the n^{th} order system (1,2) has been transformed to a balanced realization, an error criterion for model reduction based on Hankel singular values can be derived.[19] A truncation of the balanced system is performed in which the first n_r states are retained, resulting in a reduced-order model of the form (3,4). The Hankel singular values of the neglected states give an error bound on the output:

$$||\mathbf{y}(t) - \mathbf{y}_r(t)|| \;\leq\; 2 \sum_{i=n_r+1}^{n} \sigma_i \; ||\mathbf{u}(t)||, \tag{14}$$

where $||.||$ denotes the L_2 norm.

For large systems, it is not practical to explicitly compute the grammians using (7) and (8). It is more convenient to note that W_c and W_o satisfy the Lyapunov equations

$$\begin{aligned} AW_c + W_c A^* + BB^* &= 0 \quad (15)\\ A^* W_o + W_o A + C^* C &= 0. \quad (16)\end{aligned}$$

Methods have been suggested for solving (15) and (16) when the systems are large using approximate subspace iteration,[2] least squares approximation,[3] and Krylov subspace methods[4,5] to obtain low-rank approximations of the grammians. However, all of these techniques are expensive for very high-order systems and have only been demonstrated for much smaller problems than those encountered in complicated fluid dynamic applications. In this paper we will introduce an efficient method for calculating very low-rank approximations to the grammians using the concepts of the POD which are outlined in the following section.

Proper Orthogonal Decomposition

The POD has been widely used to determine efficient bases for dynamic systems. It was introduced for the analysis of turbulence by Lumley,[12] and is also known as the Karhunen-Loéve decomposition[11] and principal component analysis.[20] The basis vectors $\mathbf{\Psi}$ are chosen so as to maximise the following cost:[14]

$$\max_{\mathbf{\Phi}} \frac{\langle |(\mathbf{x}, \mathbf{\Phi})|^2 \rangle}{(\mathbf{\Phi}, \mathbf{\Phi})} = \frac{\langle |(\mathbf{x}, \mathbf{\Psi})|^2 \rangle}{(\mathbf{\Psi}, \mathbf{\Psi})}, \tag{17}$$

where $(\mathbf{x}, \mathbf{\Psi})$ denotes the scalar product of the basis vector $\mathbf{\Psi}$ with the field $\mathbf{x}(\theta, t)$ which depends on the spatial coordinates θ and time t, and $\langle\,\rangle$ represents a time-averaging operation. It can be shown that a necessary condition for (17) to hold is that $\mathbf{\Psi}$ is an eigenfunction of the kernel K defined by

$$K(\theta, \theta') = \langle \mathbf{x}(\theta, t) \;\, \mathbf{x}^*(\theta', t) \rangle. \tag{18}$$

Sirovich introduced the method of snapshots as a way of determining the eigenfunctions $\mathbf{\Psi}$ without explicitly calculating the kernel K.[13] The kernel can be approximated as

$$K(\theta, \theta') = \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}_i(\theta) \mathbf{x}_i^*(\theta'), \tag{19}$$

where $\mathbf{x}_i(\theta)$ is the instantaneous system state or "snapshot" at a time t_i and the number of snapshots m is sufficiently large. The eigenvectors of K are of the form

$$\mathbf{\Psi} = \sum_{i=1}^{m} \beta_i \mathbf{x}_i, \tag{20}$$

where the constants β_i can be seen to satisfy the eigenvector equation

$$R\beta = \Lambda \beta \tag{21}$$

and R is now the correlation matrix

$$R_{ik} = \frac{1}{m} (\mathbf{x}_i, \mathbf{x}_k). \tag{22}$$

Rather than performing a set of simulations to obtain the snapshots $\mathbf{x}_i$, the POD basis vectors can be obtained much more efficiently by taking advantage of linearity and the frequency domain. For a linear system, any general forcing function can be considered as a superposition of sinusoidally time-varying components each at a frequency ω:

$$\mathbf{u}(t) = Re \left\{ \int_{-\infty}^{\infty} \overline{\mathbf{u}}(\omega) e^{j\omega t} d\omega \right\}. \tag{23}$$

Because the system is linear, the component of forcing at frequency ω induces a response which is also harmonic with frequency ω, that is $\mathbf{x}(t) = \overline{\mathbf{x}} e^{j\omega t}$ and $\mathbf{y} = \overline{\mathbf{y}} e^{j\omega t}$. The response due to each harmonic component could be computed separately and then recombined appropriately to obtain the overall response to the general forcing function. Considering a single temporal harmonic, ω_k, the state-space system (1,2) can be written in the frequency domain as

$$\begin{aligned} \overline{\mathbf{x}}_k &= (j\omega_k I - A)^{-1} B\overline{\mathbf{u}}_k & (24) \\ \overline{\mathbf{y}}_k &= C\overline{\mathbf{x}}_k. & (25) \end{aligned}$$

The POD snapshots can be obtained by choosing a set of sample frequencies $\{\omega_k\}$ based on the frequency content of problems of interest.[21] The frequency domain system (24) can then be solved to obtain the responses $\{\overline{\mathbf{x}}_k\}$. The real and imaginary parts of each $\overline{\mathbf{x}}_k$ constitute the snapshots.

Balanced Truncation via the Method of Snapshots

Lall et al.[18] describe the connection between the POD and balanced truncation. Note the similarity between the POD kernel function K defined by (18) and the controllability grammian W_c defined by (9). In fact, as noted by Lall, if the fields $\mathbf{x}(\theta, t)$ in (18) are obtained by exciting the system with impulsive inputs, then the POD results in the construction of the controllability grammian. Further insight can be gained by thinking about the problem in the frequency domain. We begin with the frequency domain definition of the controllability grammian written as:[22]

$$W_c = \frac{1}{2\pi}\int_{-\infty}^{\infty} (j\omega I - A)^{-1} BB^* (-j\omega I - A^*)^{-1} d\omega. \tag{26}$$

We note from (24) that the term $(j\omega I - A)^{-1}B$ is the response of the linear system to sinusoidal forcing at a frequency ω. Since an impulsive input contains an equal amount of all frequencies ($\delta(t) = \int_{-\infty}^{\infty} e^{j\omega t} d\omega$), equation (26) is simply another way of thinking about (9). Although the kernel is never explicitly computed in the POD frequency domain analysis, by choosing a finite set of discrete frequencies for the snapshots, (19) can be written

$$K = \frac{1}{m}\sum_{i=1}^{m} (j\omega_i I - A)^{-1} BB^* (-j\omega_i I - A^*)^{-1}. \tag{27}$$

By comparing equations (26) and (27), we can see that in the case of general inputs, the POD kernel is therefore an approximation to the controllability grammian over a chosen, restricted frequency range. The subspace spanned by the POD basis vectors approximates the controllability subspace.

It is a natural extension to consider a POD analysis which approximates the observability subspace. Furthermore, to obtain a balanced representation of the system, we can then use concepts from a traditional control balanced truncation. Lall et al. used the direct POD method to obtain approximations to the system grammians. For a system of order n, this results in the construction of two $n \times n$ matrices. Clearly for very large systems this approach will be computationally infeasible, especially given that the matrices will not be sparse. Here we present an alternative approach which uses the POD method of snapshots to approximate the grammians, so that the large matrices need never be explicitly computed.

By obtaining snapshots of the dual system (10,11), and performing the POD method of snapshots analysis described above, we can calculate p eigenmodes of the observability kernel function. Let these eigenvectors be contained in the columns of the matrix X, with corresponding eigenvalues on the diagonal entries of the matrix Λ_o. Similarly, let the eigenvectors of the conventional (controllability) kernel K be contained in the columns of the matrix Y, with corresponding eigenvalues on the diagonal entries of the matrix Λ_c. Low-rank approximations to the controllability and observability grammians can then be made as follows:

$$\begin{aligned} W_c^p &= Y\Lambda_c Y^* & (28) \\ W_o^p &= X\Lambda_o X^*, & (29) \end{aligned}$$

where the superscript p denotes a pth order approximation.

Through use of an efficient eigenvalue solver, the eigenmodes of the product $W_c^p W_o^p$ can then be calculated. In this work, ARPACK[23] was used to determine the eigenvalues. This package requires the user to supply only matrix-vector multiplications, hence the large matrices W_c^p and W_o^p need never be explicitly formed.

The balancing algorithm can therefore be summarized as:

1. Use method of snapshots to obtain p POD eigenmodes (Y, Λ_c) for the primal system.
2. Use method of snapshots to obtain p POD eigenmodes (X, Λ_o) for the dual system.
3. Formulate the low-rank approximations $W_c^p = Y\Lambda_c Y^*$ and $W_o^p = X\Lambda_o X^*$ (the $n \times n$ matrices are never explicitly calculated).
4. Obtain the eigenvectors of the product $W_c^p W_o^p$ to determine the balancing transformation T.
5. Retain only those eigenvectors in the reduced-space basis which correspond to large Hankel singular values.

Multiple Input/Output Case

The concept extends readily to the MIMO case, however it is important to treat the system in the correct manner if the correlation with the grammians is to be maintained.

Consider a system with q inputs

$$\mathbf{u} = [u_1 \ u_2 \ \ldots u_q]^T. \tag{30}$$

The matrix B in equation (1) can be written

$$B = [\mathbf{b}_1 \ \mathbf{b}_2 \ \ldots \mathbf{b}_q]. \tag{31}$$

We now inspect the form of the controllability grammian W_c defined by equation (7). Due to the nature of the outer product, the grammian of the multiple-input system can be written as a sum of grammian components which correspond to each of the inputs as follows:

$$W_c = \int_0^\infty e^{At}\mathbf{b}_1\mathbf{b}_1^* e^{A^*t}dt + \int_0^\infty e^{At}\mathbf{b}_2\mathbf{b}_2^* e^{A^*t}dt + \ldots + \int_0^\infty e^{At}\mathbf{b}_q\mathbf{b}_q^* e^{A^*t}dt. \quad (32)$$

The kernel function (18) should therefore be written as

$$K(\theta,\theta') = \langle \mathbf{x}^1(\theta,t)\ \mathbf{x}^{1*}(\theta',t) + \mathbf{x}^2(\theta,t)\ \mathbf{x}^{2*}(\theta',t) + \ldots \mathbf{x}^q(\theta,t)\ \mathbf{x}^{q*}(\theta',t)\rangle, \quad (33)$$

where $\mathbf{x}^j$ is the response of the system to forcing in u_j only. As described earlier, the eigenfunctions $\mathbf{\Psi}$ of the kernel can be written as linear combinations of snapshots

$$\mathbf{\Psi} = \sum_{j=1}^{q}\sum_{i=1}^{m_j} \beta_i^j \mathbf{x}_i^j, \quad (34)$$

where the number of snapshots m_j can vary for different inputs j. Following the derivation in Sirovich,[13] we obtain an eigenvalue problem for the coefficients β_i^j. We find that the resulting system has an identical form to (21), with the total number of snapshots now being given by $m = \sum_{j=1}^{q} m_j$.

The POD can therefore be applied to a multiple-input problem, and the approximation of the controllability grammian will be maintained provided snapshots are obtained for each input in turn. The resulting collection of snapshots, $\mathbf{x}_i^j, i = 1...m_j, j = 1...q$, is then treated in the same way as for the SISO case. Analogous arguments can be applied to the dual problem.

Results and Discussion

In this section, the performance of the method will be illustrated with two examples. The first is a randomly generated, moderately sized problem for which the exact balanced realization can be computed. The second is a realistic high-order fluid dynamic problem.

Randomly Generated State-Space System

In this example we analyze a randomly generated single-input single-output system of size $n = 100$. The matrix was chosen to be diagonal with eigenvalues distributed uniformly over the interval $[-1\ 0]$. The vectors $\mathbf{B}$ and $\mathbf{C}$ were also randomly generated. By solving the Lyapunov equations, the exact balanced realization of the system was determined. The Hankel singular values were computed, and the first ten are plotted in Figure 1. As the figure shows, the magnitudes of the Hankel singular values decrease very rapidly. This indicates that a balanced truncation of the system could provide a very accurate representation with just a few states.

The approximate balancing method was then applied to the system. POD snapshots were taken from the primal and dual systems at frequency intervals of $\Delta\omega = 0.05$ from $\omega = 0$ to $\omega = 1$. Twenty POD basis vectors were then calculated for each system and used to form the approximations to the appropriate grammian. The resulting square roots of the eigenvalues of the grammian product are also plotted in Figure 1, and we see that the method approximates the dominant Hankel singular values very well.

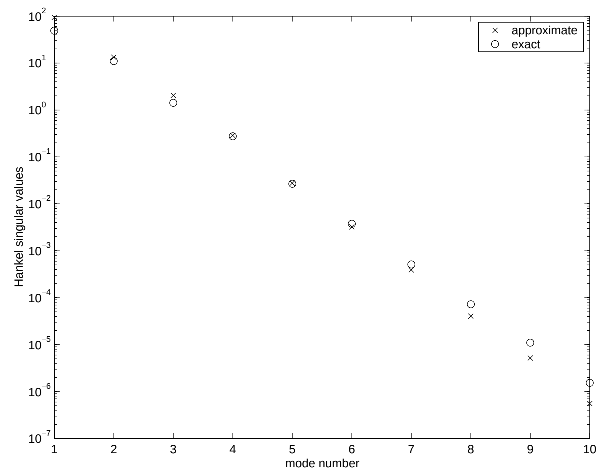

Fig. 1 Hankel singular values (o) and square roots of eigenvalues of the approximate grammian product (x), $n = 100$, $p = 20$, $m = 40$.

CFD Model

Results will now be presented for a two-dimensional NACA 0012 airfoil operating in unsteady plunging motion with a steady-state Mach number of 0.755. The flow is assumed to be inviscid, so the governing equations are the linearized Euler equations. The steady-state pressure contours for this problem are shown in Figure 2. The CFD mesh has 3482 grid points, which corresponds to a total of $n = 13928$ unknowns in the linear state-space system.

POD snapshots were obtained by causing the airfoil to plunge in sinusoidal motion at selected frequencies.[21] Frequencies were selected at 0.1 increments from $\omega = 0.1$ to $\omega = 2.0$. Snapshots were obtained at each frequency by solving the frequency domain equations (24) and the equivalent frequency domain equations for the dual system with a preconditioned complex GMRES algorithm. Thus, forty snapshots were obtained for each problem (two per frequency). The correlation matrices were calculated and the POD process used to determine the kernel eigenfunctions.

The low-rank approximations to the grammians were formed by taking fifteen eigenmodes for each ($p = 15$). ARPACK was then used to calculate the first ten eigenmodes of the grammian product. As

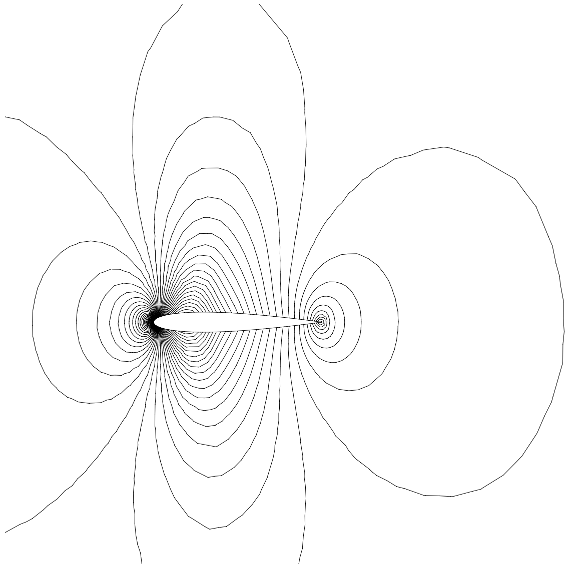

Fig. 2 Steady-state pressure contours for NACA 0012 airfoil. 3482 nodes, $M = 0.755$, $\alpha = 0.016°$.

mentioned previously, it is not necessary to explicitly form the grammians or the product. Instead, for each problem, fifteen eigenvectors, each of size n, and fifteen eigenvalues were stored and the matrix multiplications were computed as necessary. Because of the extremely low-rank approximation of the matrix, the eigenvalue solver converged very quickly. The resulting first ten eigenvalues of the grammian product are plotted with crosses in Figure 3. These eigenvalues approximate the squares of the Hankel singular values of the system. In a balanced truncation, only those states are retained which correspond to large Hankel singular values. From Figure 3, we see that the magnitudes of eigenvalues drop off very sharply, indicating that the reduced-order model will require only a few states. In fact, the first state contains most of the system "energy".

The accuracy of the reduced-order model obtained from the balanced truncation can be assessed via simulation results. Forced response of the airfoil to a pulse input in plunge is considered and the results are compared to those obtained both with the high-order CFD code and with a reduced-order model derived via conventional POD. Static corrections were also included in the reduced-order models to aid in capturing high-frequency dynamics.[21] The plunge displacement of the airfoil was prescribed to be

$$h(t) = e^{-g(t-t_0)^2}, \tag{35}$$

where g is a parameter which determines how sharp the pulse is, and thus the value of the maximum significant frequency present. Figure 4 shows the results for $g = 0.01$ which corresponds to $\omega_{\max} = 0.48$ based on a 1% level. The solid line represents the force generated on the airfoil as a function of time as calculated with

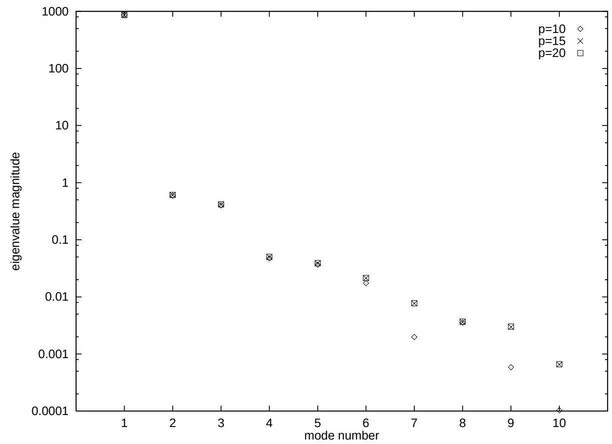

Fig. 3 Eigenvalues of the approximate grammian product (approximate the squares of the Hankel singular values of the system). $n = 13928$, $m = 40$, $p = 10, 15, 20$.

the high-order CFD model. The two dashed lines are the results obtained using a conventional POD model with eight states and the new balanced model with one state. As the figure shows, with just a single degree of freedom, the balanced reduced-order model captures the response almost exactly.

The same test was performed for a higher frequency pulse with $g = 0.1$. In this case the highest significant frequency present at a 1% level is $\omega_{\max} = 1.34$. The frequency content in this pulse input therefore spans most of the range sampled by the snapshots. Figure 5 shows the results for the CFD model, along with the reduced-order models with eight (conventional) and three (balanced) degrees of freedom. While the balanced model with three states captures the response extremely accurately, even with eight states the conventional model shows a significant error. This was the highest order model which could be obtained using conventional POD since including additional basis vectors caused the reduced-order model to become unstable. In order to more accurately capture the response, it would be necessary to include more snapshots in the conventional POD analysis.

Discussion

The computation of reduced-order models via the balancing method is slightly more than twice as expensive as conventional POD. The greatest cost occurs in the system solves required to obtain the snapshots. Because we must solve the dual system, twice as many snapshots are required for the balanced POD approach. Additionally there is the small cost associated with calculating the eigenmodes of the grammian product. In many applications however, one is less concerned with the cost of obtaining the reduced-order model, and more so with the resulting size and quality of the model. As the results above show, by

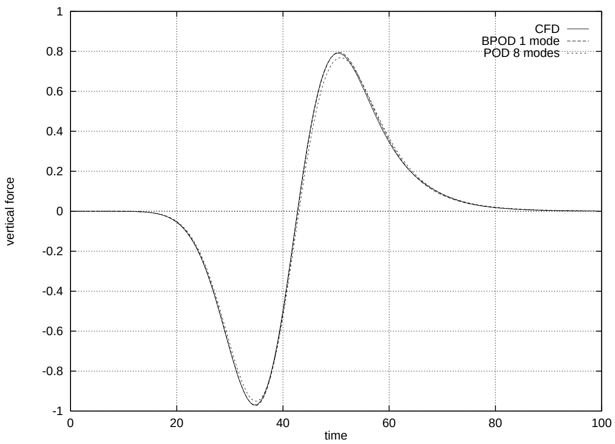

Fig. 4 Response of NACA 0012 airfoil to a pulse input in plunge. $M = 0.755$, $\alpha = 0.016°$, $g = 0.01$. **Results from CFD model (13928 states, solid line), conventional POD reduced-order model with eight states (small dash) and balanced reduced-order model with one state (larger dash). The results for the balanced model agree almost exactly with the CFD model.**

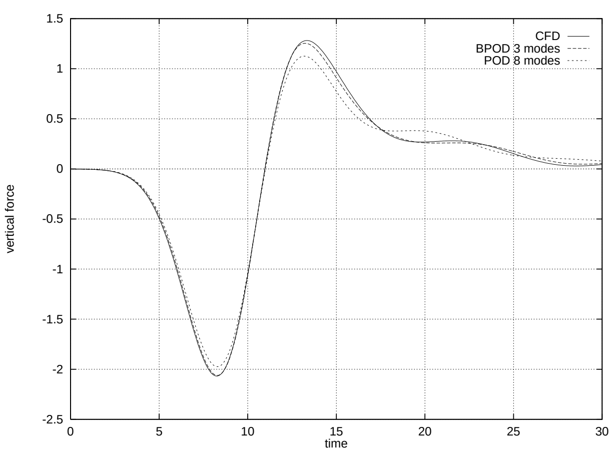

Fig. 5 Response of NACA 0012 airfoil to a pulse input in plunge. $M = 0.755$, $\alpha = 0.016°$, $g = 0.1$. **Results from CFD model (13928 states, solid line), conventional POD reduced-order model (eight states, small dash) and balanced reduced-order model (three states, larger dash).**

incorporating both input and output information very accurate models can be obtained which have a minimal number of states. The method outlined here is far more efficient than other attempts to calculate system grammians, and accurately provides the required level of information. Also, since the snapshots can be obtained efficiently in the frequency domain, this method is appropriate for problems with spatial symmetry, such as turbomachinery flows.

Throughout the algorithm, there are several arbitrary decisions to be made. Firstly, the snapshots must be selected. The range over which the sampling is performed is determined by assessing the important frequency range in the problems at hand. To determine the specific snapshot locations within this range, one uses a combination of experience and intuition. Often the required density of snapshots will be determined a posteriori from the performance of the reduced-order model. If the desired dynamics cannot be accurately captured, more snapshots must be included in the POD process and the basis vectors recalculated. It is therefore important to validate the models against known results (in this case against the CFD model). Secondly, p, the number of POD eigenmodes to be used in the low-rank approximation of the grammians, must be chosen. Again this will depend on the frequency range of interest, as well as on the number of modes to be retained in the reduced-order model. A fairly low number (fifteen) was chosen for the results presented here, however the performance of the models was found to be fairly robust with respect to this parameter. In Figure 3 the eigenvalues of the grammian product are also plotted for $p = 10$ and $p = 20$. By increasing the number of POD vectors to twenty, very little variation was seen in the eigenvalues. If only ten POD vectors were used, the first five eigenvalues were virtually unchanged while the next five showed some movement. This result is to be expected: in order to accurately resolve q eigenvalues of the grammian product, we should choose $p > q$ in the approximation of the matrices.

The method is flexible in that it can be applied as described to any linearized system. The approach also extends to nonlinear systems using concepts similar to those discussed by Lall et al.[18] It is relatively straightforward to obtain an approximation to the controllabillity subspace. For example, consider the nonlinear system

$$\dot{\mathbf{x}} = f\left(\mathbf{x}(\theta, t), \mathbf{u}(t)\right) \quad (36)$$

$$\mathbf{y} = g\left(\mathbf{x}(\theta, t)\right). \quad (37)$$

The POD eigenfunctions would be calculated using snapshots from simulation of the nonlinear system (36,37). The difficulty arises with the approximation of the observability subspace. The concept of a dual system does not exist in a nonlinear setting. Two possibilities present themselves. The first is to linearize the nonlinear system (36,37) and formulate the dual linearized system (the adjoint). The snapshots would then be obtained from a combination of nonlinear (primal) and linearized (dual) systems. The second approach is to follow the method outlined in[18] which defines an empirical observability grammian based on system outputs for various initial conditions. For the high-order systems encountered in CFD applications, this second approach, although more accurate, would be computationally very expensive. Work is underway to develop a better approach for handling large nonlinear systems.

Once the grammians have been suitably approximated, the balancing method can then be used to calculate a linear transformation of the nonlinear state to obtain the nonlinear reduced-order model

$$\dot{\mathbf{x}}_r = T^{-1} f\left(T\mathbf{x}_r(\theta,t), \mathbf{u}(t)\right) \quad (38)$$

$$\mathbf{y}_r = g\left(T\mathbf{x}_r(\theta,t)\right). \quad (39)$$

Conclusions

A new method for computing an approximate balanced truncation of a linear state-space system has been presented. By using the method of snapshots to perform a POD analysis of the primal and dual systems, low-rank, reduced-range approximations to the controllability and observability grammians are obtained very efficiently. This POD analysis can be performed either in the time or frequency domain. By incorporating information pertaining to both inputs and outputs, the resulting reduced-order models capture the desired system dynamics with a very low number of states. The required size of the models is significantly lower than for those developed using a conventional POD approach. Results have been presented for a very high-order system, and the method has been shown to work extremely effectively. The concept is applicable to general linearized systems and with some modifications can be extended to nonlinear systems.

Acknowledgments

Support for this work has been provided by the Singapore MIT Alliance.

References

[1]Moore, B., "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.

[2]Baker, M., Mingori, D., and Goggins, P., "Approximate Subspace Iteration for Constructing Internally Balanced Reduced Order Models of Unsteady Aerodynamic Systems," AIAA Paper 96-1441, 1996.

[3]Scottedward Hodel, A., "Least Squares Approximate Solution of the Lyapunov Equation," Proceedings of the 30th IEEE Conference on Decision and Control, December 1991.

[4]Jaimoukha, I. and Kasenally, E., "Krylov Subspace Methods for Solving Large Lyapunov Equations," *SIAM Journal of Numerical Analysis*, Vol. 31, No. 1, February 1994, pp. 227–251.

[5]Gudmundsson, T. and Laub, A., "Approximate Solution of Large Sparse Lyapunov Equations," *IEEE Transactions on Automatic Control*, Vol. 39, No. 5, May 1994, pp. 1110–1114.

[6]Gallivan, K., Grimme, E., and Van Dooren, P., "Padé Approximation of Large-Scale Dynamic Systems with Lanczos Methods," Proceedings of the 33rd IEEE Conference on Decision and Control, December 1994.

[7]Romanowski, M. and Dowell, E., "Using Eigenmodes to Form an Efficient Euler Based Unsteady Aerodynamics Analysis," Proceedings of the Special Symposium on Aeroelasticity and Fluid/Structure Interaction Problems, ASME International Mechanical Engineering Congress and Exposition, AD-Vol. 44, pp. 147-160, 1994.

[8]Mahajan, A., Dowell, E., and Bliss, D., "Eigenvalue Calculation Procedure for an Euler/Navier-Stokes Solver with Application to Flows over Airfoils," *Journal of Computational Physics*, Vol. 97, 1991, pp. 398–413.

[9]Hall, K., "Eigenanalysis of Unsteady Flows About Airfoils, Cascades and Wings," *AIAA Journal*, Vol. 32, No. 12, December 1994, pp. 2426–2432.

[10]Bisplinghoff, R. and Ashley, H., *Principles of Aeroelasticity*, Dover Publications, New York, NY, 1975.

[11]Loéve, M., *Probability Theory*, D. Van Nostrand Company Inc., New York, 1955.

[12]Lumley, J., "The Structures of Inhomogeneous Turbulent Flow," *Atmospheric Turbulence and Radio Wave Propagation*, 1967, pp. 166–178.

[13]Sirovich, L., "Turbulence and the Dynamics of Coherent Structures. Part 1 : Coherent Structures," *Quarterly of Applied Mathematics*, Vol. 45, No. 3, October 1987, pp. 561–571.

[14]Berkooz, G., Holmes, P., and Lumley, J., "The Proper Orthogonal Decomposition in the Analysis of Turbulent Flows," *Annual Review of Fluid Mechanics*, Vol. 25, 1993, pp. 539–575.

[15]Kim, T., "Frequency-Domain Karhunen-Loeve Method and Its Application to Linear Dynamic Systems," *AIAA Journal*, Vol. 36, No. 11, 1998, pp. 2117–2123.

[16]Hall, K. C., Thomas, J. P., and Dowell, E. H., "Reduced-Order Modelling of Unsteady Small-Disturbance Flows Using a Frequency-Domain Proper Orthogonal Decomposition Technique," AIAA Paper 99-0655, 1999.

[17]Willcox, K., Paduano, J., Peraire, J., and Hall, K., "Low Order Aerodynamic Models for Aeroelastic Control of Turbomachines," AIAA Paper 99-1467, 1999.

[18]Lall, S., Marsden, J., and Glavaski, S., "Empirical Model Reduction of Controlled Nonlinear Systems," Proceedings of the IFAC World Congress, 1999.

[19]Enns, D., *Model Reduction for Control System Design*, Ph.D. thesis, Stanford University, 1984.

[20]Hotelling, H., "Analysis of a Complex of Statistical Variables with Principal Components," *Journal of Educational Psychology*, Vol. 24, 1933, pp. 417–441,498–520.

[21]Willcox, K. E., *Reduced-Order Aerodynamic Models for Aeroelastic Control of Turbomachines*, Ph.D. thesis, Dept. of Aeronautics and Astronautics, MIT, February 2000.

[22]Zhou, K., Salomon, G., and Wu, E., "Balanced Realization and Model Reduction for Unstable Systems," *International Journal of Robust and Nonlinear Control*, Vol. 9, 1999, pp. 183–198.

[23]Lehoucq, R., Sorensen, D., and Yang, C., *ARPACK Users Guide: Solution of Large Scale Eigenvalue Problems by Implicitly Restarted Arnoldi Methods*, Rice University, October 1997.