Model Reduction for Large-Scale Applications in Computational Fluid Dynamics

K. Willcox*

1 Introduction

Recent years have seen considerable progress in solution and optimization methods for partial differential equations (PDEs), leading to advances across a broad range of engineering applications. Improvements in methodology, together with a substantial increase in computing power, are such that real-time simulation and optimization of systems governed by PDEs is now an attainable goal; however, a number of challenges remains for applications such as real-time control of dynamic processes. In many cases, computational models for such applications yield very large systems that are computationally intensive to solve. A critical element towards achieving a real-time simulation capability is the development of accurate, efficient models that can be solved sufficiently rapidly to permit control decisions in real time.

Model reduction is a powerful tool that allows the systematic generation of cost-efficient representations of large-scale systems resulting from discretization of PDEs. Reduction methodology has been developed and applied for many different disciplines, including controls, fluid dynamics, structural dynamics, and circuit design. Considerable advances in the field of model reduction for large-scale systems have been made and many different applications have been demonstrated with success; however, a number of open issues remain, including the reliability of reduction techniques, guarantees associated with the quality of the reduced models, and validity of the model over a range of operating conditions. The cost of performing the reduction may also be an issue if there is a need to adapt the reduced-order model

*K. Willcox, Associate Professor of Aeronautics and Astronautics, Massachusetts Institute of Technology, kwillcox@mit.edu

in real time.

In this paper, model reduction of computational fluid dynamic (CFD) systems will be discussed, although the techniques presented are applicable to general systems of PDEs. Two reduction methods will be discussed: the proper orthogonal decomposition (POD), and Fourier model reduction (FMR). These methods will be compared in the context of an active flow control application. Their relative advantages and disadvantages will be presented, with particular focus on the issues relevant to real-time implementation. Finally, the paper concludes with a discussion of outstanding issues and the open question of model reduction for nonlinear systems.

1.1 Problem Statement

The system of PDEs governing a general fluid flow can be discretized in the spatial domain using a CFD formulation to yield a set of nonlinear ordinary differential equations (ODEs). For unsteady flows, these ODEs can be linearized about the steady-state solution to obtain a linear CFD model that is valid for small deviations of the flow from steady-state conditions. A general linearized CFD model can be written as

$$G: \quad \frac{d}{dt}x(t) = Ax(t) + Bu(t), \qquad y(t) = Cx(t) + Du(t) \tag{1}$$

where $x(t) \in \mathrm{R}^n$ contains the n unknown perturbation flow quantities at each point in the computational grid. For example, for two-dimensional, compressible, inviscid flow, which is governed by the Euler equations, the unknowns at each grid point are the perturbations in flow density, Cartesian momentum components and flow energy. The vectors $u(t)$ and $y(t)$ in (1) contain the system inputs and outputs respectively. The definition of inputs and outputs will depend upon the problem at hand. For active control applications, the output might monitor a flow condition at a particular location which varies in response to a disturbance in the incoming flow, while inputs describe both actuation mechanisms and flow disturbances.

The linearization matrices A, B, C, and D in (1) are evaluated at steady-state conditions. In comparison with the nonlinear equations, the system (1) is relatively efficient since a time discretization, such as backward Euler, can be applied and the resulting large $n \times n$ system matrix factored just once for a time-dependent calculation. However, the order of the system is still prohibitively high for many applications, and the cost to solve the system is too large for implementation in real time. We therefore consider the task of finding a low-order, stable, linear time invariant (LTI), state-space model

$$\hat{G}: \quad \frac{d}{dt}\hat{x}(t) = \hat{A}\hat{x}(t) + \hat{B}u(t), \qquad \hat{y}(t) = \hat{C}\hat{x}(t) + \hat{D}u(t) \tag{2}$$

which approximates well the given stable model (1). Typically, A in (1) is a sparse, square matrix of very large dimension $n > 10^4$, and the desired order k of $\hat{G}$ is less than 50.

The quality of $\hat{G}$ as an approximation of G is defined as the H-Infinity norm of the difference between their transfer functions:

$$\|\hat{G} - G\|_\infty = \sup_{\omega \in \mathrm{R}} |\hat{G}(j\omega) - G(j\omega)|, \tag{3}$$

which in turn equals the square root of the maximal energy of the difference $e = \hat{y} - y$, given by

$$\|\hat{y} - y\|_2^2 = \int_t |\hat{y}(t) - y(t)|^2 dt. \tag{4}$$

1.2 Projection Framework

No efficient (polynomial time) solution is known for the problem of minimizing $\|\hat{G} - G\|_\infty$ subject to the order and stability constraints imposed on $\hat{G}$. Optimal Hankel model reduction and balanced truncation are polynomial-time algorithms that produce suboptimal reduced models with strong guarantees of quality. However, the computational burden of these methods is such that their direct application is impractical for $n > 10^3$. Instead, most model reduction techniques for large-scale systems use a projection framework. The full state vector x is represented in a reduced-space basis

$$x = V\hat{x} \tag{5}$$

where the columns of the matrix $V \in \mathrm{R}^{n \times k}$ contain k basis vectors. Defining a left projection space W so that $W^T V = I$, the governing equations (1) can be projected onto the reduced space to yield an m^{th}-order model of the form

$$\hat{G}: \quad \frac{d}{dt}\hat{x}(t) = W^T A V \hat{x}(t) + W^T B u(t), \qquad \hat{y}(t) = CV\hat{x}(t) + \hat{D}u(t). \tag{6}$$

The reduction task is then to find a suitable basis W, V so that $k << n$ and the reduction error, defined by (4), is small. Several methods for computing the basis exist, including Krylov-subspace methods and the POD.

2 Proper orthogonal decomposition

The POD has been widely used to determine efficient bases for dynamic systems [1]. It was introduced for the analysis of turbulence by Lumley [2], and is also known as singular value decomposition, Karhunen-Loéve decomposition [3], and principal component analysis [4]. POD basis vectors are computed empirically using a set of data that samples the range of relevant system dynamics.

2.1 Time-Domain POD Basis

The basis vectors Ψ are chosen so as to maximize the following cost [5]:

$$\Psi = \arg\max \frac{\langle |(x, \Psi)|^2 \rangle}{(\Psi, \Psi)}, \tag{7}$$

where (x, Ψ) denotes the scalar product of the basis vector with the field $x(\theta, t)$ which depends on problem geometry θ and time t, and $\langle\ \rangle$ represents a time-averaging operation. It can be shown that a necessary condition for (7) to hold is that Ψ is an eigenfunction of the kernel K defined by

$$K(\theta, \theta') = \langle x(\theta, t)\ x^*(\theta', t) \rangle, \tag{8}$$

where x^* denotes the complex conjugate transpose of x.

Sirovich introduced the method of snapshots as a way of determining the eigenfunctions Ψ without explicitly calculating the kernel K [6]. The kernel can be approximated as

$$K(\theta, \theta') = \frac{1}{m} \sum_{i=1}^{m} x_i(\theta) x_i^*(\theta') \tag{9}$$

where $x_i(\theta)$ 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

$$\Psi = \sum_{i=1}^{m} \beta_i x_i, \tag{10}$$

where the constants β_i satisfy the eigenvector equation

$$R\beta = \Lambda\beta \tag{11}$$

and R is now the correlation matrix

$$R_{ij} = \frac{1}{m}(x_i, x_j) \tag{12}$$

which contains the inner product between every pair of snapshots. The magnitude of the j^{th} eigenvalue, λ_j, describes the relative importance of the j^{th} POD basis vector for reconstruction of the data contained in the snapshot ensemble. For a basis containing the first p POD modes, the least squares error of data reconstruction is given by the eigenvalues corresponding to the neglected modes:

$$\sum_{i=1}^{m} \|x_i - \tilde{x}_i^p\|_2^2 = \sum_{j=p+1}^{m} \lambda_j \tag{13}$$

where $\tilde{x}_i^p$ is the representation of the i^{th} snapshot, x_i, in the p^{th}-order POD basis.

2.2 Frequency-Domain POD Basis

Often, the POD snapshots are obtained from a time simulation of the CFD model. One issue with this approach is an appropriate choice of input to the simulation. This input choice is critical, since the resulting basis will capture only those dynamics present in the snapshot ensemble. This can be a problematic issue for many applications, such as flow control design, where the dynamics of the controlled and uncontrolled systems might differ significantly. An alternative approach is to apply

the POD in the frequency domain [7]. Rather than selecting a time-dependent input function, one selects a set of sample frequencies. The corresponding flow snapshots can then be obtained by solving the frequency domain CFD equations

$$X(\omega) = [j\omega_i I - A]^{-1} B \tag{14}$$

where $u(t) = e^{j\omega_i t}$, $x(t) = Xe^{j\omega_i t}$, and ω_i is the i^{th} sample frequency.

Frequency domain POD approaches typically yield better results; however, the computational cost of the method is high. The n^{th}-order system given by (14) must be solved for each frequency selected. In a typical CFD application, a large number of frequency points is required to obtain satisfactory models. In particular, for three-dimensional applications, the cost of this approach can be prohibitive. While not discussed in this paper, Krylov-based reduction techniques are computationally more efficient than the POD. These techniques are widely used in integrated circuit applications, and have also been applied to CFD systems. It can be shown that the approach of using the Arnoldi method with multiple interpolation points has strong connections to the frequency-domain POD approach [8].

2.3 POD Reduced-Order Model

Once the set of POD basis vectors has been computed, using time- or frequency-domain snapshots, the reduced-order model is obtained using the projection (5). It is important to note that, while the POD basis is optimal in the sense that it provides the most efficient representation of the data contained in the snapshot ensemble, one can make no statement regarding the quality of the resulting reduced-order model. No estimate or bound for the reduction error is available. Since the POD basis is orthonormal, that is $W = V$, the projection $\hat{A} = V^T AV$ preserves definiteness; however, in general, this is not sufficient to preserve stability. In practice, unstable models are often generated, and some trial and error is required to determine an appropriate snapshot ensemble and basis size.

The POD is a useful reduction technique and has been shown to yield satisfactory results for a wide range of applications (see, for example, [9]); however, its lack of rigorous guarantees should not be overlooked.

2.4 Balanced POD

The concept of a balanced realization of a system was first introduced by Moore [10]. 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. The controllability and observability gramians of the linear system (1) are defined respectively as

$$W_c = \int_0^\infty e^{At} BB^* e^{A^* t} dt \tag{15}$$

and

$$W_o = \int_0^\infty e^{A^* t} C^* C e^{At} dt. \tag{16}$$

The above matrices describe the controllable and observable subspaces of the linear system (1). These two subspaces can be interpreted geometrically as described in [11]. The controllable subspace is that set of states which can be obtained with zero initial state and a given input $u(t)$, while the observable subspace comprises those states which as initial conditions could produce a certain output $y(t)$ with no external input.

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

$$W_{co} = T^{-1} \Lambda T \tag{17}$$

The columns of T then contain the basis vectors which describe the balancing transformation. The eigenvalues λ_i contained in the diagonal matrix Λ are positive, real numbers, and $\sigma_i = \sqrt{\lambda_i}$ are known as the Hankel singular values of the system. These values are independent of the particular realization of the system and describe the importance of the corresponding state for transmission of input to output. In a balanced truncation, only those states are retained which correspond to large Hankel singular values. A bound on the error of the k^{th}-order balanced reduced model, $\hat{G}_k$, is given by the Hankel singular values of the truncated modes

$$||G - \hat{G}_k||_\infty \leq 2 \sum_{i=k+1}^{m} \sigma_i \tag{18}$$

Approximate Balanced Realization via the Method of Snapshots

For large systems, it is not practical to explicitly compute the gramians using (15) and (16); however, it can be shown that there is a strong connection between the controllability gramian and the POD kernel function [12, 13]. By noting that for a SISO system the quantity $x_\delta(t) = e^{At}B$ is simply the impulse response of the system (set $u(t) = \delta(t)$ in (1)), the controllability gramian can also be written

$$W_c = \int_0^\infty x_\delta(t) x_\delta^*(t) dt \tag{19}$$

and compared with the POD kernel function K defined by (9). Similarly, the observability gramian can be written as

$$W_o = \int_0^\infty \overline{x}_\delta(t) \overline{x}_\delta^*(t) dt \tag{20}$$

where $\overline{x}_\delta(t) = e^{A^* t} C^*$ is the impulse response of the dual SISO system which is given by

$$\overline{G}: \quad \frac{d}{dt}\overline{x}(t) = A^*\overline{x}(t) + C^*\overline{u}(t), \qquad \overline{y}(t) = B^*\overline{x}(t) \tag{21}$$

By considering general, non-impulsive inputs, the POD computes the most controllable modes of the system within a certain restricted range of dynamics. It is a natural extension to consider a POD analysis which determines the most observable modes. Furthermore, to obtain a balanced representation of the system, concepts from a traditional control balanced realization can then be used. The direct POD method can be used to obtain approximations to the system gramians for small systems [12]. For large-scale systems, the POD method of snapshots can be used to approximate the gramians in a computationally efficient manner that does not require computation of large $n \times n$ matrices [13].

By obtaining snapshots of the primal and dual systems (1) and (21), and performing the POD method of snapshots analysis to determine the POD basis vectors, p eigenmodes of the controllability and observability kernel functions are obtained, respectively. Let the eigenvectors of the controllability kernel K be contained in the columns of the matrix V, with corresponding eigenvalues on the diagonal entries of the matrix Λ_c. Similarly, let the eigenvectors of the observability kernel be contained in the columns of the matrix X, with corresponding eigenvalues on the diagonal entries of the matrix Λ_o. Low-rank approximations to the controllability and observability gramians can then be made as follows:

$$W_c^p = V \Lambda_c V^* \tag{22}$$
$$W_o^p = X \Lambda_o X^*, \tag{23}$$

where the superscript p denotes a p^{th} order approximation.

Through use of an efficient eigenvalue solver, the eigenmodes of the product $W_c^p W_o^p$ can then be calculated using 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 (V, Λ_c) for the primal system.

2. Use method of snapshots to obtain p POD eigenmodes (X, Λ_o) for the dual system.

3. Calculate the low-rank approximations $W_c^p = V \Lambda_c V^*$ and $W_o^p = X \Lambda_o X^*$.

4. Obtain the eigenvectors of the product $W_c^p W_o^p$ to determine the balancing transformation T.

Once again, it is important to note the limitations of this reduction approach. Balanced truncation is a rigorous method for small systems that yields reduced-order models with strong guarantees of quality and a computable error bound given by (18). The approximate method using the POD draws upon a strong analogy with balanced truncation and has been shown to work effectively for many cases; however, as for the conventional POD, this method does not offer rigorous error or stability guarantees.

In the next section, a different approach to model reduction of large-scale systems that does not use a projection framework will be described. This method has associated with it a rigorous, although non-computable, error bound.

3 Fourier Model Reduction

The FMR method is described in [14] and uses discrete-time Fourier coefficients of the CFD system to form a reduced-order model with a rigorous error bound. Many coefficients can be calculated, which results in a very accurate representation of the system dynamics, but only a single factorization of the large system is required. As shown in Figure 1, FMR can be combined with an efficient second reduction step using explicit formulae for balanced truncation, which allows use of the Hankel singular values to rigorously select the number of reduced states. FMR yields very accurate, low-order models when the original transfer function is smooth, and thus is an attractive method for reduction of CFD systems.

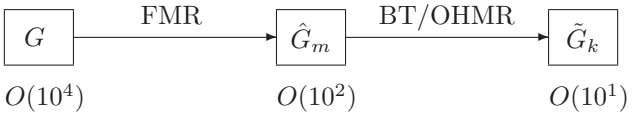

Figure 1. *Two-step model reduction process: FMR with rigorous but non-computable error bound followed by balanced truncation (BT) or optimal Hankel model reduction (OHMR).*

3.1 Fourier Series of Discrete Time Systems

Consider the discrete-time (DT) model g corresponding to the system (1), which is defined by the difference equations

$$g: \quad x(t+1) = ax(t) + bu(t), \quad y(t) = cx(t) + du(t) \tag{24}$$

where a, b, c, d are the DT matrices. The transfer function

$$g(z) = d + c(zI - a)^{-1}b \tag{25}$$

has the Fourier decomposition

$$g(z) = \sum_{k=0}^{\infty} g_k z^{-k} \tag{26}$$

where

$$g_0 = d, \qquad g_k = ca^{k-1}b \ (k = 1, 2, \ldots) \tag{27}$$

The Fourier expansion converges exponentially for $|z| > \rho(a)$, where $\rho(a)$ denotes the spectral radius of a, defined as the maximal absolute value of its eigenvalues. Note that the first m Fourier coefficients g_k are easy to calculate using the "cheap" iterative process

$$g_k = ch_{k-1}, \ h_k = ah_{k-1} \ (k = 1, \ldots, m), \quad \text{where } h_0 = b. \tag{28}$$

Let $\hat{g}_m$ denote the m^{th}-order approximation of g based on the Fourier series expansion:

$$\hat{g}_m(z) = \sum_{k=0}^{m} g_k z^{-k} \tag{29}$$

The following simple result provides an estimate of the approximation error

$$\|g - \hat{g}_m\|_\infty = \max_{|z|=1} |g(z) - \hat{g}_m(z)| \tag{30}$$

which ties it to the smoothness of G as follows.

Theorem 1. *For $q = 1, 2, \ldots$*

$$\|g - \hat{g}_m\|_\infty^2 \leq \frac{m^{1-2q}}{2\pi(2q-1)} \int_{-\pi}^{\pi} |g^{(q)}(e^{j\tau})|^2 d\tau, \tag{31}$$

where $g^{(q)}$ is the q^{th} derivative of g with respect to τ.
Proof: see [14].

3.2 Fourier Series of Continuous Time Systems

Consider the full continuous time LTI system model G defined by the system (1), where $u(t), y(t)$ are scalar input and output. It will be assumed that G is stable, i.e. that all roots of the characteristic equation $\det(sI - A) = 0$ have negative real part, and that $C(sI - A)^{-1}B$ remains bounded as $s \to \infty$.

Let $\omega_0 > 0$ be a fixed positive real number. The transfer function

$$G(s) = D + C(sI - A)^{-1}B \tag{32}$$

has the Fourier decomposition

$$G(s) = \sum_{k=0}^{\infty} G_k \left(\frac{s - \omega_0}{s + \omega_0} \right)^k \tag{33}$$

This decomposition is suggested in [15] and an approximate FFT algorithm is used to calculate the Fourier coefficients G_k. In [14], an efficient iterative procedure is proposed to directly calculate the Fourier coefficients as follows.

Consider the identity

$$G(s) = g(z) = d + c(zI - a)^{-1}b \quad \text{for} \quad z = \frac{s + \omega_0}{s - \omega_0}$$

which allows one to apply the observations and theorem from the previous subsection to this case. Note that by comparing (26) and (33), it can be seen that $G_k = g_k$. The Fourier coefficients are therefore given by the following formulae

$$G_0 = d, \quad G_k = ca^{k-1}b \ (k = 1, 2, \ldots) \tag{34}$$

$$d = D + C(\omega_0 I - A)^{-1}B \tag{35}$$

$$a = (\omega_0 I + A)(\omega_0 I - A)^{-1} \tag{36}$$

$$c = 2\omega_0 C(\omega_0 I - A)^{-1} \tag{37}$$

$$b = -(\omega_0 I - A)^{-1}B \tag{38}$$

which are relatively straightforward to obtain using algebraic manipulations as described in [14].

3.3 Reduced Model Construction

To construct an m^{th}-order reduced model, one first calculates the Fourier coefficients, $g_0, g_1, ..., g_m$. The DT reduced model is then given by

$$\begin{aligned} \hat{g}: \quad \hat{x}[t+1] &= \hat{a}\hat{x}[t] + \hat{b}u[t] \\ \hat{y}[t] &= \hat{c}\hat{x}[t] + \hat{d}u[t], \end{aligned} \tag{39}$$

where

$$\hat{a} = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots \\ 1 & 0 & 0 & 0 & \dots \\ 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & \ddots & \end{bmatrix} \quad \hat{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$
$$\hat{c} = \begin{bmatrix} g_1 & g_2 & \dots & g_m \end{bmatrix} \qquad \hat{d} = g_0. \tag{40}$$

It should be noted that, if the original system g is stable, then the reduced system $\hat{g}$ is also guaranteed to be stable. This is due to the orthogonality properties of the Fourier expansion; taking a truncated number of Fourier terms, as in equation (29), automatically results in a stable approximation of a stable system.

Different alternatives could be chosen for the DT system representation. The controller canonical form above was selected, as in [15], for the purpose of an efficient second step of reduction. As shown in Figure 1, an effective approach is to use the efficient iterative procedure to calculate several hundred coefficients, resulting in an intermediate reduced model of the form (39). A second reduction step using balanced truncation can now be performed easily, since the expressions for the gramians are known explicitly. For the DT reduced model (39), the controllability matrix is the identity matrix and the observability matrix is the Hankel matrix that has $\hat{c}$ as its first row. The balancing vectors can therefore be obtained by computing the singular vectors of the m^{th}-order Hankel matrix

$$\Gamma = \begin{bmatrix} g_1 & g_2 & g_3 & \dots & g_{m-1} & g_m \\ g_2 & g_3 & g_4 & \dots & g_m & 0 \\ g_3 & g_4 & g_5 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ g_{m-1} & g_m & 0 & \dots & 0 & 0 \\ g_m & 0 & 0 & \dots & 0 & 0 \end{bmatrix}. \tag{41}$$

The Hankel singular values, $\sigma_i,\ i = 1, 2, ..., m$, of the intermediate reduced system are given by the singular values of Γ and can be used to quantitatively guide the second reduction step via the error bound given in (18).

3.4 FMR Algorithm

The FMR algorithm is summarized in the following steps.

1. Choose a value of ω_0. The value of ω_0 should reflect the frequency range of interest. The nominal value is unity; however, if the response at high frequencies is of interest, a higher value of ω_0 should be chosen. One can visualize the transformation from continuous to discrete time as a mapping of the imaginary axis in the $s-$plane to the unit circle in the $z-$plane. The value of ω_0 then describes the compression of frequencies around the unit circle.

2. Calculate $m+1$ Fourier coefficients using (34)-(38). Using the iterative procedure, any number of coefficients can be calculated with a single n^{th}-order matrix factorization.

3. Using (41), calculate the m^{th}-order Hankel matrix. Calculate its singular values and singular vectors. Note that in the general case of p inputs and q outputs, each entry g_k will be a block of size $q \times p$.

4. Using balanced truncation, construct a k^{th}-order DT system, $\tilde{g}$. The value of k is chosen according to the distribution of Hankel singular values of the intermediate system.

5. Convert the k^{th}-order, DT reduced model to a continuous-time model using the relationships

$$\hat{A} = \omega_0 \left(\hat{a} - I\right)^{-1} \left(\hat{a} + I\right), \tag{42}$$

$$\hat{B} = 2\omega_0 \left(\hat{a} - I\right)^{-1} \hat{b}, \tag{43}$$

$$\hat{C} = -\hat{c} \left(\hat{a} - I\right)^{-1}, \tag{44}$$

$$\hat{D} = \hat{d} - \hat{c} \left(\hat{a} - I\right)^{-1} \hat{b} \tag{45}$$

The error bounds corresponding to the above algorithm are given in (31) and (18) for the first and second stages of the reduction, respectively. The Hankel singular values of the intermediate system provide a straightforward, quantitative means to choose k, the size of the final reduced model; however, the error bound given in (31) is not readily computable. One way to determine the appropriate number of Fourier coefficients, m, is to monitor the magnitudes of the coefficients, g_k, which tend to decrease quickly with increasing index k. A tolerance on the coefficient magnitude can thus be used to set a stopping criterion for the iterative procedure. Note that the exact distribution of coefficients will depend upon the transfer function under consideration. For near resonant systems, the magnitudes of the Fourier coefficients will not decrease quickly and many coefficients will be needed to gain an accurate representation. For such systems, other algorithms, such as the POD or Krylov-subspace methods, may yield better results.

4 Active Flow Control of a Supersonic Diffuser

Figure 2 shows the contours of Mach number at steady-state conditions inside the fixed geometry of a supersonic diffuser that operates at a freestream Mach number of 2.2. In steady-state operation, a shock forms downstream of the throat. In practice, the incoming supersonic flow is subject to perturbations, such as atmospheric density disturbances. Such perturbations in the flow may cause the shock to move upstream of the throat, and eventually to be expelled from the diffuser. This phenomenon, known as inlet unstart, causes huge losses in engine performance and thus is highly undesirable. In order to prevent inlet unstart for this diffuser, an active control mechanism of the shock is required.

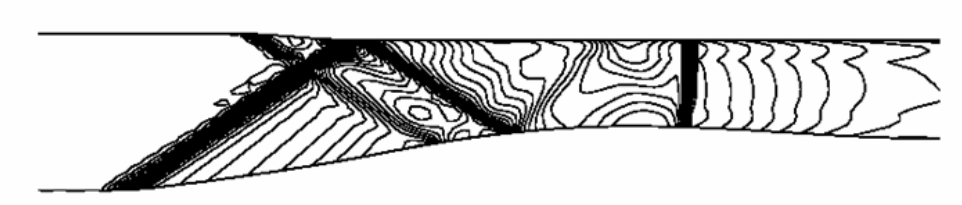

Figure 2. *Contours of Mach number for steady flow through supersonic diffuser. Steady-state inflow Mach number is 2.2.*

Figure 3 presents a schematic of the actuation mechanism. Incoming flow with possible disturbances enters the inlet and is sensed using pressure sensors. The controller then adjusts the bleed upstream of the throat in order to control the position of the shock and to prevent it from moving upstream. In simulations, it is difficult to automatically determine the shock location. The average Mach number at the diffuser throat provides an appropriate surrogate that can be easily computed. A CFD model provides an accurate representation of the complicated flow dynamics, but is not suitable for controller design purposes.

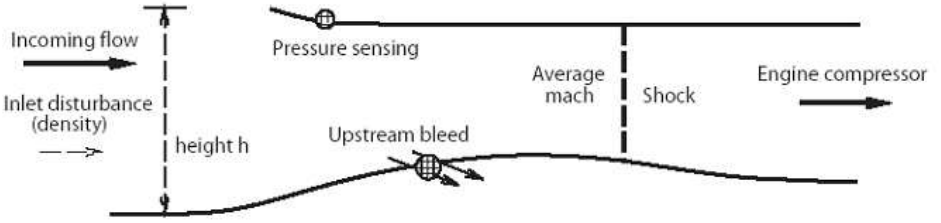

Figure 3. *Supersonic diffuser active flow control problem setup.*

The governing equations considered are the two-dimensional unsteady Euler equations, linearized about steady-state conditions. The CFD model considered

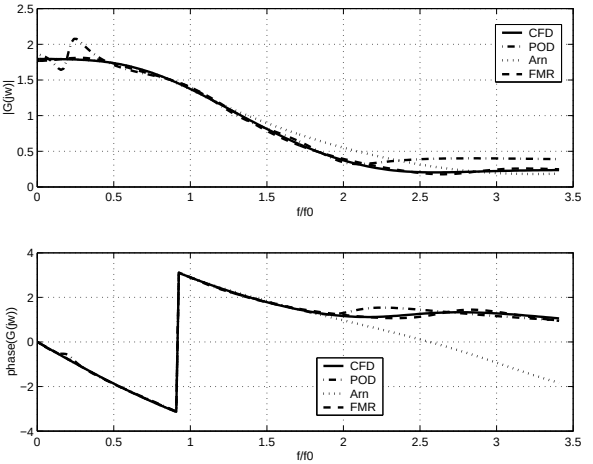

Figure 4. *Transfer function from bleed actuation to average throat Mach number for supersonic diffuser. Results from CFD model ($n = 11,730$) are compared to FMR, POD and Arnoldi models with $k = 10$ states.*

here has 3078 grid points and 11,730 unknowns and is described in [16]. The first transfer function of interest is that between bleed actuation and average Mach number at the throat. Bleed occurs through small slots located on the lower wall between 46% and 49% of the inlet overall length. Frequencies of practical interest lie in the range $f/f_0 = 0$ to $f/f_0 = 2$, where $f_0 = a_0/h$, a_0 is the freestream speed of sound and h is the height of the diffuser.

Figure 4 shows the magnitude and phase of this transfer function as calculated by the CFD model and three reduced-order models each of size $k = 10$. The FMR model was calculated by using 201 Fourier coefficients (calculated at the cost of a single CFD matrix inversion) with $\omega_0 = 5$ to construct the Hankel matrix in (41). This 200^{th}-order system was then further reduced to ten states using explicit balanced truncation. The POD model was obtained by computing 41 snapshots at 21 equally-spaced frequencies from $f/f_0 = 0$ to $f/f_0 = 2$. For comparison, a tenth-order Arnoldi-based model, derived using the methodology in [16], is also shown in Figure 4.

It can be seen from Figure 4 that the FMR model matches the CFD results well over the entire frequency range plotted, with a small discrepancy at higher frequencies. The Arnoldi model matches well for low frequencies, but shows considerable error for $f/f_0 > 1.3$. The POD model has some undesirable oscillations at low frequencies, and strictly is only valid over the frequency range sampled in the snapshot ensemble ($f/f_0 < 2$).

The performance of the POD and Arnoldi models can be improved by increasing the size of the reduced-order models. Figure 5 shows the results using 30 Arnoldi vectors and 15 POD basis vectors. The agreement at low frequencies is now

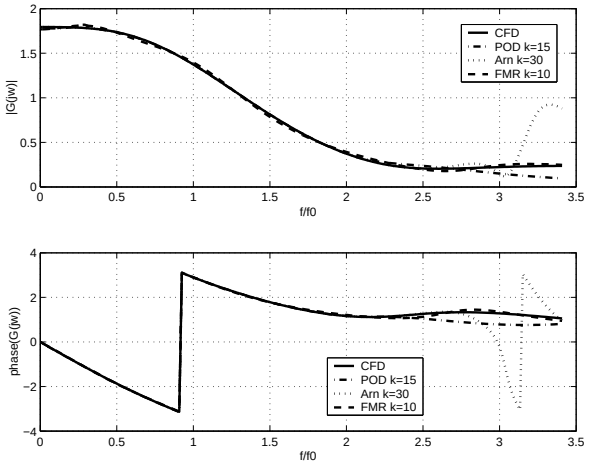

Figure 5. *Transfer function from bleed actuation to average throat Mach number for supersonic diffuser. Results from CFD model ($n = 11,730$) are compared to FMR with $k = 10$ states, POD with $k = 15$ states, and Arnoldi with $k = 30$ states.*

very good for all models, but the POD and Arnoldi models still show discrepancy at higher frequencies. The POD model could be further improved by including more snapshots in the ensemble; however, each additional frequency considered requires an n^{th}-order complex matrix inversion. The Arnoldi model could be further improved by increasing the size of the basis; however, this was found to result in unstable reduced-order models. This result highlights one of the major problems with the commonly used POD and Krylov-based reduction techniques. Because no rigorous statement about model quality can be made, the reduction becomes an ad hoc process that requires trial and error to obtain accurate, stable reduced-order models. In particular, for the POD, the choice of snapshot ensemble is critical.

FMR is also applied to the transfer function between an incoming density perturbation and the average Mach number at the diffuser throat. This transfer function represents the dynamics of the disturbance to be controlled and is shown in Figure 6. As the figure shows, the dynamics contain a delay, and are thus more difficult for the reduced-order model to approximate. Results are shown for FMR with $m = 200$ and $\omega_0 = 5, 10$. With $\omega_0 = 5$, the model has significant error for frequencies above $f/f_0 = 2$. Choosing a higher value of ω_0 improves the fit, although some discrepancy remains. These higher frequencies are unlikely to occur in typical atmospheric disturbances, however if they are thought to be important, the model could be further improved by either evaluating more Fourier coefficients, or by choosing a higher value of ω_0. The $\omega_0 = 10$ model is further reduced via balanced truncation to a system with thirty states without a noticeable loss in accuracy.

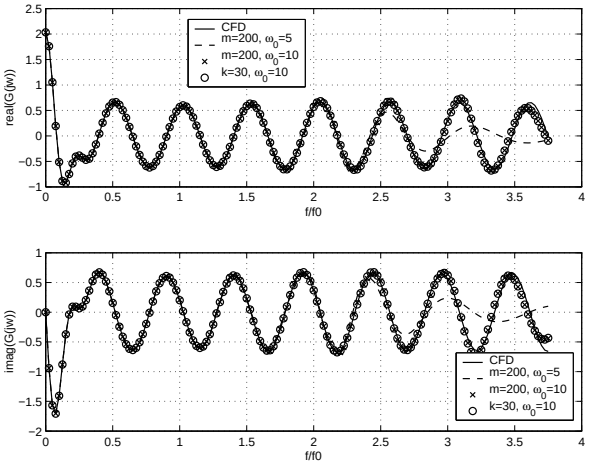

Figure 6. *Transfer function from incoming density perturbation to average throat Mach number for supersonic diffuser. Results from CFD model ($n = 11,730$) are compared to 200^{th}-order FMR models with $\omega_0 = 5, 10$. The $\omega_0 = 10$ model is further reduced to $k = 30$ via balanced truncation.*

5 Conclusion

Model reduction is an essential component to achieving real-time simulation and control of PDEs. Several model reduction techniques are available and in use across a broad range of applications. For large-scale systems, the POD and Krylov-based methods have been used with considerable success. The three orders of magnitude reduction from $O(10^4)$ states to $O(10^1)$ states demonstrated in the supersonic diffuser example is representative of what can be achieved in many applications. However, it is important to note that many open questions and unresolved issues remain. The popular reduction approaches for large-scale systems do not offer rigorous guarantees regarding the quality of the reduced-order model, and remain ultimately ad hoc. FMR goes some way to addressing this issue by providing a rigorous, although non-computable, error bound. Using this error bound to combine FMR with a more rigorous technique, such as balanced truncation, yields an efficient, systematic, two-step reduction process.

In addition, reduction of large-scale nonlinear systems remains an open question. Direct projection of the nonlinear governing equations onto a reduced subspace yields a model that has low order but that cannot be implemented efficiently. A trajectory piecewise-linear approach that can be used efficiently in conjunction with large-scale reduction methods has been proposed and demonstrated for integrated circuit applications [17]. This approach has also been shown to combine effectively with the POD for nonlinear CFD applications [18]. Many open questions remain regarding the robustness and rigor of this approach; however, it represents a significant step towards achieving efficient, low-order, nonlinear models.

Bibliography

[1] Holmes, P., Lumley, J.L. and Berkooz, G., *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, Cambridge University Press, Cambridge, UK (1996).

[2] Lumley, J.L., *The Structures of Inhomogeneous Turbulent Flow*, in Atmospheric Turbulence and Radio Wave Propagation (1967) pp. 166-178.

[3] Loéve, M., *Probability Theory*, D. Van Nostrand Company Inc., New York, NY (1955).

[4] Hotelling, H., *Analysis of a Complex of Statistical Variables with Principal Components*, in Journal of Educational Psychology, 24 (1933), pp. 417-441,498-520.

[5] Berkooz, G., Holmes, P. and Lumley, J.L., *The Proper Orthogonal Decomposition in the Analysis of Turbulent Flows*, in Annual Review of Fluid Mechanics, 25 (1993), pp. 539-575.

[6] Sirovich, L., *Turbulence and the Dynamics of Coherent Structures. Part 1 : Coherent Structures*, in Quarterly of Applied Mathematics, 45 (1987), pp. 561-571.

[7] Kim,T., *Frequency-Domain Karhunen-Loeve Method and Its Application to Linear Dynamic Systems*, in AIAA Journal, 36 (1998), pp. 2117-2123.

[8] Willcox, K., Peraire, J. and White, J., *An Arnoldi approach for generation of reduced-order models for turbomachinery*, in Computers and Fluids, 31 (2002), pp. 369-89.

[9] Dowell, E.H. and Hall, K.C., *Modeling of fluid-structure interaction*, in Annual Review of Fluid Mechanics, 33 (2001), pp. 445-90.

[10] Moore, B.C., *Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction*, in IEEE Transactions on Automatic Control, AC-26 (1981), pp. 17-31.

[11] Enns, D.F., *Model Reduction for Control System Design*, PhD Thesis, Dept. of Aeronautics and Astronautics, Stanford University (1984).

[12] Lall, S., Marsden, J.E. and Glavaski, S., *A subspace approach to balanced truncation for model reduction of nonlinear control systems*, in International Journal on Robust and Nonlinear Control, 12 (2002), pp. 519-535.

[13] Willcox, K. and Peraire, J., *Balanced Model Reduction via the Proper Orthogonal Decomposition*, in AIAA Journal, 40 (2002), pp. 2323-2330.

[14] Willcox, K. and Megretski, A., *Fourier Series for Accurate, Stable, Reduced-Order Models for Linear CFD Applications*, AIAA Paper 2003-4235, to appear in SIAM J. Scientific Computing (2004).

[15] Gu, G., Khargonekar, P.P. and Lee, E.B., *Approximation of Infinite-Dimensional Systems*, in IEEE Transactions on Automatic Control, 34 (1989), pp. 610-18.

[16] Lassaux, G., *High-Fidelity Reduced-Order Aerodynamic Models: Application to Active Control of Engine Inlets.*, SM Thesis, Dept. of Aeronautics and Astronautics, MIT (2002).

[17] Rewienski, M., *A Trajectory Piecewise-Linear Approach to Model Order Reduction of Nonlinear Dynamical Systems*, PhD Thesis, Dept. of Electrical Engineering and Computer Science, MIT (2003).

[18] Gratton, D. and Willcox, K., *Reduced-Order, Trajectory Piecewise-Linear Models for Nonlinear Computational Fluid Dynamics*, AIAA Paper 2004-2329 (2004).