MODEL REDUCTION FOR LARGE-SCALE LINEAR APPLICATIONS

Karen Willcox [*] **Alexandre Megretski** [**]

[*] *Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139*
[**] *Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139*

Abstract: Three model reduction methods are considered in the context of large scale fluid dynamic applications: the widely-used proper orthogonal decomposition, the Arnoldi method and a new Fourier method. The new method uses a Fourier expansion of the transfer function in discrete frequency to efficiently calculate reduced models with guaranteed stability and accuracy properties. Each method is described and then applied to the case of flow through a supersonic diffuser. The Fourier model reduction approach is found to be superior in all aspects; it is computationally more economical, preserves the stability of the original system, uses both input and output information to yield efficient models, and is valid over a wide range of frequencies.

Keywords: Large-Scale, Model Reduction, Aerospace Systems, Computational Fluid Dynamics

1. INTRODUCTION

In the past decade, model reduction has become popular throughout the fluid dynamics community. While computational fluid dynamic (CFD) methods produce accurate models for problems of interest, their size and computational expense render them unsuitable for many applications. In particular, model reduction has been widely used for aeroelastic applications where the flow model must be coupled to a structural model (Dowell and Hall, 2001). While several different reduction techniques have been considered for CFD applications, including eigenmodes (Hall, 1994) and the Arnoldi method (Willcox *et al.*, 2002), the proper orthogonal decomposition (POD) is by far the most widely used in the fluid dynamics community (Holmes *et al.*, 1996).

First introduced in the context of turbulence (Lumley, 1967), the POD method of snapshots has been developed as a way to apply the technique to large-scale systems (Sirovich, 1987). This approach requires a set of flow solutions, or ”snapshots” from a CFD simulation. These snapshots are then used to create a reduced-space basis, onto which the CFD governing equations are projected. While the POD basis is optimal in the sense that it minimizes the error between the snapshots and their projection in the reduced space, there are no guarantees as to the quality of the reduced-order model as an approximation of the original CFD system. In particular, one can not even guarantee that POD-based reduction of a stable CFD model will result in a stable reduced-order model.

Many effective, more rigorous reduction techniques have been developed in a controls context. The quality of a reduced-order system $\hat{G}$ as an approximation of the original system G is defined as the H-Infinity norm of the difference between their transfer functions:

$$\|\hat{G} - G\|_\infty = \sup_{\omega \in \mathbf{R}} |\hat{G}(j\omega) - G(j\omega)| \qquad (1)$$

No polynomial-time algorithm is known for determining the optimal reduced-order model, that is, one which minimizes the above norm. Algorithms such as Hankel model reduction (Adamjan *et al.*, 1971; Bettayeb *et al.*, 1980; Kung and Lin, 1981) and balanced truncation (Moore, 1981) have been widely used throughout the controls community to generate suboptimal reduced models with strong guarantees of quality. These algorithms can be performed in polynomial time; however, the computational requirements make them impractical for application to large systems such as those encountered in CFD applications. Several methods have been developed for computing approximations to the grammians for large systems, including the approximate subspace iteration (Baker *et al.*, 1996), least squares approximation (Scottedward Hodel, 1991) and Krylov subspace methods (Jaimoukha and Kasenally, 1994; Gudmundsson and Laub, 1994; Sorensen and Antoulas, 2002); however, these algorithms are complicated and computationally intensive.

In Willcox and Megretski (2003), a new technique is described for model reduction of large-scale systems. Fourier model reduction (FMR) uses an efficient iterative procedure to calculate Fourier coefficients of the transfer function in the discrete frequency domain. The resulting reduced-order models are guaranteed to be stable and the approximation error satisfies a known bound, which depends on the smoothness of the original transfer function.

In this paper, FMR is compared to the POD and Arnoldi methods for a CFD application of flow through a supersonic diffuser. In this example, reduced-order models are required in order to derive active control strategies. In the following section, each of the three algorithms, POD, Arnoldi, and FMR, are briefly outlined. Results are then presented for the chosen example and finally, conclusions are drawn.

2. MODEL REDUCTION FOR CFD

We consider the task of finding a low-order, state-space model

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

which approximates well the given stable model

$$\begin{aligned} G: \quad E\frac{d}{dt}x(t) &= Ax(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t). \end{aligned} \tag{3}$$

For the case of CFD applications, $x(t) \in \mathbf{R}^n$ is the state vector containing the n unknown perturbation flow quantities at each point in the computational grid, while $\hat{x}(t) \in \mathbf{R}^k$ is the k^{th}-order reduced state vector. The vectors $u(t)$ and $y(t)$ contain the system inputs and outputs respectively. For simplicity, single-input, single-output systems will be considered here; however, all algorithms extend to the multiple-input, multiple-output case. The matrices E, A, B, C and D in (3) arise from the CFD formulation, and are evaluated at steady-state flow conditions. Typically, A and E are sparse matrices of very large dimension ($n > 10^4$). The descriptor matrix E is included for generality, and may contain some zero rows, which arise from implementation of flow boundary conditions. On solid walls, a condition is imposed on the flow velocity, while at farfield boundaries certain flow parameters are specified, depending on the nature of the boundary (inflow/outflow) and the local flow conditions (subsonic/supersonic). Although these prescribed quantities could be condensed out of (3) to obtain a smaller state-space system, such a manipulation is often complicated and can destroy the sparsity of the system. The more general form of the system is therefore considered.

2.1 *Proper Orthogonal Decomposition*

The POD has been widely used as a method of performing model reduction for large, CFD systems (Holmes *et al.*, 1996). The method of snapshots (Sirovich, 1987) is a way to construct a reduced-space basis using flow solutions or ”snapshots”. If x^i is the flow solution at a time t_i, then the POD basis vectors can be constructed as follows.

The correlation matrix R is first formed by computing the inner product between every pair of snapshots

$$R_{ik} = \frac{1}{m}\left(x^i, x^k\right), \tag{4}$$

where m is the number of snapshots and $\left(x^i, x^k\right)$ denotes the inner product between x^i and x^k. The eigenvalues λ_i and eigenvectors ψ^i of R are then computed. The j^{th} POD basis vector, Φ_j, is given by a linear combination of snapshots

$$\Phi_j = \sum_{i=1}^{m} \psi_i^j x^i, \tag{5}$$

where ψ_i^j denotes the i^{th} element of the j^{th} eigenvector. The magnitude of the j^{th} eigenvalue, λ_j, describes the relative importance of the j^{th} POD basis vector.

Once the orthonormal set of POD basis vectors has been computed, the reduced-order model is obtained by projecting the CFD solution onto the reduced-space basis:

$$x(t) = \sum_{i=1}^{m} \hat{x}_i(t)\Phi_i. \tag{6}$$

Substituting this expression into the original system (3) and using orthogonality, we obtain the reduced-order system (2).

Often, the POD snapshots are obtained from a 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 (Kim, 1998). Rather than selecting a time-dependent input function, one selects a set of sample frequencies. The corresponding flow solutions can then be obtained by solving the frequency domain CFD equations

$$X(\omega) = \left[j\omega_i E - A\right]^{-1} B \tag{7}$$

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 (7) must be solved for each frequency selected. In a typical CFD application, a large number of frequency points are required to obtain satisfactory models. In particular, for three-dimensional applications, the cost of this approach is often prohibitive.

2.2 *Arnoldi Method*

The Arnoldi method is one of a set of moment-matching model reduction techniques. Consider a Taylor series expansion of the transfer function

$$G(s) = C\left[sE - A\right]^{-1} B + D \tag{8}$$

about the point $s = s_0$. We can write

$$G(s) = \sum_{j=0}^{\infty} m_j (s - s_0)^j, \tag{9}$$

where

$$m_0 = D + C\tilde{A}^{-1}B, \tag{10}$$

$$m_j = C\left(-\tilde{A}^{-1}E\right)^j \tilde{A}^{-1}B \;(j = 1, 2, \ldots) \tag{11}$$

and $\tilde{A} = (s_0E - A)$. The coefficient m_j is known as the j^{th} moment of $G(s)$ about $s = s_0$.

The Arnoldi method uses an efficient iterative process to generate a set of vectors that spans the k^{th}-order Krylov subspace defined by

$$\begin{aligned}\mathcal{K}_k(\tilde{A}^{-1}E, \tilde{A}^{-1}B) = span\{&\tilde{A}^{-1}B, (\tilde{A}^{-1}E)\tilde{A}^{-1}B,\\ &..., (\tilde{A}^{-1}E)^{k-1}\tilde{A}^{-1}B\},\end{aligned} \tag{12}$$

The reduced-order model is obtained by projecting onto the k^{th}-order Krylov subspace and matches k moments of the system transfer function about s_0. The flow solution $x(t)$ is represented by an expansion of the form (6), where Φ_i now represents the i^{th} Arnoldi vector.

2.3 *Fourier Model Reduction*

A new method of model reduction for large-scale applications is described in Willcox and Megretski (2003). A transfer function expansion concept similar to that described for the Arnoldi method is used; however, the expansion is now performed in the discrete frequency domain. Using the identity

$$G(s) = g(z) = d + c(zI - a)^{-1}b \quad \text{for} \quad z = \frac{s + \omega_0}{s - \omega_0}, \tag{13}$$

where

$$d = D + C(\omega_0 E - A)^{-1}B, \tag{14}$$

$$a = -(\omega_0 E + A)(\omega_0 E - A)^{-1}, \tag{15}$$

$$c = 2\omega_0 C(\omega_0 E - A)^{-1}, \tag{16}$$

$$b = -E(\omega_0 E - A)^{-1}B, \tag{17}$$

and ω_0 is some fixed positive real number, the transfer function $G(s)$ has the Fourier decomposition

$$G(s) = \sum_{j=0}^{\infty} G_j \left(\frac{s - \omega_0}{s + \omega_0}\right)^j, \tag{18}$$

where

$$G_0 = d, \quad G_j = ca^{j-1}b \;(j = 1, 2, \ldots). \tag{19}$$

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. The first m Fourier coefficients are easy to calculate using the efficient iterative process

$$G_j = ch_{j-1},\; h_j = ah_{j-1} \;(j = 1, \ldots, m), \quad \text{where } h_0 = b, \tag{20}$$

which is expected to be “stable” since g is stable, i.e. $\rho(a) < 1$.

The calculated $m + 1$ Fourier coefficients are used to construct an m^{th}-order discrete time reduced model, which can then be converted to an m^{th}-order state-space system. An effective approach is to use the above iterative procedure to form the intermediate discrete time system

$$\begin{aligned}\hat{g}: \quad \hat{x}[t+1] &= \hat{a}\hat{x}[t] + \hat{b}u[t],\\ \hat{y}[t] &= \hat{c}\hat{x}[t] + \hat{d}u[t],\end{aligned} \tag{21}$$

where

$$\hat{a} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & \ddots & \end{bmatrix} \quad \hat{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

$$\hat{c} = \begin{bmatrix} G_1 & G_2 & \ldots & G_m \end{bmatrix} \qquad \hat{d} = G_0 \tag{22}$$

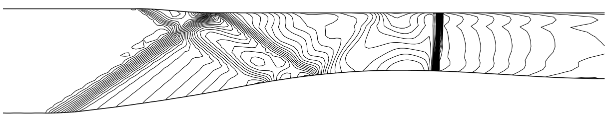

Fig. 1. Mach contours for steady flow through supersonic diffuser. Steady-state inflow Mach number is 2.2.

using several hundred states. This intermediate model can then be further reduced via balanced truncation. This second reduction step can be applied efficiently, since the Hankel matrix of the system (21) is known to be

$$\Gamma = \begin{bmatrix} G_1 & G_2 & G_3 & \dots & G_{m-1} & G_m \\ G_2 & G_3 & G_4 & \dots & G_m & 0 \\ G_3 & G_4 & G_5 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ G_m & 0 & 0 & \dots & 0 & 0 \end{bmatrix}. \quad (23)$$

3. RESULTS

3.1 *Supersonic Diffuser*

Results will be presented for a CFD application of active control of a supersonic diffuser as shown in Figure 1. The diffuser operates at a nominal Mach number of 2.2; however, it is subject to perturbations in the incoming flow, which may be due (for example) to atmospheric variations. In nominal operation, there is a strong shock downstream of the diffuser throat, as can be seen from the Mach contours plotted in Figure 1. Incoming disturbances can cause the shock to move forward towards the throat. When the shock sits at the throat, the inlet is unstable, since any disturbance that moves the shock slightly upstream will cause it to move forward rapidly, leading to unstart of the inlet. This is extremely undesirable, since unstart results in a large loss of thrust. In order to prevent unstart from occurring, one option is to actively control the position of the shock. This control may be effected through flow bleeding upstream of the diffuser throat. In order to derive effective active control strategies, it is imperative to have low-order models which accurately capture the relevant dynamics.

The CFD formulation for this problem is described fully in Lassaux (2002). The governing equations considered are the two-dimensional Euler equations, which are linearized for unsteady flows. The CFD model considered here has 3078 grid points and 11,730 unknowns.

We consider the transfer function 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 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; a wider range will be plotted to gain further insight to the performance of the models.

Figure 2 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 (23). This 200^{th}-order system was then further reduced to ten states using explicit balanced truncation. One might consider a similar approach using the Arnoldi method: first calculate 200 Arnoldi vectors (at the cost of a single CFD matrix inversion), and then further reduce the resulting system with balanced truncation. Although this two-step Arnoldi approach might be effective for some applications, for this case it could not be applied, since the model resulting from projection onto the 200^{th}-order Arnoldi basis was unstable. The Arnoldi model was therefore constructed directly by projection onto the reduced-space basis spanned by the first ten Arnoldi vectors computed about $s_0 = 0$. Finally, the POD model was obtained by computing 41 snapshots at 21 equally-spaced frequencies from $f/f_0 = 0$ to $f/f_0 = 2$. This required the inversion of one real and 20 complex n^{th}-order matrices.

It can be seen from Figure 2 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 3 shows the results using 30 Arnoldi vectors and 15 POD basis vectors. The agreement at low frequencies is now 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.

Not only is the accuracy of the FMR model in Figures 2 and 3 better than the other models, but also the tech-

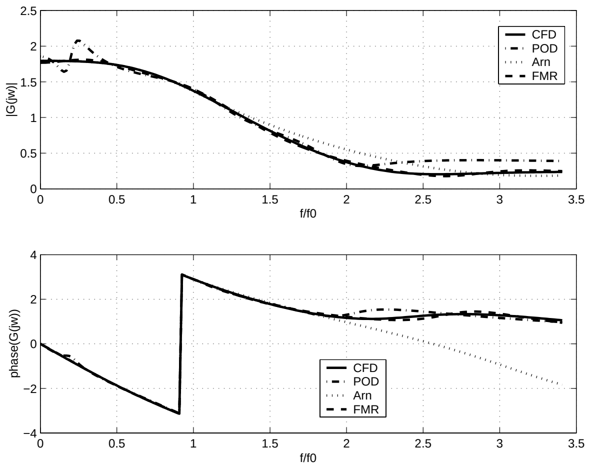

Fig. 2. Transfer function from bleed actuation to average throat Mach number for supersonic diffuser. Results from CFD model ($n = 11,730$) are compared to FMR, POD and Arnoldi models with $k = 10$ states.

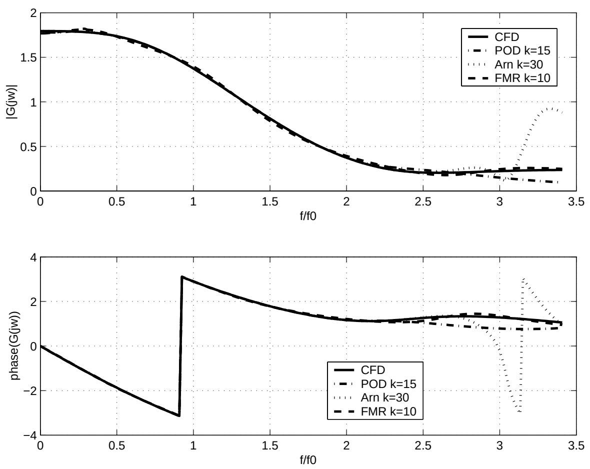

Fig. 3. Transfer function from bleed actuation to average throat Mach number for supersonic diffuser. Results from CFD model ($n = 11,730$) are compared to FMR with $k = 10$ states, POD with $k = 15$ states, and Arnoldi with $k = 30$ states.

nique compares favorably to the alternative methods in other aspects. One significant advantage of FMR is that, if the original CFD system is stable, the reduced-order models are also guaranteed to be stable. This is not the case for the Arnoldi and POD methods, and in practice, these techniques often result in unstable reduced models. For the supersonic diffuser considered here, this was observed for the Arnoldi model if too many basis vectors were used.

Secondly, the computational cost of deriving the models is of the same order for both the FMR and Arnoldi methods, that is, on the order of one n^{th}-order matrix inversion. The FMR requires some extra computation to transform from discrete to continuous time, plus the extra step of balanced truncation on the intermediate reduced model; however, the number of matrix solves typically dominates the reduction cost for large-scale applications. The computational cost of the POD method is much higher than for the other methods.

Finally, the FMR model yields more accurate results over a large frequency range, with fewer states. This is partly due to the fact that FMR considers both inputs and outputs in the reduction process, while Arnoldi and POD consider only system inputs and therefore yield inefficient models. Moreover, POD models are restricted to the frequency range contained within the snapshot ensemble. Arnoldi models are restricted to a frequency range close to that chosen for the Taylor series expansion, in this case $s_0 = 0$. The frequency range of the FMR model is controlled by the choice of the parameter ω_0 (see Willcox and Megretski (2003) for details). For this case, choosing $\omega_0 = 5$ enabled a large frequency range to be approximated accurately.

4. CONCLUSIONS

Three reduction methods have been compared for a large CFD system: the POD, Arnoldi method and Fourier model reduction. The most effective approach to reduction is found to be use of FMR to derive an intermediate Hankel matrix of order several hundred, followed by balanced truncation to obtain the final reduced-order model. While POD is the most commonly used method for reduction of CFD systems, it has several drawbacks. Firstly, the POD method is expensive, since a complex CFD system solve is required for each frequency considered; moreover, a large number of frequencies must be included if the reduced model is to yield accurate results. Secondly, the POD method only accounts for inputs when performing the reduction, and thus the resulting reduced models are often inefficient. Finally, no guarantees are available for the quality of the reduced-order model; in particular, the models may be unstable.

The Arnoldi method is better than the POD in terms of computational expense; however, this approach also does not consider outputs in the reduction process and is not guaranteed to produce a stable reduced-order model. This issue of stability was found to be a problem in the supersonic diffuser flow considered in this paper. The new FMR procedure addresses all these issues and provides an efficient, reliable method for model reduction of large scale linear systems.

5. REFERENCES

Adamjan, V.M., D.Z. Arov and M.G. Krein (1971). Analytic Properties of Schmidt Pairs for a Hankel Operator and the Generalized Schur-Takagi Problem. *Math. USSR Sbornik* **15**, 31–73.

Baker, M.L., D.L. Mingori and P.J. Goggins (1996). Approximate Subspace Iteration for Constructing Internally Balanced Reduced Order Models of Unsteady Aerodynamic Systems. AIAA Paper 96-1441.

Bettayeb, M., L.M. Silverman and M.G. Safonov (1980). Optimal Approximation of Continuous-Time Systems. *Proceedings of the 19th IEEE Conference on Decision and Control*.

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

Gudmundsson, T. and A.J. Laub (1994). Approximate Solution of Large Sparse Lyapunov Equations. *IEEE Transactions on Automatic Control* **39**(5), 1110–1114.

Hall, K.C. (1994). Eigenanalysis of Unsteady Flows About Airfoils, Cascades and Wings. *AIAA Journal* **32**(12), 2426–2432.

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

Jaimoukha, I.M. and E.M. Kasenally (1994). Krylov Subspace Methods for Solving Large Lyapunov Equations. *SIAM Journal of Numerical Analysis* **31**(1), 227–251.

Kim, T. (1998). Frequency-Domain Karhunen-Loeve Method and Its Application to Linear Dynamic Systems. *AIAA Journal* **36**(11), 2117–2123.

Kung, S-Y. and D.W. Lin (1981). Optimal Hankel-Norm Model Reductions: Multivariable Systems. *IEEE Transactions on Automatic Control* **AC-26**(1), 832–52.

Lassaux, G. (2002). High-Fidelity Reduced-Order Aerodynamic Models: Application to Active Control of Engine Inlets.. Master's thesis. Dept. of Aeronautics and Astronautics, MIT.

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

Moore, B.C. (1981). Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction. *IEEE Transactions on Automatic Control* **AC-26**(1), 17–31.

Scottedward Hodel, A. (1991). Least Squares Approximate Solution of the Lyapunov Equation. Proceedings of the 30th IEEE Conference on Decision and Control.

Sirovich, L. (1987). Turbulence and the Dynamics of Coherent Structures. Part 1 : Coherent Structures. *Quarterly of Applied Mathematics* **45**(3), 561–571.

Sorensen, D.C. and A.C. Antoulas (2002). The sylvester equation and approximate balanced reduction. Linear Algebra and Its Applications, Fourth Special Issue on Linear Systems and Control, edited by Blondel et al., 351-352, pp. 671-700.

Willcox, K. and A. Megretski (2003). Fourier Model Reduction for Large-Scale Applications in Computational Fluid Dynamics. To be presented at SIAM Conference on Applied Linear Algebra, Williamsburg, VA.

Willcox, K.E., J. Peraire and J. White (2002). An Arnoldi approach for generation of reduced-order models for turbomachinery. *Computers and Fluids* **31**(3), 369–89.