Reduced-order models of linearized channel flow using balanced truncation

Clarence W. Rowley and Milos Ilak

Abstract—Fluid systems often exhibit inherently low-dimensional behavior, even though the governing equations are complex and high-dimensional. At this time, full 3D discretizations of the Navier-Stokes equations are too computationally intensive to be used for control synthesis, so low-order models of the flow physics are desirable. This paper compares two different model-reduction procedures for the linearized flow in a plane channel: the method of Proper Orthogonal Decomposition and Galerkin projection, popular in the fluid mechanics community; and balanced truncation, a common method for model reduction of linear systems. Standard methods of computing balancing transformations are computationally intractable for systems of this size, so we use a numerical approximation of balanced truncation using empirical Gramians computed from simulations of the linearized and adjoint systems. For the channel flow considered here, the subspaces spanned by POD modes and balancing modes are very close, but reduced-order models from approximate balanced truncation perform much better than the standard POD models, for the same order of truncation.

I. INTRODUCTION

A better understanding of the mechanisms leading to transition from laminar to turbulent flow, and the potential for actuators to affect these mechanisms, may lead to drag reduction and substantial fuel savings for aircraft and other applications. This paper discusses low-dimensional models for a transitional flow, and compares two different techniques, focusing on the effects of actuation on the flow.

The incompressible flow in a plane channel is considered, and in particular this study focuses on the linearized equations, as in several previous studies [1], [2], [3], [4], [5]. We begin with the linearized Navier-Stokes equations and obtain reduced-order models using two separate approaches: projection of the governing equations onto empirical basis functions found from Proper Orthogonal Decomposition (POD) of data from simulations; and balanced truncation. While the POD/Galerkin method is computationally tractable even for very large systems, and has been used on many fluids problems, including boundary layers [6], cavity flows [7], [8], cylinder wakes [9], airfoils [10], and many others, balanced truncation has been used for fluids problems by only a few researchers [3], [2], and even then, in only one spatial dimension, because the computations rapidly become intractable as the number of states increases. In a typical fluids problem, the number of states is on the order of $10^5$ or more, and since computing balancing transformations involves solving $n \times n$ Lyapunov equations (where $n > 10^5$ is the number of states) as well as correspondingly large eigenvalue problems, traditional approaches to balanced truncations may not be feasible. Recently, snapshot-based methods for computing balanced truncations have been suggested, both to extend the concepts to nonlinear systems [11], [12] and to large linear systems [13]. In the present paper, we apply these methods to models of plane channel flow, and compare the resulting models to POD/Galerkin models. In §II we give a brief overview of these model reduction methods; in §III we describe the channel flow studied; and in §IV we compare the low-order models obtained by the various methods.

II. MODEL REDUCTION METHODS

Here, we review the two model reduction methods we will compare: Galerkin projection onto basis functions determined by Proper Orthogonal Decomposition (POD) of a particular dataset [14], [15]; and balanced truncation, using empirical Gramians built from snapshots from simulations of impulse responses [11], [12], [13].

A. POD and Galerkin projection

The idea of Galerkin projection is, given dynamics

$$\dot{x} = f(x), \quad x(t) \in \mathcal{X},$$

where $\mathcal{X}$ is a high-dimensional Hilbert space, to project onto a low-dimensional subspace $S \subset \mathcal{X}$, i.e., obtaining a reduced order model for a variable $r \in S$ as

$$\dot{r} = P_S f(r),$$

where $P_S : \mathcal{X} \rightarrow S$ is the orthogonal projection. One way of choosing this subspace $S$ is to gather a set of data $\{\mathbf{x}_j \in \mathcal{X} \mid j = 1, \ldots, m\}$ representative of typical behavior of the system, and to determine the subspace of a fixed dimension that optimally spans this dataset:

$$S = \arg \min_S \sum_{j=1}^{m} \|\mathbf{x}_j - P_S \mathbf{x}_j\|^2. \quad (1)$$

Proper Orthogonal Decomposition determines an orthogonal basis for such a subspace, which may be computed by finding the left singular vectors of the data matrix, whose columns are the snapshots $\mathbf{x}_j$.

Advantages of the POD/Galerkin method are that it is computationally tractable even for very large datasets, for instance, even when the dimension of $\mathcal{X}$ is on the order of millions, as is often the case for typical fluids simulations. However, disadvantages are that the subspace determined by (1) truncates the lowest-energy modes, and low-energy

This work was supported by the National Science Foundation, award CMS-0347239.
C.W. Rowley and M. Ilak are with the Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ 08544, USA cwrowley@princeton.edu, milak@princeton.edu
modes may be important for the dynamics. For linear systems, another interpretation is that the POD modes of data arising from the input-state impulse response are the most controllable modes of the linear system (eigenvectors corresponding to largest eigenvalues of the controllability Gramian) [13], so projecting onto these may neglect important modes that may be weakly controllable but strongly observable. Another disadvantage is that the method depends on the choice of inner product, and different inner products can drastically change the behavior of the models [13], and even the stability type of equilibria [7]. Ideally, one could imagine optimizing over not just the choice of the subspace, but also the choice of inner product. The method described in the next section achieves this goal at least partially: while not optimal, balanced truncation essentially does select both the subspace for projection and the inner product for projection, not through an optimization, but through a heuristic procedure that typically performs very well, and has provable error bounds.

B. Snapshot-based balanced truncation

Balanced truncation is typically used for model reduction of stable linear input-output systems of the form

$$\dot{x} = Ax + Bu \quad y = Cx$$

where $x(t) \in X$ is the state, $u(t) \in U$ is the input, and $y(t) \in Y$ is the output, and $X, U, Y$ are Hilbert spaces. If the system is both controllable and observable, one can determine a coordinate system in which the controllability and observability Gramians, defined respectively by

$$W_c = \int_0^\infty e^{At}BB^*e^{A^*t} dt, \quad W_o = \int_0^\infty e^{At}C^*Ce^{A^*t} dt,$$

are equal and diagonal (equal to the matrix of Hankel singular values). One then truncates the states that are least controllable and observable, corresponding to the smallest Hankel singular values [16], [17]. This procedure, while not actually optimal in any known sense, has $H_\infty$ error bounds that are typically close to the best attainable by any reduced order model: recall that the error between a full model $G(s)$ and any reduced order model $G_r(s)$ with state dimension $r$ is bounded by

$$\|G - G_r\|_\infty \geq \sigma_{r+1},$$

where $\sigma_j$ is the $j$-th Hankel singular value (in decreasing order). Extensions are also available for unstable [18], [19] and nonlinear systems [11], [20].

Here, we are interested in the case where the dimension of $X$ is very large, on the order of millions, so direct computation of the full Gramians is not possible. In addition, the output space $Y$ may be very large as well: for instance, we may be interested in capturing an entire fluid flow field accurately, so that the output is the entire state (though possibly with a different inner product). In order to compute the balancing transformation in this case, we construct empirical Gramians, as in [11], and compute the balancing transformation using the method of snapshots described in [13], in which one uses a single SVD of a data matrix formed by inner products of snapshots from forward and adjoint simulations, as described below.

1) Empirical Gramians: Here, we will assume that the number of inputs $u$ is small. We first perform simulations of the system (2), with $x(0) = 0$, and impulsive inputs on each component of $u$. After [11], the empirical controllability Gramian may be computed by assembling the snapshots from these simulations (solutions $x(t_j)$ at different times $t_j$, $j = 1, \ldots, m$) as columns of a matrix $X : \mathbb{R}^m \to \mathcal{X}$, appropriately weighted by quadrature coefficients so that $W_c = XX^*$ approximates the integral (3). One may compute the empirical observability Gramian analogously, by computing solutions to the adjoint system

$$\dot{z} = A^*z + C^*v, \quad v \in \mathcal{Y}$$

forward in time, with $z(0) = 0$ and impulses on each component of $v$. Here $C^* : \mathcal{Y} \to \mathcal{X}$ is the adjoint of $C : \mathcal{X} \to \mathcal{Y}$, which of course depends on the inner products in both $\mathcal{X}$ and $\mathcal{Y}$, so care must be taken if nonstandard inner products are used (as they will be in §III). From these simulations, one obtains $l$ snapshots $z_j$ which are appropriately weighted and assembled as columns of a data matrix $Y : \mathbb{R}^l \to \mathcal{X}$, so that the observability Gramian in (3) is approximated by $YY^*$.

Note that if the number of outputs is large, the computation of the adjoint snapshots may not be tractable, since one simulation is needed for each component of the output. Below, we will consider the output $y = x$, as we wish our model to reproduce the full flow information, so the number of outputs is indeed large, and we require an alternate approach. In particular, we first project the output onto a low-dimensional subspace, i.e., taking $y = P_rCx$, where $P_r$ is an orthogonal projection onto an $r$-dimensional subspace of $\mathcal{Y}$. Conveniently, the projection $P_r$ that minimizes the 2-norm of the difference between the original transfer function and the output-projected transfer function is given simply by POD of the set of impulse responses, which has already been computed in the previous step [13]. Writing this projection as $P_r = \Phi_r\Phi_r^*$, where columns of $\Phi_r : \mathbb{R}^r \to \mathcal{Y}$ are POD modes, one needs only to compute $r$ impulse responses of the system

$$\dot{z} = A^*z + C^*\Phi_r w, \quad w \in \mathbb{R}^r.$$  

2) Balancing transformation: Once the primal and adjoint snapshots are computed, as described in [13], the balancing transformation that diagonalizes the empirical Gramians may be found by an SVD of the matrix $Y^*X : \mathbb{R}^m \to \mathbb{R}^l$, whose dimensions are the number of primal snapshots × the number of adjoint snapshots, which is typically manageable (e.g., $< 10^4$). Forming the SVD

$$Y^*X = U\Sigma V^*,$$

one easily shows $\Sigma \in \mathbb{R}^{r \times p}$ is the diagonal matrix of nonzero Hankel singular values, and $U \in \mathbb{R}^{l \times p}$ and $V \in \mathbb{R}^{m \times p}$.
satisfy $U^*U = V^*V = I_p$. The balancing transformation is then found by computing matrices

$$T = XV\Sigma^{-1/2}, \quad S = \Sigma^{-1/2}U^*Y^*$$

which satisfy $ST = I_p$. If $p = n$ (i.e., if the system is controllable and observable, and enough snapshots have been taken), then $T = S^{-1}$ is the transformation that diagonalizes the Gramians. If $p < n$, as will be the case here, then as shown in [13], $T$ block diagonalizes the empirical Gramians, extracting the first $p$ Hankel singular values.

Knowing both $S$ and $T$, one may obtain balanced truncations directly, without literally transforming the entire state to new coordinates (which would be computationally intractable for systems of this order) and subsequently truncating. If $q$ is the desired order of the reduced-order model, one lets $T_q$ denote the first $q$ columns of $T$, and $S_1$ the first $q$ rows of $S$, and then the reduced-order model is given by

$$\dot{z} = S_1AT_1z + S_1B$$
$$\dot{y} = CT_1z$$

Note, however, that this is not equivalent to orthogonal projection onto the first $q$ columns of $T$, since these columns are not orthogonal with respect to the standard inner product. (Instead, the columns of $T$ and the rows of $S$ form a bi-orthogonal set, as discussed in [13].)

III. PLANE CHANNEL FLOW

We consider the flow between two parallel plates, as indicated in Fig. 1. The domain is periodic in the streamwise $x$ and spanwise $z$ directions, and we linearize about a laminar base flow. (Here, $x$, $y$, and $z$ denote coordinates in physical space, not to be confused with the state, output, and adjoint variables from the previous section.) Non-dimensionalizing by the channel half-height $\delta$ and the centerline velocity $U_{cl}$, the base flow has the form $U(y) = 1 - y^2$.

A. Linearized equations

For this geometry, the linearized equations may be conveniently written in terms of the wall-normal velocity $v$ and the wall-normal vorticity $\eta$ [1]. The other variables (e.g., streamwise and spanwise velocities $u$ and $w$) may then be computed using the continuity equation $\partial_x u + \partial_y v + \partial_z w = 0$. In these coordinates, the linearized (non-dimensional) equations have the form

$$\frac{\partial}{\partial t} \begin{bmatrix} \Delta & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} v \\ \eta \end{bmatrix} = \begin{bmatrix} L_{OS} & 0 \\ -U'\partial_z & L_{SQ} \end{bmatrix} \begin{bmatrix} v \\ \eta \end{bmatrix}$$

where $\Delta = \partial_x^2 + \partial_y^2 + \partial_z^2$ is the Laplacian, and

$$L_{OS} = U\partial_z \Delta - U''\partial_x - \frac{1}{R}\Delta^2$$
$$L_{SQ} = -U\partial_x + \frac{1}{R}\Delta$$

are the Orr-Sommerfeld and Squire operators, respectively. Here, $R = U_{cl}\delta/\nu$ is the Reynolds number, where $\nu$ is the kinematic viscosity. It has been shown numerically that the above linearized system is stable up to $R \approx 5772$, though severe non-normality exists, and large transient response is observed [21], [11], [5].

B. Inner product and adjoint equations

To determine the corresponding adjoint equations, one first needs to define an inner product on the vector space $\mathcal{X}$ of flow variables $(v, \eta)$. It will be convenient to define the inner product

$$\langle (v_1, \eta_1), (v_2, \eta_2) \rangle = \int_\Omega (-v_1\Delta v_2 + \eta_1\eta_2) \, dx \, dy \, dz,$$

where $\Omega$ denotes the fluid volume. Note that, letting $M : \mathcal{X} \to \mathcal{X}$ denote the operator on the left hand side of (8), this is just the $L_2$ inner product of $(v_1, \eta_1)$ with $M(v_2, \eta_2)$.

With this definition of the inner product, the adjoint equations are easily found by integration by parts:

$$\frac{\partial}{\partial t} \begin{bmatrix} -\Delta & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} v \\ \eta \end{bmatrix} = \begin{bmatrix} L_{OS}^* & 0 \\ 0 & L_{SQ}^* \end{bmatrix} \begin{bmatrix} v \\ \eta \end{bmatrix}$$

where

$$L_{OS}^* = -U\partial_x \Delta - 2U'\partial_z \partial_y - \frac{1}{R}\Delta^2$$
$$L_{SQ}^* = U\partial_x + \frac{1}{R}\Delta.$$

C. Numerical method

Both the linearized (8) and adjoint (9) equations are solved numerically, using Fourier modes in the $x$ and $z$ directions and Chebyshev polynomials in the wall-normal ($y$) direction, using a method similar to that in [22]. Two different simulations are considered here: first, for validation purposes, streamwise-constant perturbations are considered, at a grid resolution that is small enough that exact balanced truncation may be computed using standard routines in Matlab (16 × 17 gridpoints, state dimension $n = 480$). Next, full 3D perturbations are considered at a grid resolution for which exact balanced truncation is intractable (16 × 17 × 16 grid, $n = 7680$). For both cases, the $R = 100$, with a domain $[0, 2\pi]$ in the $x$ and $z$ directions. To compute the POD and balancing modes, both simulations use 1000 snapshots equally spaced in the interval $t \in [0, 200]$, and the timestep for the simulations is 0.002 for both linearized and adjoint simulations. Though the grid is coarse, the simulations are fully resolved at this Reynolds number, with the first gridpoint from the wall at $y^+ = 0.27$. 

Fig. 1. Flow geometry for plane channel flow.
IV. RESULTS

A. Streamwise-constant perturbations

Figure 2 shows $H_\infty$ norms for the error between the full order model, whose transfer function is denoted $G(s)$, and reduced-order models of state dimension $r$, whose transfer function is denoted $G_r(s)$. Here, the system is small enough ($n = 480$) that exact balanced truncation may be computed using standard routines, and the error from the snapshot-based approximate balanced truncation is almost identical to that from the full balanced truncation. For most values of $r$, the errors from balanced truncation, either exact or snapshot-based, are smaller than the error from the POD/Galerkin model, and are close to the lower bound for any reduced-order model (obtained from the magnitude of the first truncated Hankel singular value [17]). For further results on the streamwise-constant case, including plots of the modes themselves, see [13] (in this reference, a slightly different forcing term was used, but results are qualitatively the same).

B. Three-dimensional perturbations

Full three-dimensional perturbations were introduced using a localized body force in a small region in the center of the domain. Sample velocity fields at two separate times ($t = 9$ and $t = 30$) are shown in Fig. 3, which illustrate how the localized perturbation becomes stretched in the streamwise direction, and eventually resembles the streamwise-constant perturbations considered earlier.

Fig. 4 shows the first three POD modes from snapshots of the impulse response for this simulation, and the eigenvalues are shown in Fig. 5. Note that mode 1 is approximately streamwise constant, and modes 2 and 3 are approximately phase-shifted versions of each other, corresponding to a traveling structure in the streamwise direction. The POD eigenvalues fall off slower than in the streamwise-constant case, but the decay is still exponential, as expected. The first 5 modes capture 69% of the energy, while the first 10 modes capture 88%. In what follows, we consider snapshot-based balanced truncation, with the output projected on the first 10 POD modes, as in (5).

Ten adjoint simulations are performed, one for each (projected) output, and from these snapshots and the forward snapshots already computed, the Hankel singular values $\sigma_j$ are computed from (6), and plotted in Fig. 5. Note that $\sigma_2$ and $\sigma_3$ are approximately equal, as we would expect for an oscillator (analogous to the behavior of POD modes 2 and 3, whose amplitudes oscillate, forming a traveling structure). For comparison, the values of $\sigma_j$ computed using only a 5-mode output projection are also shown, and these agree with the 10-mode output projection only up to about $j = 4$.

The primal and dual modes from this procedure (what we call balanced POD modes) are shown in Figs. 6 and 7. The primal modes are columns of the matrix $T$ in (7), and the adjoint modes are rows of the matrix $S$. The primal modes look strikingly similar to the POD modes from Fig. 4: indeed, the subspaces spanned by these modes are very close. However, the balanced POD modes are of course orthogonal with respect to a different inner product, and their adjoint modes shown in Fig. 7 look very different from the POD modes. Of course, for standard POD, the adjoint modes are equal to the primal modes. Thus, for this example, the main difference between POD and balanced truncation (or balanced POD) is not in the subspace one projects onto, but in the inner product used in the projection of the dynamics.

One might expect that, since the subspaces are so close, that the behavior of the reduced-order models from POD and balanced truncation might be about the same. However, as shown in Fig. 8, the models from balanced truncation actually perform significantly better. This figure shows the norm of the solution vs. time, with zero initial conditions and an impulsive input, illustrating the non-normal transient growth. The 5-mode POD model performs very poorly, completely missing the transient growth, while the 5-mode balanced POD model captures this transient quite well, about
Fig. 3. Contours and level sets of wall-normal vorticity $\eta$ from the response to an impulsive body force in a small region in the center of the domain, at time $t = 9$ (left) and $t = 30$ (right), showing the evolution of streamwise structures. Flow is left to right, and walls are at the top and bottom.

Fig. 4. First three POD modes from impulse response of 3D simulation, showing contours and level sets of wall-normal velocity $v$ (top) and vorticity $\eta$ (bottom).

Fig. 6. First three primal modes from balanced POD, showing wall-normal velocity $v$ (top) and vorticity $\eta$ (bottom).
as well as the 10-mode POD model. The curve for 10-mode balanced POD is almost identical to that of the full simulation.

REFERENCES


