

Construction of Energy-Stable Projection-Based Reduced Order Models

Irina Kalashnikova^a, Matthew F. Barone^b, Srinivasan Arunajatesan^b, Bart G. van Bloemen Waanders^c

^aComputational Mathematics Department, Sandia National Laboratories, P.O. Box 5800, MS 1320, Albuquerque, NM 87185-1320.

^bAerosciences Department, Sandia National Laboratories, P.O. Box 5800, MS 0825, Albuquerque, NM 87185-0825.

^cOptimization and Uncertainty Quantification Department, Sandia National Laboratories, P.O. Box 5800, MS 1318, Albuquerque, NM 87185-1318.

Abstract

An approach for building energy-stable Galerkin reduced order models (ROMs) for linear hyperbolic or incompletely parabolic systems of partial differential equations (PDEs) using continuous projection is developed. This method is an extension of earlier work by the authors specific to the equations of linearized compressible inviscid flow. The key idea is to apply to the PDEs a transformation induced by the Lyapunov function for the system, and to build the ROM in the transformed variables. For linear problems, the desired transformation is induced by a special inner product, termed the “symmetry inner product”, which is derived herein for several systems of physical interest. Connections are established between the proposed approach and other stability-preserving model reduction methods, giving the paper a review flavor. More specifically, it is shown that a discrete counterpart of this inner product is a weighted L^2 inner product obtained by solving a Lyapunov equation, first proposed by Rowley *et al.* and termed herein the “Lyapunov inner product”. Comparisons between the symmetry inner product and the Lyapunov inner product are made, and the performance of ROMs constructed using these inner products is evaluated on several benchmark test cases.

Keywords: Reduced order model (ROM), proper orthogonal decomposition (POD)/Galerkin projection, linear hyperbolic/incompletely parabolic systems, linear time-invariant (LTI) systems, numerical stability, Lyapunov equation.

1. Introduction

Numerous modern-day engineering problems require the simulation of complex systems with tens of millions or more unknowns. Despite improved algorithms and the availability of massively parallel computing resources, “high-fidelity” models are, in practice, often too computationally expensive for use in a design or analysis setting. The continuing push to incorporate into modeling efforts the quantification of uncertainties, critical to many science and engineering applications, can present an intractable computational burden due to the high-dimensional systems that arise. This

situation has prompted researchers to develop reduced order models (ROMs): models constructed from high-fidelity simulations that retain the essential physics and dynamics of their corresponding full order models (FOMs), but have a much lower computational cost. Since ROMs are, by construction, small, they can enable uncertainty quantification (UQ) as well as on-the-spot decision making and/or control.

In order to serve as a useful predictive tool, a ROM should possess the following properties: consistency (with respect to its corresponding high-fidelity model), stability, and convergence (to the

solution of its corresponding high-fidelity model). The second of these properties, namely numerical stability, is particularly important, as it is a prerequisite for studying the convergence and accuracy of a ROM. It is well-known that popular model reduction approaches known as the proper orthogonal decomposition (POD) method [25; 26; 19] and the balanced proper orthogonal decomposition (BPOD) method [31; 23] lack, in general, an *a priori* stability guarantee. In [30], Amsallem *et al.* suggest that POD ROMs constructed for linear time-invariant (LTI) systems in descriptor form tend to possess better numerical stability properties than POD ROMs constructed for LTI systems in non-descriptor form. Although heuristics such as these exist, it is in general unknown *a priori* if a ROM constructed using POD or BPOD will preserve the stability properties of the high-fidelity system from which the model was constructed. There *does* exist a model reduction technique that has a rigorous stability guarantee, namely balanced truncation [29; 10]; however, the computational cost of this method, which requires the computation and simultaneous diagonalization of infinite controllability and observability Gramians, makes balanced truncation computationally intractable for systems of very large dimensions (i.e., systems with more than 10,000 degrees of freedom [24]).

The importance of obtaining stable ROMs has been recognized in recent years by a number of authors. It is shown by Patera, Veroy and Rozza in [27; 28] that a stable ROM can be constructed using the reduced basis method. In [24], Rowley *et al.* show that Galerkin projection preserves the stability of an equilibrium point at the origin if the ROM is constructed in an “energy-based” inner product. In [6; 7], Barone *et al.* demonstrate that a symmetry transformation leads to a stable formulation for a Galerkin ROM for the linearized compressible Euler equations [6; 7] with solid wall and far-field boundary conditions. In [1], Serre *et al.* propose applying the stabilizing projection developed by Barone *et al.* in [6; 7] to a skew-symmetric system constructed by augmenting a given linear system with its adjoint system. This approach yields a

ROM that is stable at finite time even if the solution energy of the full-order model is growing.

The methods described above derive (*a priori*) a stability-preserving model reduction framework that is specific to a particular equation set. There exist, in addition to these techniques, approaches which stabilize an unstable ROM through a post-processing (*a posteriori*) stabilization step applied to an unstable algebraic ROM system. Ideally, the stabilization is such that it will only minimally modify the ROM. In [5], Amsallem *et al.* propose a method for stabilizing projection-based linear ROMs through the solution of a small-scale convex optimization problem. In [39], a set of linear constraints for the left-projection matrix, given the right-projection matrix, are derived by Bond *et al.* to yield a projection framework that is guaranteed to generate a stable ROM. An approach for stabilizing unstable ROMs for LTI systems, termed ROM stabilization via optimization-based eigenvalue reassignment, was proposed by Kalashnikova *et al.* in the recent work [56]. In this approach, the unstable eigenvalues of an unstable ROM are modified through the numerical solution of a constrained nonlinear least-squares optimization problem formulated such that the error in the stabilized ROM output is minimal. In [40], a ROM stabilization methodology that achieves improved accuracy and stability through the use of a new set of basis functions representing the small, energy-dissipation scales of turbulent flows is derived by Balajewicz *et al.* In [35], Zhu *et al.* derive some large eddy simulation (LES) closure models for POD ROMs for the incompressible Navier-Stokes equations, and demonstrate numerically that the inclusion of these LES terms yields a ROM with increased numerical stability (albeit at the sacrifice of consistency of the ROM with respect to the direct numerical simulation (DNS) from which the ROM is constructed).

In this article, several approaches to building stable ROMs for linear systems, both in the continuous as well as in the discrete projection setting, are presented, connected and extended. The article has

a review flavor, but contains several new contributions, most notably the following:

- The energy-stable continuous projection ROM method developed specifically for the equations of linearized compressible inviscid flow in [6; 7] is extended to generic systems of PDEs of the hyperbolic or incompletely parabolic type.
- A stability preserving symmetry inner product is derived for several physical systems (the wave equation, the linearized shallow water equations, the linearized compressible Euler equations, the linearized compressible Navier-Stokes equations).
- Connections between the proposed energy-stable continuous projection method and other model reduction techniques with an *a priori* stability guarantee, e.g., a discrete projection approach involving a Lyapunov equation-based inner product introduced by Rowley *et al.* in [24], are established using the concept of energy stability.
- Numerical studies evaluating the performance of ROMs constructed in the energy inner products described herein are performed.

The remainder of this paper is organized as follows. The first part consists of some preliminaries: projection-based model reduction (in particular, the POD¹/Galerkin method) is overviewed (Section 2), and several notions of stability (energy-stability, Lyapunov stability, asymptotic stability, exponential stability, time-stability) are defined (Section 3). Attention is then turned to the construction of energy-stable ROMs for linear systems of PDEs using continuous projection (Section 4). The energy-stability preserving model reduction

¹For concreteness, it is assumed herein that the reduced basis is constructed via the POD method, as the POD is a popular method for computing reduced bases that is feasible even for very large systems but can give rise to unstable ROMs. It is emphasized that the energy-stability results discussed herein hold for *any* choice of reduced basis, not just the POD basis, however.

approach developed specifically for the equations of linearized compressible inviscid flow in [6; 7] is generalized. Examples of this inner product are given for several systems of physical interest, and some numerical results are presented. Next, it is shown that a certain transformation applied to a generic linear hyperbolic or incompletely parabolic set of PDEs and induced by the Lyapunov function for these equations will yield a Galerkin ROM that is stable for *any* choice of reduced basis. It is then shown that, for many PDEs, the desired transformation is induced by a special weighted L^2 inner product, termed the “symmetry inner product”. It is also demonstrated that a discrete weighted L^2 inner product first derived by Rowley *et al.* in [24] and termed herein the “Lyapunov inner product” is a discrete counterpart of the symmetry inner product. The weighting matrix that defines the Lyapunov inner product can be computed in a black-box fashion for a stable LTI system arising from the discretization of a linear system of PDEs in space. Numerical studies of POD ROMs constructed in the Lyapunov inner product are performed. A unifying summary of the energy-stability preserving model reduction approaches described within this paper is given Section 6, along with some conclusions. It is anticipated that this discussion will aid the reader in selecting the most appropriate model reduction methodology for his/her application.

2. Projection-based model reduction

In this section, several approaches to building projection-based reduced order models are reviewed. Attention is restricted to LTI systems. A system is called time-invariant if the output response for a given input does not depend on when that input is applied [15].

At the continuous level, an LTI system can be represented by a PDE (or system of PDEs) of the form

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathcal{L}(\mathbf{x}(t)) + \mathcal{L}_c(\mathbf{u}(t)), \\ \mathbf{y}(t) &= \mathcal{L}_o(\mathbf{x}(t)),\end{aligned}\tag{1}$$

in Ω . Here, t denotes time, $\mathbf{x} \in \mathbb{R}^n$ is called the state vector, $\mathbf{u} \in \mathbb{R}^p$ represents the vector of con-

control variables, $\mathbf{y} \in \mathbb{R}^q$ is the measured signal or output, Ω is a bounded domain, and the ‘ \cdot ’ symbol denotes differentiation with respect to time, i.e., $\dot{\mathbf{x}} \equiv \frac{\partial \mathbf{x}}{\partial t}$. The operator $\mathcal{L} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a smooth linear spatial-differential operator, and $\mathcal{L}_c : \mathbb{R}^p \rightarrow \mathbb{R}^n$ and $\mathcal{L}_o : \mathbb{R}^n \rightarrow \mathbb{R}^q$ are smooth linear mappings. The abstract operators \mathcal{L} , \mathcal{L}_c and \mathcal{L}_o are introduced to keep the discussion as general as possible, and used in subsequent analysis.

Suppose the PDE system (1) has been discretized in space using some numerical scheme, e.g., the finite element method. The result will be a semi-discrete LTI system of the form:

$$\begin{aligned}\dot{\mathbf{x}}_N(t) &= \mathbf{A}\mathbf{x}_N(t) + \mathbf{B}\mathbf{u}_P(t) \\ \mathbf{y}_{QN}(t) &= \mathbf{C}\mathbf{x}_N(t).\end{aligned}\quad (2)$$

Here, $\mathbf{x}_N \in \mathbb{R}^N$ is the discretized state vector, $\mathbf{u}_P \in \mathbb{R}^p$ is the discretized vector of control variables, and $\mathbf{y}_{QN} \in \mathbb{R}^q$ is the discretized output; $\mathbf{A} \in \mathbb{R}^{N \times N}$, $\mathbf{B} \in \mathbb{R}^{N \times p}$ and $\mathbf{C} \in \mathbb{R}^{q \times N}$ are constant matrices (in particular, they are not functions of time t).

The general approach to projection-based model reduction consists of three steps, described below.

Step 1: Calculation of reduced trial and test bases, denoted by $\Phi_M = (\phi_1, \dots, \phi_M)$ and $\Psi_M = (\psi_1, \dots, \psi_M)$ respectively, each of order M , with $M \ll N$.

Step 2: Approximation of the solution to (1) by

$$\mathbf{x}(t) \approx \sum_{i=1}^M x_{M,i}(t) \phi_i = \Phi_M \mathbf{x}_M(t), \quad (3)$$

where $x_{M,i}(t)$ are the unknown ROM coefficients or modal amplitudes, to be determined in solving the ROM.

Step 3: Substitution of the approximation (3) into the governing system ((1) or (2)) and projection of this system onto the reduced test basis.

The result of this procedure is a ‘‘small’’ (size $M \ll N$) dynamical system that, for a suitable choice of reduced bases, accurately describes the dynamics of the full order system for some set of conditions. The reduced bases $\Phi_M \in \mathbb{R}^{N \times M}$ and $\Psi_M \in \mathbb{R}^{N \times M}$ are functions of space but not time,

and are assumed to have full column rank. In the case that $\Psi_M \neq \Phi_M$, the projection is referred to as a Petrov-Galerkin projection. Otherwise, if $\Psi_M = \Phi_M$, the projection is referred to as a Galerkin projection. This terminology is introduced here as it will be shown later that the energy-stable model reduction approaches derived in this work are effectively Petrov-Galerkin methods.

2.1. Calculation of the reduced bases (Step 1)

There exist a number of approaches for calculating the reduced basis modes (*Step 1* of the model reduction), e.g., the POD method [25; 26; 19], the BPOD method [31; 23], the balanced truncation method [29; 10], the reduced basis method [27; 28]; also methods based on goal-oriented bases [21], generalized eigenmodes [38], and Koopman modes [41]. Attention is restricted here to the POD basis, but it is noted that the energy-stability results derived in this paper hold for *any* choice of reduced basis. The reason for the choice of the POD reduced basis is two-fold. First, the POD is a widely used approach for computing efficient bases for dynamical systems. Moreover, ROMs constructed via the POD/Galerkin method lack in general an *a priori* stability guarantee (meaning POD/Galerkin ROMs would benefit from stability-preserving model reduction approaches such as those developed herein).

Discussed in detail in Lumley [16] and Holmes *et al.* [19], POD is a mathematical procedure that, given an ensemble of data and an inner product, denoted generically by (\cdot, \cdot) , constructs a basis for the ensemble. This basis is optimal in the sense that it describes more energy (on average) of the ensemble in the chosen inner product than any other linear basis of the same dimension M . The ensemble $\{\mathbf{x}^k : k = 1, \dots, K\}$ is typically a set of K instantaneous snapshots of a numerical solution field, taken for K values of a parameter of interest, or at K different times. Mathematically, POD seeks an M -dimensional ($M \ll K$) subspace spanned by the set $\{\phi_i\}$ such that the projection of the difference between the ensemble \mathbf{x}^k and its projection onto the

reduced subspace is minimized on average. It is a well-known result [6; 19; 34; 33] that the solution to the POD optimization problem reduces to the eigenvalue problem

$$\mathbf{R}\boldsymbol{\phi} = \lambda\boldsymbol{\phi}, \quad (4)$$

where \mathbf{R} is a self-adjoint and positive semi-definite operator with its (i, j) entry given by $R_{ij} = \frac{1}{K} (\mathbf{x}^i, \mathbf{x}^j)$ for $1 \leq i, j \leq K$. If it is assumed that \mathbf{R} is compact, then there exists a countable set of non-negative eigenvalues λ_i with associated eigenfunctions $\boldsymbol{\phi}_i$. It can be shown [19; 16] that the set of M eigenfunctions, or POD modes, $\{\boldsymbol{\phi}_i : i = 1, \dots, M\}$ corresponding to the M largest eigenvalues of \mathbf{R} is precisely the desired basis. This is the so-called ‘‘method of snapshots’’ for computing a POD basis [25].

2.2. Approximation of solution in reduced basis (Step 2)

Once the reduced basis is computed, the solution $\mathbf{x}(t)$ is approximated as a linear combination of the reduced basis modes (3) (Step 2). Given this approximation, the following error formula can be shown for the POD [19; 34]:

$$\frac{1}{K} \sum_{i=1}^K \left\| \mathbf{x}^i - \sum_{j=1}^M (\mathbf{x}^i, \boldsymbol{\phi}_j) \boldsymbol{\phi}_j \right\|^2 = \sum_{k=M+1}^J \lambda_k, \quad (5)$$

where $J = \dim(\{\mathbf{x}^1, \dots, \mathbf{x}^K\})$, and where $\lambda_1 \geq \dots \geq \lambda_J > 0$ are the positive eigenvalues of the operator \mathbf{R} (4).

Typically, the size of the reduced basis is chosen based on an energy criterion. That is, M is selected to be the minimum integer such that

$$E_{POD}(M) \geq \text{tol}, \quad (6)$$

where $0 \leq \text{tol} \leq 1$ represents the snapshot energy represented by the POD basis, and

$$E_{POD}(M) \equiv \frac{\sum_{i=1}^M \lambda_i}{\sum_{i=1}^K \lambda_i}. \quad (7)$$

2.3. Projection (Step 3)

There are two approaches for performing Step 3 of the model reduction: continuous and discrete projection. These approaches are described, as well as compared and contrasted, in the present subsection. Stability-preserving methods for constructing ROMs using these approaches will be detailed in Sections 4 and 5.

2.3.1. Model reduction via continuous projection

In the continuous projection approach [6; 7], the continuous system of PDEs (1) is projected onto a continuous reduced test basis $\{\boldsymbol{\psi}_i\}_{i=1}^M \in \mathbb{R}^n$ in a continuous inner product (\cdot, \cdot) , for example, the usual L^2 inner product²

$$(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = \int_{\Omega} \mathbf{x}^{(1)T} \mathbf{x}^{(2)} d\Omega, \quad (8)$$

where the $x_{M,i}(t)$ are the unknown ROM coefficients or modal amplitudes (so that $\mathbf{x}_M^T \equiv (x_{M,1}, \dots, x_{M,M})$), to be determined in solving the ROM dynamical system (derived below).

Substituting (3) into (1), the following is obtained

$$\begin{aligned} \sum_{i=1}^M \dot{x}_{M,i}(t) \boldsymbol{\phi}_i &= \mathcal{L}(\sum_{i=1}^M x_{M,i}(t) \boldsymbol{\phi}_i) \\ &+ \mathcal{L}_c(\mathbf{u}(t)), \\ \mathbf{y}_{QM}(t) &= \mathcal{L}_o(\sum_{i=1}^M x_{M,i}(t) \boldsymbol{\phi}_i), \end{aligned} \quad (9)$$

where $\mathbf{y}_{QM}(t)$ is the reduced approximation of the output.

Next, a reduced test basis $\{\boldsymbol{\psi}_i\}_{i=1}^M \in \mathbb{R}^n$ is introduced, and the system of PDEs (9) is projected onto the reduced test basis modes $\boldsymbol{\psi}_j$ for $j = 1, 2, \dots, M$ in the inner product (\cdot, \cdot) to yield

$$\begin{aligned} \sum_{i=1}^M \dot{x}_{M,i}(t) (\boldsymbol{\psi}_j, \boldsymbol{\phi}_i) &= \left(\boldsymbol{\psi}_j, \mathcal{L}(\sum_{i=1}^M x_{M,i}(t) \boldsymbol{\phi}_i) \right) \\ &+ \left(\boldsymbol{\psi}_j, \mathcal{L}_c(\mathbf{u}(t)) \right), \\ \mathbf{y}_{QM}(t) &= \mathcal{L}_o(\sum_{i=1}^M x_{M,i}(t) \boldsymbol{\phi}_i), \end{aligned} \quad (10)$$

for $j = 1, 2, \dots, M$. Typically, the reduced trial and test bases $\boldsymbol{\phi}_i$ and $\boldsymbol{\psi}_i$ are chosen to be orthonormal

²Weighted variants of the L^2 inner product are considered later in this work.

in the inner product (\cdot, \cdot) , so that $(\boldsymbol{\psi}_j, \boldsymbol{\phi}_i) = \delta_{ij}$, where δ_{ij} denotes the Kröner delta function. Invoking this property, as well as the linearity property of the operators \mathcal{L} and \mathcal{L}_o , (10) simplifies to

$$\begin{aligned}\dot{x}_{M,j}(t) &= \sum_{i=1}^M x_{M,i}(t) (\boldsymbol{\psi}_j, \mathcal{L}(\boldsymbol{\phi}_i)) \\ &+ (\boldsymbol{\psi}_j, \mathcal{L}_c(\mathbf{u}(t))), \\ \mathbf{y}_{QM}(t) &= \sum_{i=1}^M x_{M,i}(t) \mathcal{L}_o(\boldsymbol{\phi}_i),\end{aligned}\quad (11)$$

for $j = 1, 2, \dots, M$. The equations (11) define a set of M time-dependent ODEs for the modal amplitudes $x_{M,i}(t)$ in (3).

2.3.2. Model reduction via discrete projection

In the discrete projection approach, the FOM ODE system (2) (the PDE system discretized in space) is projected onto a discrete reduced test basis in a discrete inner product. Suppose this discrete inner product is the following weighted L^2 inner product:

$$\left(\mathbf{x}_N^{(1)}, \mathbf{x}_N^{(2)} \right)_{\mathbf{P}} = \mathbf{x}_N^{(1)T} \mathbf{P} \mathbf{x}_N^{(2)}, \quad (12)$$

where $\mathbf{P} \in \mathbb{R}^{N \times N}$ is a symmetric positive-definite matrix. Let $\boldsymbol{\Phi}_M \in \mathbb{R}^{N \times M}$ and $\boldsymbol{\Psi}_M \in \mathbb{R}^{N \times M}$ denote the reduced trial and reduced test bases for (2), respectively. Assume these matrices have full column rank, and are orthonormal in the inner product (12), so that $\boldsymbol{\Psi}_M^T \mathbf{P} \boldsymbol{\Phi}_M = \mathbf{I}_M$, where \mathbf{I}_M denotes the $M \times M$ identity matrix. The first step in constructing a ROM for (2) using discrete projection is to approximate the solution $\mathbf{x}_N(t)$ by (3). Substituting (3) into (2), and projecting this system onto the reduced test basis, the following $M \times M$ LTI system is obtained:

$$\begin{aligned}\dot{\mathbf{x}}_M(t) &= \mathbf{A}_M \mathbf{x}_M(t) + \mathbf{B}_M \mathbf{u}_P(t), \\ \mathbf{y}_{QM}(t) &= \mathbf{C}_M \mathbf{x}_M(t),\end{aligned}\quad (13)$$

where

$$\begin{aligned}\mathbf{A}_M &= \boldsymbol{\Psi}_M^T \mathbf{P} \mathbf{A} \boldsymbol{\Phi}_M, \\ \mathbf{B}_M &= \boldsymbol{\Psi}_M^T \mathbf{P} \mathbf{B}, \\ \mathbf{C}_M &= \mathbf{C} \boldsymbol{\Phi}_M,\end{aligned}\quad (14)$$

and where \mathbf{y}_{QM} is a reduced approximation of the output.

2.3.3. Continuous vs. discrete projection

In the majority of applications of reduced order modeling, the discrete projection approach is employed in constructing the ROM. This discrete approach has the advantage that boundary condition terms present in the discretized equation set are often (depending on the implementation) inherited by the ROM; that is, the ROM solution will satisfy the boundary conditions of the FOM. Certain properties of the numerical scheme used to solve the full equations may be inherited by the ROM as well. The discrete approach can be black-box, at least for linear systems of the form (2): it operates on the matrices \mathbf{A} , \mathbf{B} and \mathbf{C} , so that access to the high-fidelity code that was used to generate these matrices or the governing PDEs is not required provided these matrices can be written out from the high-fidelity code. In contrast, the continuous projection approach is tied to the governing PDEs – the continuous problem (1) needs to be translated to the discrete setting, e.g., by interpolating the reduced basis modes and evaluating the continuous inner products in (11) using a numerical quadrature [6]. Although the continuous approach is inherently an embedded method, its similarity to spectral numerical approximation methods allows the use of analysis techniques employed by the spectral methods community [37; 7].

Which of the two projection approaches described above (continuous vs. discrete projection) is preferred depends on the application and the type of model reduction approach sought (e.g., embedded vs. black-box). The discussion in the remainder of this paper is intended to aid the reader in selecting one of these approaches for his or her problem of interest.

Note that, regardless of which projection approach is used to build the ROM, the ROM dynamical system will have the form (13), as (11) has this form when written as a matrix problem. The solution to the ROM is obtained by advancing (13) forward in time using a time-integration scheme. Since the system considered here is linear, the projection terms in (11) are not time-dependent. Hence, these

terms can be pre-computed and stored in the offline stage of the model reduction – in particular, they need not be re-computed at each time step of the online time-integration stage of the ROM.

3. Stability definitions

As stated in Section 1, one of the objectives of this paper is to present and establish connections between some model reduction techniques that have an *a priori* stability guarantee. Before beginning this discussion, some general definitions of stability that will be used in subsequent analysis are reviewed.

3.1. Energy-stability

The concept of energy-stability originated in the literature involving the numerical analysis of spectral and finite difference discretizations to time-dependent PDEs [48; 8; 12]. It has also appeared in the Galerkin finite element method literature, e.g., [4; 2], where the energy-method was employed to derive stable Galerkin methods for hyperbolic conservation laws. It is well-known that physical systems admit a certain energy structure. The basic idea behind building energy-stable ROMs is that a ROM constructed for such systems should preserve this energy structure. Among the authors who have explored the concept of energy-stability in the context of model reduction are Rowley *et al.* [23] and Kwasniok [3]. In [23], Rowley *et al.* introduced a family of “energy-based” inner products for the purpose of constructing stable Galerkin ROMs for fluid problems. In [3], Kwasniok recognized the role of energy conservation in ROMs of nonlinear, incompressible fluid flow for atmospheric modeling applications, and proposed a Galerkin projection approach in which the ROM conserves turbulent kinetic energy or turbulent enstrophy.

The concept of energy-stability will be introduced in the context of a specific canonical model problem, then generalized. Consider, without loss of

generality, the following scalar initial value problem, known as a Cauchy problem [20]:

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathcal{L}(\mathbf{x}(t)), \quad t \geq 0 \\ \mathbf{x}(0) &= \mathbf{f}.\end{aligned}\tag{15}$$

Here, \mathcal{L} denotes a linear differential operator with constant coefficients (e.g., the linear operator in (1)), $\mathbf{f} \in \mathbb{R}^n$ is the initial condition, and $\mathbf{x}(t) \in \mathbb{R}^n$ is the system state at time $t \geq 0$. The operator \mathcal{L} is said to be semi-bounded with respect to an inner product (\cdot, \cdot) if it satisfies the following inequality for all sufficiently smooth functions $\mathbf{w} \in L^2$:

$$(\mathbf{w}, \mathcal{L}\mathbf{w}) \leq \alpha(\mathbf{w}, \mathbf{w}),\tag{16}$$

where $\alpha \in \mathbb{R}$. The following theorem (quoted from [20]) states the conditions under which the Cauchy problem (15) is well-posed.

Theorem 3.1.1 [20]: The Cauchy problem (15) is well-posed if and only if the operator \mathcal{L} is semi-bounded with respect to an inner product (\cdot, \cdot) which corresponds to a norm equivalent to the L^2 norm.

Consider now a Galerkin approximation to (15), denoted here by \mathbf{x}_N , and satisfying

$$(\dot{\mathbf{x}}_N, \boldsymbol{\phi}) = (\mathcal{L}(\mathbf{x}_N), \boldsymbol{\phi}),\tag{17}$$

for all $\boldsymbol{\phi}$ sufficiently smooth, and suppose \mathcal{L} is semi-bounded with respect to (\cdot, \cdot) . Setting $\boldsymbol{\phi} = \mathbf{x}_N$ in (17) leads to the following energy estimate for the Galerkin approximation:

$$\frac{dE_N}{dt} \leq 2\alpha E_N,\tag{18}$$

where $E_N \equiv \frac{1}{2}\|\mathbf{x}_N\|^2$ denotes the energy of the Galerkin approximation \mathbf{x}_N , and $\|\cdot\|$ is the norm induced by the inner product (\cdot, \cdot) . Applying Gronwall’s lemma ((71) in Appendix A.1) to (18) gives the inequality

$$\|\mathbf{x}_N(t)\| \leq e^{\frac{1}{2}\alpha t} \|\mathbf{x}_N(0)\|.\tag{19}$$

The result (19) says that the energy of the numerical solution to (17) is bounded in a way that is consistent with the behavior of the energy of the exact

solution to the original differential equation (15), i.e., the numerical solution is energy-stable. This definition can be extended to a ROM LTI system of the form (13).

Definition 3.1.2 (Energy-Stability [12]): A ROM LTI system (13) is called energy-stable if

$$E_M(t) \leq e^{\alpha t} E_M(0), \quad (20)$$

for some constant $\alpha \in \mathbb{R}$, where

$$E_M \equiv \frac{1}{2} \|\mathbf{x}_M\|^2 \quad (21)$$

is the system energy of the ROM numerical solution \mathbf{x}_M to (13), and $\|\cdot\|$ is a norm equivalent to the L^2 norm.

In general, a ROM LTI system (13) is not guaranteed to satisfy Definition 3.1.2 even if the PDE system (1) is well-posed and the full order LTI system arising from the discretization of these PDEs in space (2) is stable. However, it is often possible to ensure (20) holds for the ROM LTI system through a careful selection of the reduced trial and test bases Φ_M and Ψ_M and/or the inner product in which the projection step of the model reduction is performed (Sections 4 and 5).

3.2. Lyapunov, asymptotic and exponential stability

The concept of energy-stability can be related to classical notions of stability, namely Lyapunov stability, asymptotic stability and exponential stability. Consider an autonomous nonlinear dynamical system:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^n, \quad (22)$$

where $\mathbf{f} \in \mathbb{R}^n$ is a given function, subject to some initial condition $\mathbf{x}(0) = \mathbf{x}_0$. Let \mathbf{x}_e be an equilibrium point of the system (22), meaning $\mathbf{f}(\mathbf{x}_e) = \mathbf{0}$ for all $t \geq 0$.

Definition 3.2.1 (Lyapunov, asymptotic and exponential stability) [15]: The equilibrium point \mathbf{x}_e of (22) is said to be:

- (a) *Lyapunov stable* if $\forall \varepsilon > 0$ there exists a $\delta(\varepsilon) > 0$ such that if $\|\mathbf{x}(0) - \mathbf{x}_e\| < \delta$, then $\|\mathbf{x}(t) - \mathbf{x}_e\| < \varepsilon \forall t \geq 0$.
- (b) *Asymptotically stable* if there exists a $\delta > 0$ such that if $\|\mathbf{x}(0) - \mathbf{x}_e\| < \delta$, then $\lim_{t \rightarrow \infty} \|\mathbf{x}(t) - \mathbf{x}_e\| = 0$.
- (c) *Exponentially stable* if there exist $\alpha, \beta, \delta > 0$ such that if $\|\mathbf{x}(0) - \mathbf{x}_e\| < \delta$, then $\|\mathbf{x}(t) - \mathbf{x}_e\| \leq \alpha \|\mathbf{x}(0) - \mathbf{x}_e\| e^{-\beta t} \forall t \geq 0$.

In other words, if an equilibrium point of (22) is Lyapunov stable, solutions within a distance $\delta > 0$ from it will remain a distance $\varepsilon > 0$ from it for all time; if it is asymptotically stable, solutions within this distance will eventually converge to the equilibrium; if it is exponentially stable, the solutions will not only converge, but at an exponential rate. In general, exponential stability implies asymptotic stability, and asymptotic stability implies Lyapunov stability.

The following theorem, known as the Lyapunov stability theorem [15], can be used to characterize the stability of the stability of an equilibrium point \mathbf{x}_e for (22).

Theorem 3.2.2 (Lyapunov Stability Theorem) [15]: Let V be a non-negative function on \mathbb{R}^n and let \dot{V} represent the time derivative of V along trajectories of the system dynamics (22), i.e., $\dot{V} = \frac{\partial V}{\partial \mathbf{x}} \dot{\mathbf{x}} = \frac{\partial V}{\partial \mathbf{x}} \mathbf{f}(\mathbf{x})$. Let $B_r = B_r(\mathbf{x}_e)$ be a ball of radius r around an equilibrium point \mathbf{x}_e of (22). If there exists an $r > 0$ such that V is positive definite and \dot{V} is negative semi-definite for all $\mathbf{x} \in B_r$, then \mathbf{x}_e is Lyapunov stable.

The function V defined in Theorem 3.2.2 above is known as the Lyapunov function for the system (22). Observe that the numerical energy E_N defined in (18) satisfies the definition of a Lyapunov function (Theorem 3.2.2) if (23) holds. Thus, if an LTI ROM (2) is energy-stable with $\alpha = 0$ (Definition 3.1.2), then the ROM is Lyapunov stable. In Section 5, it is shown how Theorem 3.1.2 can be used to define a stability-preserving inner product for building stable ROMs for (2).

The stability concepts introduced above simplify for the specific case of LTI systems of the form (2). It is straightforward to verify that for linear systems, asymptotic and exponential stability are equivalent. Moreover, the following result holds.

Theorem 3.2.3 (Asymptotic Stability Theorem for LTI Systems) [15]: An LTI system (2) is asymptotically (and exponentially) stable if and only if all the eigenvalues of \mathbf{A} have strictly negative real parts.

Theorem 3.2.3 is commonly used to check numerically (*a posteriori*) the stability of an LTI system (2) or a ROM (13) constructed for an LTI system (Section 5.2).

3.3. Time-stability

Another form of stability is what is referred to herein as “time-stability”. Essentially, a system that is time-stable is one whose solution will not “blow up” (i.e., produce an unbounded output) given a finite input and/or non-zero initial condition. For a general nonlinear system, exponential stability implies time-stability, but time-stability is a stronger notion than asymptotic stability [54]. Since exponential and asymptotic stability are equivalent for LTI systems, asymptotic stability *does* imply time-stability in this special case.

The concept of time-stability can also be defined in terms of the system energy.

Definition 3.3.1 (Time-Stability [12]): A ROM LTI system (13) is called time-stable if the numerical energy of the ROM solution is non-increasing in time for an arbitrary time step, i.e., if

$$\frac{dE_N}{dt} \leq 0. \quad (23)$$

It is straightforward to demonstrate that a time-stable scheme is also energy-stable. Suppose an LTI ROM (13) is time-stable, so the ROM solution satisfies the energy estimate (23). Applying Gronwall’s lemma ((71) in Appendix A.1) to this

inequality, $E_N(t) \leq E_N(0)$. Thus, (20) holds with $\alpha = 0$.

In general, the converse of the above statement does not hold: energy-stability does not necessarily imply time-stability. This is to be expected. The practical implication of a ROM possessing the energy-stability property is that its numerical solution is bounded in a way that is consistent with the behavior of the exact solutions of the governing equations (1). It is possible in general that the governing PDEs support instabilities. In this case, an energy-stable ROM may possess unbounded solutions as $t \rightarrow \infty$, as (it can be argued) it should, if these unbounded solutions are physical.

4. Stable model reduction for LTI systems via continuous projection

In this section, an approach for building energy-stable ROMs via continuous Galerkin projection is developed for PDE systems of the form:

$$\dot{\mathbf{q}} + \mathbf{A}_i \frac{\partial \mathbf{q}}{\partial x_i} - \mathbf{K}_{ij} \frac{\partial^2 \mathbf{q}}{\partial x_i \partial x_j} + \mathbf{G} \mathbf{q} = \mathbf{f}. \quad (24)$$

In (24), $\mathbf{q} \in \mathbb{R}^n$ denotes a vector of unknowns, $\mathbf{f} \in \mathbb{R}^n$ is a source term, \mathbf{A}_i , \mathbf{K}_{ij} and \mathbf{G} are $n \times n$ matrices, where $1 \leq i, j \leq d$, with d denoting the number of spatial dimensions, and $n \in \mathbb{N}$. The matrices \mathbf{A}_i , \mathbf{K}_{ij} and \mathbf{G} could be a function of space, but they are assumed to be steady (not a function of time t). The so-called Einstein notation (implied summation on repeated indices) has been employed in (24) and subsequent expressions. Most conservation laws, as well as many PDEs of physical interest, can be written in the form (24). If $\mathbf{K}_{ij} = \mathbf{0} \forall i, j$, (24) is known as a hyperbolic system [14]. An example of a system of this form is the linearized compressible Euler system. A method for constructing energy-stable ROMs specifically for the compressible Euler system using continuous Galerkin projection was presented in [6; 7], and is extended to generic systems of the form (24) herein. Otherwise, if $\mathbf{K}_{ij} \neq \mathbf{0}$, (24) is known as an incompletely parabolic system [14]. A canonical

example of such a system is the linearized compressible Navier-Stokes system.

4.1. A stabilizing transformation

Suppose there exists a transformation

$$\begin{aligned} T : \mathbb{R}^n &\rightarrow \mathbb{R}^n, \\ \mathbf{q} &\rightarrow \mathbf{v}, \end{aligned} \quad (25)$$

such that in the new variables \mathbf{v} , the system (24) has the form

$$\dot{\mathbf{v}} + \mathbf{A}_i^S \frac{\partial \mathbf{v}}{\partial x_i} - \mathbf{K}_{ij}^S \frac{\partial^2 \mathbf{v}}{\partial x_i \partial x_j} + \mathbf{G}^S \mathbf{v} = \mathbf{f}^S, \quad (26)$$

where:

- *Property 1:* The matrices \mathbf{A}_i^S are symmetric for all $1 \leq i \leq d$.
- *Property 2:* The matrices \mathbf{K}_{ij}^S are symmetric for all $1 \leq i, j \leq d$.
- *Property 3:* The augmented viscosity matrix:

$$\mathbf{K}^S \equiv \begin{pmatrix} \mathbf{K}_{11}^S & \dots & \mathbf{K}_{1d}^S \\ \vdots & \ddots & \vdots \\ \mathbf{K}_{d1}^S & \dots & \mathbf{K}_{dd}^S \end{pmatrix} \quad (27)$$

is positive semi-definite.

Theorem 4.1.1: Suppose a ROM for (26) is constructed using continuous Galerkin projection in the $L^2(\Omega)$ inner product. Suppose the matrices in (26) satisfy *Properties 1–3* above. Suppose also that the reduced basis modes satisfy the boundary conditions of the full order system, or they are implemented weakly in the ROM in a stability-preserving way³. Let \mathbf{v}_M denote the ROM solution to (26). Then the ROM is energy-stable with energy estimate

$$\|\mathbf{v}_M(\cdot, T)\|_2 \leq e^{\frac{1}{2}\beta_S T} \|\mathbf{v}_M(\cdot, 0)\|_2, \quad (28)$$

³The reader is referred to [7] for a discussion of stability-preserving weak implementations of boundary conditions for ROMs constructed using the continuous projection approach. In general, a weak implementation of boundary conditions will be stability-preserving provided the boundary conditions are well-posed.

where β_S is an upper bound on the eigenvalues of the matrix

$$\mathbf{B}^S \equiv \frac{\partial \mathbf{A}_i^S}{\partial x_i} + \frac{\partial^2 \mathbf{K}_{ij}^S}{\partial x_i \partial x_j} - 2\mathbf{G}^S. \quad (29)$$

Moreover, this energy-stability result holds for *any* choice of reduced basis.

Proof. To prove energy-stability of a ROM constructed for (26), it is necessary to bound the energy of the ROM solution to (26) with $\mathbf{f}^S = \mathbf{0}$:

$$\begin{aligned} \frac{dE_M}{dt} &= \frac{1}{2} \frac{d}{dt} \|\mathbf{v}_M\|_2^2 \\ &= \frac{1}{2} \frac{d}{dt} (\mathbf{v}_M, \mathbf{v}_M) \\ &= \left(\mathbf{v}_M, \frac{\partial \mathbf{v}_M}{\partial t} \right) \\ &= \left(\mathbf{v}_M, -\mathbf{A}_i^S \frac{\partial \mathbf{v}_M}{\partial x_i} + \mathbf{K}_{ij}^S \frac{\partial^2 \mathbf{v}_M}{\partial x_i \partial x_j} - \mathbf{G}^S \mathbf{v}_M \right) \\ &= -\int_{\Omega} \mathbf{v}_M^T \mathbf{A}_i^S \frac{\partial \mathbf{v}_M}{\partial x_i} \partial \Omega + \int_{\Omega} \mathbf{v}_M^T \mathbf{K}_{ij}^S \frac{\partial^2 \mathbf{v}_M}{\partial x_i \partial x_j} \partial \Omega \\ &\quad - \int_{\Omega} \mathbf{v}_M^T \mathbf{G}^S \mathbf{v}_M \partial \Omega. \end{aligned} \quad (30)$$

Each of the terms in (30) will be bounded separately. First,

$$\begin{aligned} -\int_{\Omega} \mathbf{v}_M^T \mathbf{A}_i^S \frac{\partial \mathbf{v}_M}{\partial x_i} \partial \Omega &= -\frac{1}{2} \int_{\Omega} \frac{\partial}{\partial x_i} (\mathbf{v}_M^T \mathbf{A}_i^S \mathbf{v}_M) d\Omega \\ &\quad + \frac{1}{2} \int_{\Omega} \mathbf{v}_M^T \frac{\partial \mathbf{A}_i^S}{\partial x_i} \mathbf{v}_M d\Omega \\ &= -\frac{1}{2} \int_{\partial \Omega} \mathbf{v}_M^T \mathbf{A}_i^S n_i \mathbf{v}_M d\Gamma \\ &\quad + \frac{1}{2} \int_{\Omega} \mathbf{v}_M^T \frac{\partial \mathbf{A}_i^S}{\partial x_i} \mathbf{v}_M d\Omega. \end{aligned} \quad (31)$$

In (31), the property that each of the matrices \mathbf{A}_i^S is symmetric has been employed (*Property 1*). The symbol Γ has been used to denote the boundary of Ω , $\partial \Omega$.

Next, note that:

$$\mathbf{K}_{ij}^S \frac{\partial^2 \mathbf{v}_M}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i} \left(\mathbf{K}_{ij}^S \frac{\partial \mathbf{v}_M}{\partial x_j} \right) - \left(\frac{\partial \mathbf{K}_{ij}^S}{\partial x_i} \frac{\partial \mathbf{v}_M}{\partial x_j} \right). \quad (32)$$

Then,

$$\begin{aligned} \int_{\Omega} \mathbf{v}_M^T \mathbf{K}_{ij}^S \frac{\partial^2 \mathbf{v}_M}{\partial x_i \partial x_j} \partial \Omega &= \int_{\Omega} \mathbf{v}_M^T \frac{\partial}{\partial x_i} \left(\mathbf{K}_{ij}^S \frac{\partial \mathbf{v}_M}{\partial x_j} \right) d\Omega \\ &\quad - \int_{\Omega} \mathbf{v}_M^T \frac{\partial \mathbf{K}_{ij}^S}{\partial x_i} \frac{\partial \mathbf{v}_M}{\partial x_j} \partial \Omega. \end{aligned} \quad (33)$$

Again, each of the two terms in (33) will be bounded separately.

$$\begin{aligned} \int_{\Omega} \mathbf{v}_M^T \frac{\partial}{\partial x_i} \left(\mathbf{K}_{ij}^S \frac{\partial \mathbf{v}_M}{\partial x_j} \right) d\Omega &= -\int_{\Omega} \frac{\partial \mathbf{v}_M^T}{\partial x_i} \mathbf{K}_{ij}^S \frac{\partial \mathbf{v}_M}{\partial x_j} d\Omega \\ &\quad + \int_{\partial \Omega} \mathbf{v}_M^T \mathbf{K}_{ij}^S \frac{\partial \mathbf{v}_M}{\partial x_j} n_i d\Gamma \\ &\leq \int_{\partial \Omega} \mathbf{v}_M^T \mathbf{K}_{ij}^S \frac{\partial \mathbf{v}_M}{\partial x_j} n_i d\Gamma, \end{aligned} \quad (34)$$

provided the matrix (27) is positive semi-definite (*Property 3*).

Now for the second term in (33):

$$\begin{aligned}
-\int_{\Omega} \mathbf{v}_M^T \frac{\partial \mathbf{K}_{ij}^S}{\partial x_i} \frac{\partial \mathbf{v}_M}{\partial x_j} \partial \Omega &= -\frac{1}{2} \int_{\Omega} \frac{\partial}{\partial x_j} \left(\mathbf{v}_M^T \frac{\partial \mathbf{K}_{ij}^S}{\partial x_i} \mathbf{v}_M \right) d\Omega \\
&+ \frac{1}{2} \int_{\Omega} \mathbf{v}_M^T \frac{\partial^2 \mathbf{K}_{ij}^S}{\partial x_i \partial x_j} \mathbf{v}_M d\Omega \\
&= -\frac{1}{2} \int_{\partial \Omega} \mathbf{v}_M^T \frac{\partial \mathbf{K}_{ij}^S}{\partial x_i} n_j \mathbf{v}_M d\Gamma \\
&+ \frac{1}{2} \int_{\Omega} \mathbf{v}_M^T \frac{\partial^2 \mathbf{K}_{ij}^S}{\partial x_i \partial x_j} \mathbf{v}_M d\Omega.
\end{aligned} \tag{35}$$

In (35), the property that the \mathbf{K}_{ij}^S matrices and therefore their derivatives are symmetric has been employed (*Property 2*).

Finally, (31) and (33) are substituted into (30). The boundary integral terms may be neglected if the reduced basis modes satisfy the boundary conditions or the boundary conditions have been implemented in a stability-preserving way. The following bound is obtained:

$$\begin{aligned}
\frac{1}{2} \frac{d}{dt} \|\mathbf{v}_M\|_2^2 &\leq \frac{1}{2} \int_{\Omega} \mathbf{v}_M^T \left(\frac{\partial \mathbf{A}_i^S}{\partial x_i} \right) \mathbf{v}_M d\Omega \\
&+ \frac{1}{2} \int_{\Omega} \mathbf{v}_M^T \frac{\partial^2 \mathbf{K}_{ij}^S}{\partial x_i \partial x_j} \mathbf{v}_M d\Omega \\
&- \int_{\Omega} \mathbf{v}_M^T \mathbf{G}^S \mathbf{v}_M \partial \Omega \\
&= \frac{1}{2} \int_{\Omega} \mathbf{v}_M^T \mathbf{B}^S \mathbf{v}_M d\Omega,
\end{aligned} \tag{36}$$

where \mathbf{B}^S is given by (29). Applying Gronwall's inequality ((71) in Appendix A.1) to (36), it is found that:

$$\|\mathbf{v}_M(\cdot, T)\|_2 \leq e^{\frac{1}{2} \beta_S T} \|\mathbf{v}_M(\cdot, 0)\|_2, \tag{37}$$

where β_S is an upper bound on the eigenvalues of the matrix \mathbf{B}^S (29). □

The proof of Theorem 4.1.1 is one of the new contributions of this article.

Note that, if $\mathbf{G} = \mathbf{0}$ in (24) and the \mathbf{A}_i and \mathbf{K}_{ij} matrices are spatially-constant, it follows that $\beta_S = 0$ in (37). In this case, if the ROM for (24) is constructed in the variables \mathbf{v} , the ROM will be time-stable as well as stable in the sense of Lyapunov, in addition to being energy-stable. For linearized conservation laws (e.g., the linearized shallow water

equations, the linearized compressible Euler equations, the linearized compressible Navier-Stokes equations), the property that $\mathbf{G} = \mathbf{0}$ and the \mathbf{A}_i and \mathbf{K}_{ij} are spatially-constant will in general hold if the base flow is spatially uniform.

4.2. Stability-preserving “symmetry inner product” and Petrov-Galerkin connection

A key property of systems of the form (24) is that they are symmetrizable [8; 6; 7]; that is, it is possible to derive a symmetric positive-definite matrix \mathbf{H} such that:

- *Property 1**: The matrices $\mathbf{H}\mathbf{A}_i$ are symmetric for all $1 \leq i \leq d$.
- *Property 2**: The matrices $\mathbf{H}\mathbf{K}_{ij}$ are symmetric for all $1 \leq i, j \leq d$.
- *Property 3**: The augmented viscosity matrix:

$$\mathbf{K}^H \equiv \begin{pmatrix} \mathbf{H}\mathbf{K}_{11} & \dots & \mathbf{H}\mathbf{K}_{1d} \\ \vdots & \ddots & \vdots \\ \mathbf{H}\mathbf{K}_{d1} & \dots & \mathbf{H}\mathbf{K}_{dd} \end{pmatrix} \tag{38}$$

is positive semi-definite.

Since \mathbf{H} is symmetric positive-definite, the following defines a valid inner product:

$$\left(\mathbf{q}^{(1)}, \mathbf{q}^{(2)} \right)_{(\mathbf{H}, \Omega)} \equiv \int_{\Omega} \mathbf{q}^{(1)T} \mathbf{H} \mathbf{q}^{(2)} d\Omega. \tag{39}$$

Following the terminology introduced in [6; 7], the inner product (39) will be referred to as the “symmetry inner product”. It is straightforward to see that the following corollary to Theorem 4.1.1 holds.

Corollary 4.2.1: Suppose a ROM for (24) is constructed using continuous Galerkin projection in the symmetry inner product (39). Suppose *Properties 1*-3** hold. Suppose also, as in Theorem 4.1.1, that the reduced basis modes satisfy the boundary conditions of the full order system, or they are implemented weakly in the ROM in a stability-preserving way. Let \mathbf{q}_M denote the ROM solution

to (24). Then the ROM is energy-stable with energy estimate

$$\|\mathbf{q}_M(\cdot, T)\|_{(\mathbf{H}, \Omega)} \leq e^{\frac{1}{2}\beta_H T} \|\mathbf{q}_M(\cdot, 0)\|_{(\mathbf{H}, \Omega)}, \quad (40)$$

where β_H is an upper bound on the eigenvalues of the matrix

$$\mathbf{B}^H \equiv \frac{\partial(\mathbf{H}\mathbf{A}_i)}{\partial x_i} + \frac{\partial^2(\mathbf{H}\mathbf{K}_{ij})}{\partial x_i \partial x_j} - 2\mathbf{H}\mathbf{G}. \quad (41)$$

Moreover, this energy-stability result holds for *any* choice of reduced basis.

Proof. Because of simple linear transformations, the proof is analogous to the proof of Theorem 4.1.1. \square

Again, in the case that $\mathbf{G} = \mathbf{0}$ and the \mathbf{A}_i and \mathbf{K}_{ij} matrices are spatially-constant, it will follow from Corollary 4.2.1 that a ROM constructed in the symmetry inner product (39) will be time-stable and stable in the sense of Lyapunov, in addition to being energy-stable.

It is interesting to observe that a Galerkin projection of the governing (24) in the symmetry inner product (39) is equivalent to a Petrov-Galerkin projection. Let ϕ_i for $i = 1, \dots, M$ denote the reduced trial basis vector for the solution \mathbf{q} . Performing a Galerkin projection of the equations (24) onto the modes ϕ_k gives

$$\int_{\Omega} \phi_k^T \mathbf{H} \left(\dot{\mathbf{q}} + \mathbf{A}_i \frac{\partial \mathbf{q}}{\partial x_i} + \mathbf{K}_{ij} \frac{\partial^2 \mathbf{q}}{\partial x_i \partial x_j} + \mathbf{G}\mathbf{q} \right) d\Omega = \int_{\Omega} \phi_k^T \mathbf{H} \mathbf{f} d\Omega, \quad (42)$$

for $k = 1, \dots, M$. Equation (42) is equivalent to a Petrov-Galerkin projection of the equations (24) in the regular L^2 inner product

$$\int_{\Omega} \psi_k^T \left(\dot{\mathbf{q}} + \mathbf{A}_i \frac{\partial \mathbf{q}}{\partial x_i} + \mathbf{K}_{ij} \frac{\partial^2 \mathbf{q}}{\partial x_i \partial x_j} + \mathbf{G}\mathbf{q} \right) d\Omega = \int_{\Omega} \psi_k^T \mathbf{f} d\Omega, \quad (43)$$

where the reduced test basis functions are given by $\psi_k = \mathbf{H}\phi_k$, for all $k = 1, \dots, M$.

4.3. Examples of stability-preserving transformation and symmetry inner product for several physical systems

It is straightforward to derive the matrix \mathbf{H} that defines the symmetry inner product (39) for many problems of physical interest. This matrix has been derived herein by the authors for several hyperbolic and incompletely parabolic systems (the wave equation, the linearized shallow water equations, the linearized compressible Euler equations, and the linearized compressible Navier-Stokes equations), and is given below.

Example 1: Wave Equation

Consider the one-dimensional (1D) wave equation:

$$\ddot{u} = a^2 \frac{\partial^2 u}{\partial x^2}, \quad (44)$$

where $a \in \mathbb{R}$ denotes the wave speed, and $\ddot{u} \equiv \frac{\partial^2 u}{\partial t^2}$. (44) is a canonical PDE of the hyperbolic type. This equation can be written as a first order system

$$\dot{\mathbf{q}} = \mathbf{A} \frac{\partial \mathbf{q}}{\partial x}, \quad (45)$$

where

$$\mathbf{q} = \begin{pmatrix} \dot{u} \\ \frac{\partial u}{\partial x} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 0 & a^2 \\ 1 & 0 \end{pmatrix}. \quad (46)$$

Remark that if

$$\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & a^2 \end{pmatrix}, \quad (47)$$

the matrix $\mathbf{H}\mathbf{A}$ is symmetric [32].

Example 2: Linearized Shallow Water Equations

Consider the linearized form of the shallow water equations:

$$\dot{\mathbf{q}}' + \mathbf{A}_i \frac{\partial \mathbf{q}'}{\partial x_i} + \mathbf{G}\mathbf{q}' = \mathbf{0}. \quad (48)$$

These equations are obtained from the full (non-linear) shallow water equations by decomposing

the fluid vector $\mathbf{q}(\mathbf{x}, t)$ into a steady mean plus an unsteady fluctuation, i.e.,

$$\mathbf{q}(\mathbf{x}, t) = \bar{\mathbf{q}}(\mathbf{x}) + \mathbf{q}'(\mathbf{x}, t), \quad (49)$$

and linearizing the full shallow water equations around the steady mean state $\bar{\mathbf{q}}$. If $\mathbf{q}^T = (u, v, w, \phi)$, then the convective flux matrices in the hyperbolic system (48) in three-dimensions (3D) are given by:

$$\mathbf{A}_1 = \begin{pmatrix} \bar{u} & 0 & 0 & 1 \\ 0 & \bar{u} & 0 & 0 \\ 0 & 0 & \bar{u} & 0 \\ \bar{\phi} & 0 & 0 & \bar{u} \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} \bar{v} & 0 & 0 & 0 \\ 0 & \bar{v} & 0 & 1 \\ 0 & 0 & \bar{v} & 0 \\ 0 & \bar{\phi} & 0 & \bar{v} \end{pmatrix},$$

$$\mathbf{A}_3 = \begin{pmatrix} \bar{w} & 0 & 0 & 0 \\ 0 & \bar{w} & 0 & 0 \\ 0 & 0 & \bar{w} & 1 \\ 0 & 0 & \bar{\phi} & \bar{w} \end{pmatrix}, \quad (50)$$

where ϕ denotes the local height of the fluid above the equilibrium depth, and $u, v,$ and w are the components of the fluid velocity vector [32]. Each of the convective flux matrices (50) can be symmetrized by the matrix

$$\mathbf{H} = \begin{pmatrix} \bar{\phi} & 0 & 0 & 0 \\ 0 & \bar{\phi} & 0 & 0 \\ 0 & 0 & \bar{\phi} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (51)$$

Example 3: Linearized Compressible Euler Equations

Consider the linearized compressible Euler equations. These equations may be used if a compressible fluid system can be described by inviscid, small-amplitude perturbations about a steady-state mean flow. The equations are obtained from the full (non-linear) compressible Euler equations by decomposing the fluid vector $\mathbf{q}(\mathbf{x}, t)$ into a steady mean plus an unsteady fluctuation (49) and linearizing these equations around the steady mean state $\bar{\mathbf{q}}$. If $\mathbf{q}^T = (u, v, w, \zeta, p)$, where u, v and w are the three components of the velocity vector, ζ is the specific volume (the reciprocal of the density), and p is the pressure, the linearized

compressible Euler equations take the form (48). In 3D, the convective flux matrices \mathbf{A}_i in the linearized compressible Euler hyperbolic system (48) are given by:

$$\mathbf{A}_1 = \begin{pmatrix} \bar{u} & 0 & 0 & 0 & \bar{\zeta} \\ 0 & \bar{u} & 0 & 0 & 0 \\ 0 & 0 & \bar{u} & 0 & 0 \\ -\bar{\zeta} & 0 & 0 & \bar{u} & 0 \\ \gamma\bar{p} & 0 & 0 & 0 & \bar{u} \end{pmatrix},$$

$$\mathbf{A}_2 = \begin{pmatrix} \bar{v} & 0 & 0 & 0 & 0 \\ 0 & \bar{v} & 0 & 0 & \bar{\zeta} \\ 0 & 0 & \bar{v} & 0 & 0 \\ 0 & -\bar{\zeta} & 0 & \bar{v} & 0 \\ 0 & \gamma\bar{p} & 0 & 0 & \bar{v} \end{pmatrix}, \quad (52)$$

$$\mathbf{A}_3 = \begin{pmatrix} \bar{w} & 0 & 0 & 0 & 0 \\ 0 & \bar{w} & 0 & 0 & 0 \\ 0 & 0 & \bar{w} & 0 & \bar{\zeta} \\ 0 & 0 & -\bar{\zeta} & \bar{w} & 0 \\ 0 & 0 & \gamma\bar{p} & 0 & \bar{w} \end{pmatrix}.$$

Here, $\gamma = C_p/C_v$ is the ratio of specific heats. The reader may verify that if the linearized compressible Euler system (48) is pre-multiplied by the following symmetric positive definite matrix:

$$\mathbf{H} = \begin{pmatrix} \bar{\rho} & 0 & 0 & 0 & 0 \\ 0 & \bar{\rho} & 0 & 0 & 0 \\ 0 & 0 & \bar{\rho} & 0 & 0 \\ 0 & 0 & 0 & \alpha^2 \gamma \bar{\rho}^2 \bar{p} & \bar{\rho} \alpha^2 \\ 0 & 0 & 0 & \bar{\rho} \alpha^2 & \frac{(1+\alpha^2)}{\gamma \bar{\rho}} \end{pmatrix}, \quad (53)$$

where α is a real, non-zero parameter to yield the system, the convective flux matrices $\mathbf{H}\mathbf{A}_i$ are all symmetric [6; 7].

Example 4: Linearized Compressible Navier-Stokes Equations

Consider the 3D linearized compressible Navier-Stokes equations. These equations are appropriate when a compressible fluid system can be described by viscous, small-amplitude perturbations about a steady-state base flow. As with the linearized shallow water equations and linearized compressible Euler equations, to derive these equations from the full (non-linear) compressible Navier-Stokes

equations, the fluid vector $\mathbf{q}(\mathbf{x}, t)$ is written as the sum of a steady mean plus an unsteady fluctuation (49), and a linearization around the steady mean is performed. If the viscous work terms are neglected from the equations⁴ (appropriate, for example, in a low Mach number regime), the result is a linear incompletely parabolic system of the form (24). If the fluid vector is given by $\mathbf{q}^T = (u, v, w, T, \rho)$, where T and ρ denote the fluid temperature and density respectively, the convective and viscous flux matrices that appear in (24) are given by the expressions found in [8]. The reader can verify that if the system (24) is pre-multiplied by the symmetric positive definite matrix given by

$$\mathbf{H} \equiv \begin{pmatrix} \bar{\rho} & 0 & 0 & 0 & 0 \\ 0 & \bar{\rho} & 0 & 0 & 0 \\ 0 & 0 & \rho & 0 & 0 \\ 0 & 0 & 0 & \frac{\bar{\rho}R}{T(\gamma-1)} & 0 \\ 0 & 0 & 0 & 0 & \frac{RT}{\bar{p}} \end{pmatrix}, \quad (54)$$

the ‘‘symmetrized’’ convective flux matrices $\mathbf{H}\mathbf{A}_i$ and diffusive flux matrices $\mathbf{H}\mathbf{K}_{ij}$ satisfy *Properties I*–3** in Section 4.2. Here, R denotes the universal gas constant.

Note that the symmetry transformations in the examples above are not unique. For example, in [9], Abarbanel *et al.* exhibit a transformation of the form (26) for the linearized compressible Navier-Stokes equations written in the primitive variables $\mathbf{q}^T = (\rho, u, v, w, p)$.

4.4. Numerical experiments

The stability-preserving model reduction approach based on continuous projection described in Sec-

⁴To the authors’ knowledge, the viscous work terms are invariably neglected from the linearized compressible Navier-Stokes equations by researching studying energy-stability of these equations [8; 9]. The omission of these terms is justified only in the low Mach number regime, or in the case that the base flow is uniform. The extension of the energy-stability symmetrization approach presented here to the linearized compressible Navier-Stokes equations in which the viscous work terms are retained is the subject of present research.

tions 4.1–4.2 is now evaluated numerically on a test case involving a 2D inviscid acoustic pressure pulse in a 2D prismatic domain. The governing equations are the equations of linearized compressible flow, given in Section 4.3 (Example 3) above. Prior to showing these results, a stability-preserving discrete implementation of the projection step of the model reduction is outlined.

4.4.1. Stability-preserving discrete implementation

The stability analysis of Sections 4.1–4.2 has assumed that the integrals resulting from the projection of the governing equations onto the reduced basis modes are evaluated exactly in continuous form. This continuous result can be translated to the discrete setting through the use of high-precision numerical quadrature as follows. First, the snapshots and the POD basis modes are cast as a collection of continuous finite elements. It is then possible to construct a numerical quadrature operator that computes exactly (with respect to the finite element representation) all continuous inner products arising from the continuous Galerkin projection of the equations onto the POD modes. Suppose the domain Ω is broken up into n_{el} finite elements Ω_e such that $\cup_{e=1}^{n_{el}} \Omega_e = \Omega$. Suppose each of these elements have mn nodes. Then, the finite element representation of the vector \mathbf{q} in (24) in each element Ω_e is:

$$\mathbf{q}_e^h = \sum_{i=1}^{mn} N_i(\mathbf{x}) \mathbf{q}_i(\mathbf{x}), \quad \mathbf{x} \in \Omega_e. \quad (55)$$

By the discussion in Section 4.2, it is necessary to compute numerically integrals of the form:

$$\left(\mathbf{q}^{(1)}, \mathbf{q}^{(2)} \right)_{(\mathbf{H}, \Omega)} = \int_{\Omega} \mathbf{q}^{(1)T} \mathbf{H} \mathbf{q}^{(2)} d\Omega. \quad (56)$$

Suppose, without loss of generality, that the finite element shape functions are chosen to be bilinear, so $mn = 4$. The discrete representations of the vectors $\mathbf{q}^{(1)}$ and $\mathbf{q}^{(2)}$ are denoted by $\mathbf{q}^{h(1)}$ and $\mathbf{q}^{h(2)}$, respectively. The length of these vectors is equal to the number of mesh nodes N times the dimension

of the vector, r . Let \mathbf{H}_e^h be the $r \times r$ element inner product matrix, taken to be piecewise constant over each element. Then, the formula for numerical integration of (56) can be written as

$$\left(\mathbf{q}^{(1)}, \mathbf{q}^{(2)} \right)_{(\mathbf{H}, \Omega)} = \mathbf{q}^{h(1)T} \mathbf{W} \mathbf{q}^{h(2)}, \quad (57)$$

where \mathbf{W} is a sparse block matrix comprised of $N \times N$ blocks of dimension $r \times r$. The $(k, l)^{th}$ block of this matrix given by $w_{kl} \mathbf{I}$, where

$$w_{kl} = \sum_{e=1}^{n_{kl}^{el}} \mathbf{H}_e^h \sum_{j=1}^4 N_{k_e}(\mathbf{x}_{j_e}) N_{l_e}(\mathbf{x}_{j_e}) \omega_{j_e}. \quad (58)$$

Here, the outer sum is over the elements connected to the $k-l$ nodal ‘‘edge’’; the ω_{j_e} are the integration weights and the \mathbf{x}_{j_e} are the integration points.

A parallel C++ code that reads in the snapshot data written by a high-fidelity code, assembles the necessary finite element representation of the snapshots and computes the numerical quadrature necessary for evaluation of the inner products has been written by the authors. The code, known as `Spirit`, performs all the calculations in parallel using distributed matrix and vector data structures and parallel eigensolvers from the Trilinos project [49], and uses the `libmesh` finite element library [50] to compute element quadratures. The parallelism in `Spirit` allows for large data sets and a relatively large number of POD modes. The `libmesh` finite element library [50] was used to compute element quadratures. The online time-integration of the ROM system (2) (with the ROM coefficient matrix computed within `Spirit` and written to disk) is then performed using a fourth-order Runge-Kutta scheme in `MATLAB`.

4.4.2. 2D inviscid acoustic pulse example

For the sake of brevity, the proposed model reduction approach is evaluated on only one of the physics sets given in Section 4.3. The test case considered is that of a 2D inviscid acoustic pressure pulse in the following 2D prismatic domain: $\Omega = (-1, 1) \times (-1, 1) \in \mathbb{R}^2$. The governing equations

are the linearized compressible Euler equations (Example 3 in Section 4.3). The base flow is uniform, with the following values: $\bar{p} = 101,325$ Pa, $\bar{T} = 300$ K, $\bar{\rho} = \frac{\bar{p}}{RT} = 1.17$ kg/m³, $\bar{u}_1 = \bar{u}_2 = 0.0$ m/s, and $\bar{c} = 348.0$ m/s, where $\bar{c} \equiv \sqrt{\gamma R \bar{T}}$ is the mean speed of sound. The problem is initialized with a pressure pulse in the middle of the domain:

$$\begin{aligned} p'(\mathbf{x}; 0) &= 141.9 e^{-10(x^2+y^2)}, \\ \rho'(\mathbf{x}; 0) &= \frac{p'(\mathbf{x}; 0)}{RT}, \\ T'(\mathbf{x}; 0) &= 0, \\ u'_1(\mathbf{x}; 0) &= u'_2(\mathbf{x}; 0) = 0. \end{aligned} \quad (59)$$

In terms of the mean values, the amplitude of the initial pressure pulse (59) is $0.001 \bar{p} \bar{c}^2$.

For the problem considered, the high-fidelity fluid simulation data were generated using a Sandia in-house finite volume flow solver known as `SIGMA CFD`. This code is derived from `LESLIE3D` [51], a Large Eddy Simulations (LES) flow solver originally developed in the Computational Combustion Laboratory at the Georgia Institute of Technology. For a detailed description of the schemes and models implemented within `LESLIE3D`, the reader is referred to [52; 53].

As both the high-fidelity code as well as the ROM code are 3D codes, a 2D mesh of the domain Ω is converted to a 3D mesh by extruding the 2D mesh in the z -direction by one element. The computational grid for this test case is composed of 3362 nodes, cast into 9600 tetrahedral finite elements within the ROM code. A no-penetration (slip wall) boundary condition is imposed on the four sides of the domain in the x and y plane. To ensure the solution has no dynamics in the z -direction, the following values of the z -velocity component are specified: $\bar{u}_3 = 0$, $u'_3(\mathbf{x}; 0) = 0$. Symmetry boundary conditions are imposed for $z = \text{constant}$ in the high-fidelity code. The high-fidelity computational fluid dynamics (CFD) simulation from which the ROM is generated is performed until time $T = 0.01$ seconds. During this simulation, the initial pressure pulse (59) reflected from the walls of the domain

a number of times. Snapshots from this simulation were saved every 5×10^{-5} seconds, to yield a total of 200 snapshots. These snapshots were used to construct 20 mode POD bases. Two different procedures were used to generate a fluid ROM for this problem: the POD/Galerkin method with the symmetry inner product (39) with \mathbf{H} given by (53), and the POD/Galerkin method with the classical L^2 inner product. The size of the POD basis was determined using an energy criterion (6) (see Section 2.1): M was selected such that the modes capture 99.9% of the snapshot energy. Since the base flow for this example is uniform, $\mathbf{G} = \mathbf{0}$ and \mathbf{A}_i and \mathbf{K}_{ij} are spatially-constant in (24), meaning an energy-stable ROM is expected to be time-stable and stable in the sense of Lyapunov. Figure 1 shows a time history of the first two ROM modal amplitudes (circles) compared to the projection of the FOM CFD simulation onto the first two POD modes (solid lines) for the symmetry (a) and L^2 (b) ROMs. Mathematically, this figure compares as a function of time t :

$$x_{M,i}(t) \quad \text{vs.} \quad (\mathbf{q}'_{\text{FOM}}, \phi_i)_{(\mathbf{H}, \Omega)}, \quad (60)$$

for $i = 1, 2$, where \mathbf{q}'_{FOM} is the high-fidelity CFD solution from which the ROMs were constructed. The reader may observe reasonable agreement between the symmetry ROM and the full simulation (Figure 1(a)) for the time interval considered. In contrast, agreement between the L^2 ROM and the full simulation is reasonable only until approximately $t = 0.005$ seconds (Figure 1(b)). The oscillations in the L^2 ROM modal amplitudes observed for $t > 0.008$ seconds suggest the presence of an instability in the L^2 ROM. If the modal amplitudes $x_{M,i}(t)$ are plotted up to a longer time horizon (Figure 2), the instability in the L^2 ROM is apparent.

Figures 3–4 compare the FOM pressure field (a) with the field reconstructed from the symmetry (b) and L^2 (c) ROM solutions at times $t = 4.5 \times 10^{-4}$ and 7.95×10^{-3} seconds. At time $t = 4.5 \times 10^{-4}$ seconds, both the symmetry and L^2 ROM solutions are in good agreement with the high-fidelity solution (Figure 3). At the later time, 7.95×10^{-3} seconds, there is a good qualitative agreement be-

tween the high-fidelity solution and the symmetry ROM solution (Figure 4(a), (b)). The same cannot be said of the L^2 ROM solution, however. It is apparent from Figure 4(c) that the L^2 ROM solution has blown up by $t = 7.95 \times 10^{-3}$ seconds, which confirms the instability of the 20 mode L^2 ROM suggested in Figures 1–2.

5. Stable model reduction for LTI systems via discrete projection

In Section 4, a method for constructing energy-stable ROMs via continuous projection of a linear system of PDEs was presented. The discussion in Section 4 motivates the following question: can the energy inner product be determined in a black-box fashion for any given full order model system? It is shown in the present section that there is a discrete counterpart of the symmetry inner product, first derived by Rowley *et al.* [24] and termed the “Lyapunov inner product” herein. Although the Lyapunov inner product has appeared in several publications [24; 1; 30], to the authors’ knowledge, a numerical study of the properties and performance of POD ROMs constructed in the Lyapunov inner product is lacking from the literature at the present time, and one of the contributions of this work.

5.1. Stability-preserving Lyapunov inner product and Petrov-Galerkin connection

Suppose the LTI system (2) is stable in the sense of Lyapunov, i.e., all eigenvalues of the matrix \mathbf{A} have non-positive real parts (Corollary 3.4.2). Since \mathbf{A} is stable, there exists a Lyapunov function for

$$\dot{\mathbf{x}}_N(t) = \mathbf{A}\mathbf{x}_N(t). \quad (61)$$

In particular,

$$V(\mathbf{x}_N) = \mathbf{x}_N^T \mathbf{P} \mathbf{x}_N, \quad (62)$$

is a Lyapunov function for (61), where \mathbf{P} is the solution of the following Lyapunov equation:

$$\mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} = -\mathbf{Q}. \quad (63)$$

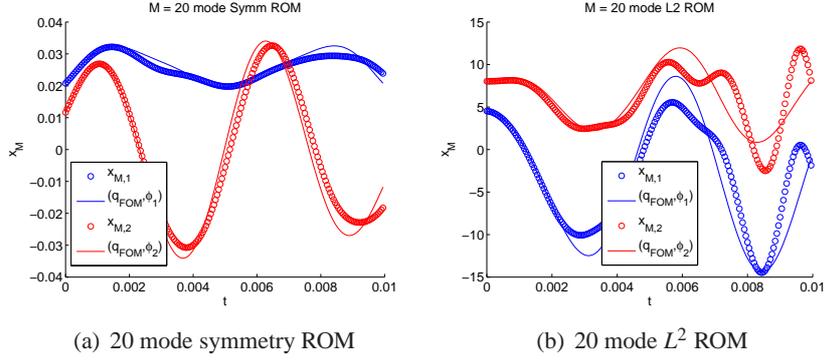


Figure 1: Time history of modal amplitudes for inviscid pressure pulse problem

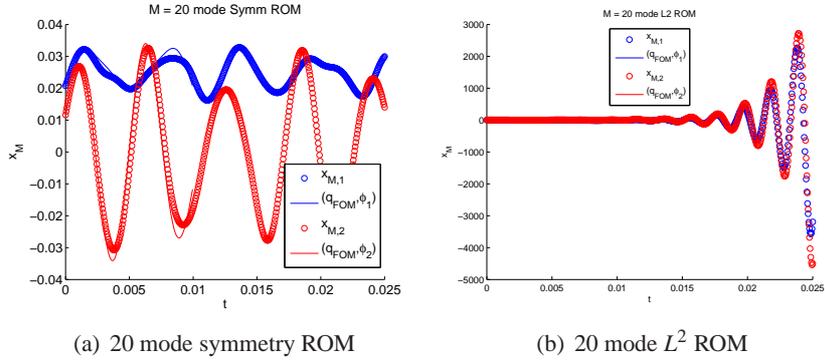


Figure 2: Time history of modal amplitudes for inviscid pressure pulse problem for longer time horizon

Here, \mathbf{Q} is some positive-definite matrix [15]. A positive definite solution \mathbf{P} to (63) exists provided \mathbf{A} is stable. Moreover, if \mathbf{Q} is symmetric, \mathbf{P} is symmetric as well. Given \mathbf{A} and \mathbf{Q} , a solution to the Lyapunov equation (63) can be obtained, for instance, using the `lyap` function in the MATLAB control toolbox [44]:

$$\mathbf{P} = \text{lyap}(\mathbf{A}', \mathbf{Q}, [\], \text{speye}(N, N)).$$

Assume the system (61) is stable and a positive-definite symmetric \mathbf{P} has been computed from (63). Since \mathbf{P} is symmetric positive-definite, the following

$$\left(\mathbf{x}_N^{(1)}, \mathbf{x}_N^{(2)} \right)_{\mathbf{P}} \equiv \mathbf{x}_N^{(1)T} \mathbf{P} \mathbf{x}_N^{(2)}, \quad (64)$$

defines an inner product. Let Φ_M be a reduced basis of size M , so that

$$\mathbf{x}_N(t) \approx \Phi_M \mathbf{x}_M(t), \quad (65)$$

where $\mathbf{x}_M(t)$ denotes the ROM solution. Theorem 5.1.1 (summarized here from [24] to keep this work self-contained) shows that (64) is the energy inner product for this system.

Theorem 5.1.1 (from [24]): Assume the linear full order system (61) is stable. Suppose a ROM for (61) is constructed via a Galerkin projection in the $(\cdot, \cdot)_{\mathbf{P}}$ inner product (64), to yield the following reduced linear system:

$$\dot{\mathbf{x}}_M = \Phi_M^T \mathbf{P} \mathbf{A} \Phi_M \mathbf{x}_M, \quad (66)$$

where it has been assumed that the basis Φ_M has been constructed to be orthonormal in the $(\cdot, \cdot)_{\mathbf{P}}$ inner product, i.e., $\Phi_M^T \mathbf{P} \Phi_M = \mathbf{I}_M$ where \mathbf{I}_M denotes the $M \times M$ identity matrix. Then, the ROM (66) is energy-stable, time-stable and stable in the sense of Lyapunov.

Proof. It is shown that the energy $E_M \equiv \frac{1}{2} \|\mathbf{x}_M\|_2^2$

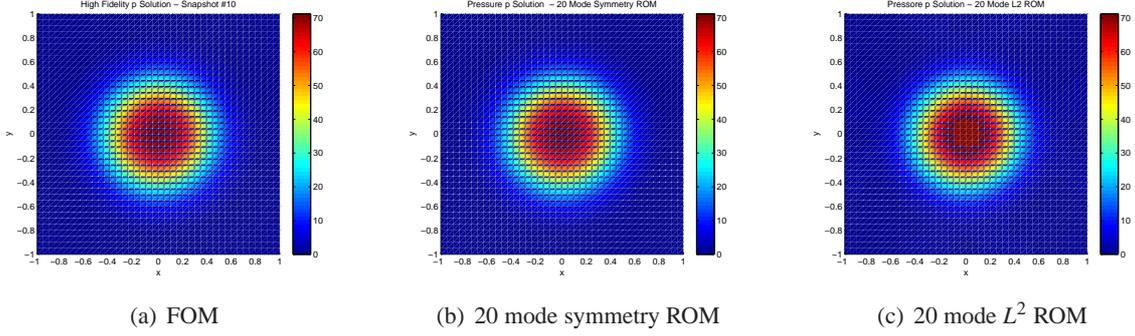


Figure 3: Pressure field at time $t = 4.5 \times 10^{-4}$ seconds

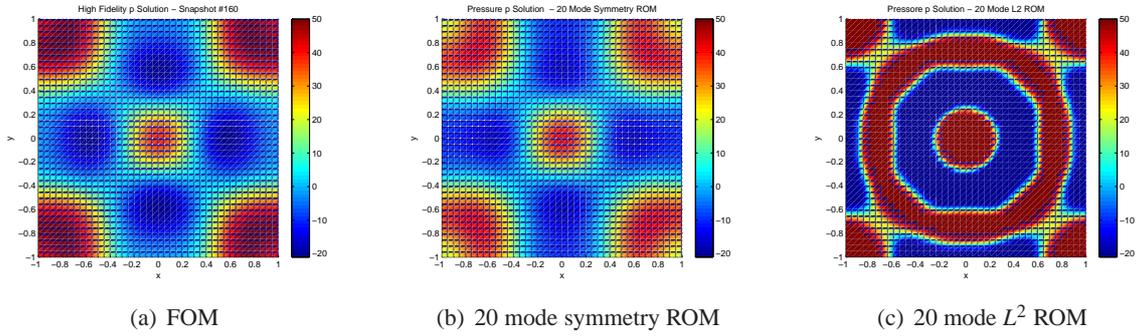


Figure 4: Pressure field at time $t = 7.95 \times 10^{-3}$ seconds

of the ROM system (66) is non-increasing:

$$\begin{aligned}
\frac{dE_M}{dt} &= \frac{1}{2} \frac{d}{dt} (\mathbf{x}_M, \mathbf{x}_M)_2 \\
&= \mathbf{x}_M^T \dot{\mathbf{x}}_M \\
&= \mathbf{x}_M^T \Phi_M^T \mathbf{P} \mathbf{A} \Phi_M \mathbf{x}_M \\
&= \mathbf{x}_M^T \Phi_M^T \left(\frac{1}{2} \mathbf{P} \mathbf{A} + \frac{1}{2} \mathbf{P}^T \mathbf{A} \right) \Phi_M \mathbf{x}_M \quad (67) \\
&= \mathbf{x}_M^T \Phi_M^T \left(\frac{1}{2} \mathbf{P} \mathbf{A} + \frac{1}{2} \mathbf{A}^T \mathbf{P} \right) \Phi_M \mathbf{x}_M \\
&= -\frac{1}{2} \mathbf{x}_M^T \Phi_M^T \mathbf{Q} \Phi_M \mathbf{x}_M \\
&< 0,
\end{aligned}$$

since $\mathbf{Q} > \mathbf{0}$. It follows that (66) is time-stable, stable in the sense of Lyapunov and energy-stable (Section 3). \square

The Lyapunov inner product (64) is a discrete counterpart of the continuous symmetry inner product (39). This inner product can be employed to construct stable Galerkin ROMs for (2) using discrete projection. An interesting question that arises is whether the matrix \mathbf{P} defining the Lyapunov inner product (64) is related in some way

to the matrix \mathbf{W} (57) that is used to perform the continuous projection in the symmetry inner product. In general, the answer is no. In particular, \mathbf{W} is by construction a sparse matrix (Figure 5(a)), whereas \mathbf{P} may be dense even if \mathbf{A} is sparse. This is clear from Figures 5(b) and (c), which show (respectively) the sparsity pattern of a sample \mathbf{A} matrix⁵, and its corresponding \mathbf{P} matrix.

One downside of the Lyapunov inner product is that the matrix \mathbf{P} which defines this inner product is admittedly expensive to compute: the cost of solving the Lyapunov equation (63) requires $\mathcal{O}(N^3)$ operations. As a consequence, the Lyapunov inner product has the same downside as another model reduction approach with an *a priori* stability guarantee, namely balanced truncation [29; 10]: it may not be practical to compute the matrix \mathbf{P} defin-

⁵The \mathbf{A} matrix whose sparsity pattern is shown in Figure 5(b) is the ‘‘PDE example’’ in the SLICOT model reduction benchmark repository [42].

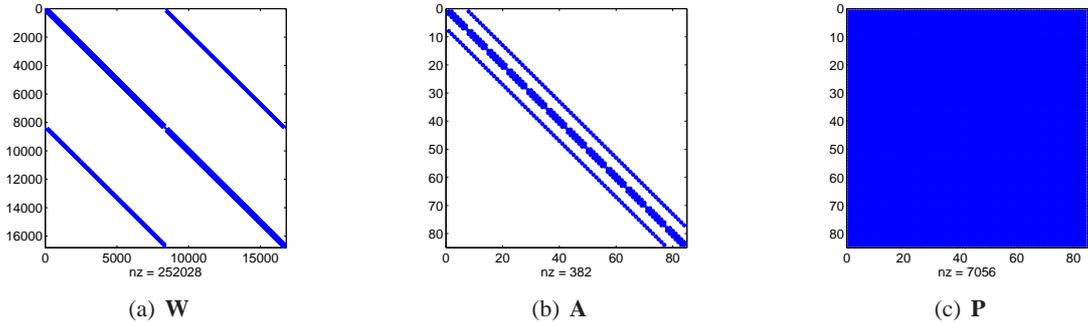


Figure 5: Sparsity structure of representative \mathbf{P} matrix for a given sparse \mathbf{A} matrix compared to sparsity structure of representative \mathbf{W} matrix

ing the Lyapunov inner product for very large systems. It is worthwhile to note that computing \mathbf{P} (63) is less computationally intensive than reducing a system using balanced truncation, which requires the solution of *two* Lyapunov equations for the so-called observability and reachability Gramians *and* the factorizations of these Gramians [29; 10] (see Appendix A.2). The computational cost of calculating the weighting matrix that defines the Lyapunov inner product relative to the computational cost of reducing a system using balanced truncation is studied numerically in Section 5.2. Note that it can be shown that the balanced truncation algorithm may be viewed as a projection algorithm in a special Lyapunov inner product [24]. A proof uncovering this connection is given in Appendix A.3.

As observed earlier for the symmetry inner product, it is clear from (66) that the Galerkin projection of the system (61) in the Lyapunov inner product (64) can be viewed as a Petrov-Galerkin projection of this system in the regular L^2 inner product, with the reduced test basis given by $\Psi_M = \mathbf{P}\Phi_M$, where Φ_M is the reduced trial basis.

5.2. Numerical experiments

The stability-preserving model reduction approach based on discrete projection presented in Section 5.1 is now evaluated on two examples: the international space station problem, and a problem involving a model of an electrostatically actuated beam. For both examples, the error in the ROM output

relative to the full order model output, defined by

$$\mathcal{E}_{rel}^o = \frac{\sum_{i=1}^{K_{max}} |\mathbf{y}_{QN}(t_i) - \mathbf{y}_{QM}(t_i)|}{\sum_{i=1}^{K_{max}} |\mathbf{y}_{QN}(t_i)|}, \quad (68)$$

is computed and reported. Here the symbol K_{max} denotes the integer such that $T_{max} = K_{max}dt_{snap}$, where T_{max} is the maximum time until which the ROM is run. The notation $|\cdot|$ in (68) denotes the absolute value, which evaluates to a scalar for the numerical examples considered herein, as they both have one output ($Q = 1$).

5.2.1. International space station (ISS) example

The first numerical example considered here involves a structural dynamics model of component 1r (Russian service module) of the international space station (ISS) [36]. The model consists of an LTI system of the form (2) with $N = 270$ and $P = Q = 3$. In the numerical test performed here, only the first input and first output is considered, so $P = Q = 1$. The matrices \mathbf{A} , \mathbf{B} and \mathbf{C} defining (2) are downloaded from the ROM benchmark repository [42]. It is verified that the FOM system is stable: the maximum real part of the eigenvalues of \mathbf{A} is -0.0031 .

To generate the snapshots from which POD bases are constructed, the full order model (2) is solved using a backward Euler time integration scheme with an initial condition of $\mathbf{x}_N(0) = \mathbf{0}$ and $\mathbf{u}_P(t) = (1 \times 10^4)\delta_{t=0}$. That is, at time $t = 0$, an impulse of magnitude 1×10^4 is applied. A total of $K = 2000$

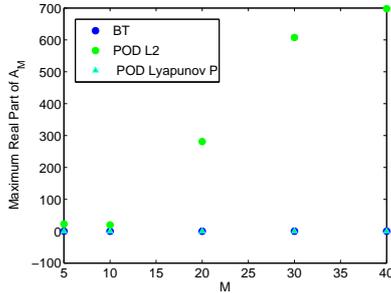


Figure 6: Maximum real part of eigenvalues of ROM system matrix \mathbf{A}_M for ISS problem

snapshots are collected, every $dt_{snap} = 5 \times 10^{-5}$ seconds, until time $t = 0.1$ seconds. These snapshots are used to construct POD bases of sizes $M = 5, 10, 20, 30,$ and 40 . For each M , a POD basis is constructed using the L^2 inner product, as well as the Lyapunov inner product (64). The matrix \mathbf{P} defining the inner product (64) is obtained using the `lyap` function in MATLAB’s control toolbox with $\mathbf{Q} = \mathbf{I}_N$, the $N \times N$ identity matrix (Section 5.1). The POD ROM solutions are compared with solutions obtained by reducing the system using balanced truncation [29; 10]. First, the eigenvalues of the ROM matrix \mathbf{A}_M for each M are computed to determine stability using Theorem 3.2.3. The maximum real part of the eigenvalues of these ROM system matrices is plotted in Figure 6 as a function of M . The reader can observe that the Lyapunov inner product POD ROMs and balanced truncation ROMs are stable for all M considered – all the real parts of the eigenvalues of these systems’ matrices are ≤ 0 . In contrast, the L^2 POD ROMs are unstable for all M .

Having checked stability, each ROM is run until a specified time T_{max} , and the average error in the output relative to the full order model (68) is computed. The relative errors (68) in the output for ROMs of different sizes run up to different values of T_{max} are summarized in Table 1. In the case a ROM went unstable and (68) overflowed, the table contains an entry of ‘-’.

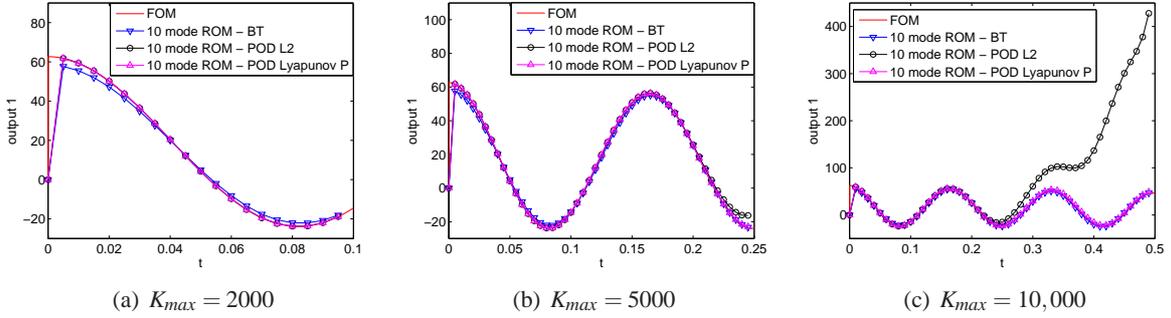
The objective of the $K_{max} = 2000$ ($T_{max} = 0.1$ seconds) run is to test how well the POD bases can reproduce the snapshots from which they were con-

structed, as exactly $K = 2000$ snapshots (taken up to $t = 0.1$ seconds) were used to generate these bases. Although the L^2 POD ROM is unstable for all values of M considered (Figure 6), this ROM still produces a reasonable solution for $M = 5$ and $M = 10$ (Figure 7(a) and Table 1). The instability manifests itself if a larger basis size is used, however. The Lyapunov ROM remains stable and accurate – orders of magnitude more accurate than the balanced truncation ROM for each M considered (Table 1).

The objective of the $K_{max} = 5000$ ($T_{max} = 0.25$ seconds) and $K_{max} = 10,000$ ($T_{max} = 0.5$ seconds) runs is to test the predictive capabilities of the POD ROMs relative to the balanced truncation ROMs for long-time simulations. The reduced order models are run for a much longer time horizon than the run used to generate the POD bases employed in building the ROMs. For $K_{max} = 5000$, The L^2 POD ROM exhibits an instability for all M considered except $M = 10$. For this value of M , the balanced truncation ROM and Lyapunov POD ROM are more accurate than the L^2 POD ROM, however (Figure 7(b) and Table 1). For $K_{max} = 10,000$, the L^2 POD ROM is unstable for all M considered. This instability is apparent in Figure 7(c). Hence, the instability identified in the earlier eigenvalue analysis (Figure 6) manifests itself if the L^2 POD ROM is run for a long enough time. For $K_{max} = 5000$ and $K_{max} = 10,000$, the Lyapunov POD ROM is more accurate than the balanced truncation ROM for small M . However, its accuracy is limited, as there does not appear to be a conver-

Table 1: Relative errors (68) \mathcal{E}_{rel}^o in ROM output for ISS problem

K_{max}	Method	M				
		5	10	20	30	40
2000	BT	9.80×10^{-2}	6.39×10^{-2}	9.56×10^{-3}	2.34×10^{-3}	8.34×10^{-4}
	POD L^2	1.09×10^{-4}	3.14×10^{-7}	—	—	—
	POD Lyapunov \mathbf{P}	8.69×10^{-6}	4.05×10^{-7}	1.13×10^{-6}	8.44×10^{-7}	9.22×10^{-7}
5000	BT	7.64×10^{-2}	4.68×10^{-2}	8.14×10^{-3}	1.87×10^{-3}	5.58×10^{-4}
	POD L^2	2.41	4.73×10^{-2}	—	—	—
	POD Lyapunov \mathbf{P}	2.88×10^{-2}	5.24×10^{-3}	1.31×10^{-2}	1.21×10^{-2}	2.86×10^{-2}
10,000	BT	6.87×10^{-2}	4.47×10^{-2}	7.08×10^{-3}	1.78×10^{-3}	5.76×10^{-4}
	POD L^2	165	3.24	—	—	—
	POD Lyapunov \mathbf{P}	5.25×10^{-2}	6.46×10^{-2}	9.92×10^{-2}	1.08×10^{-1}	9.92×10^{-2}


 Figure 7: $y_{QM}(t)$ for $M = 10$ ROMs (FOM = full order model) for ISS problem

gence with M -refinement.

5.2.2. Electrostatically actuated beam example

The second numerical example is that of an electrostatically actuated beam. One application for this model is analysis of microelectromechanical systems (MEMS) devices, such as electromechanical radio frequency (RF) filters [45]. Given a simple enough shape, these devices can be modeled as 1D beams embedded in two or three dimensional space. It is assumed that the beam deflection is small, so that geometric nonlinearities can be neglected. The resulting linear PDEs are discretized using the finite element method following the approach presented in [46; 45] to yield a ROM LTI system of the form (2). The matrices \mathbf{A} and \mathbf{B} in (2) are downloaded from the Oberwolfach model reduction benchmark collection [47]. These global matrices are then disassembled into

their local counterparts, and reassembled to yield a discretization of any desired size. In the full order model for which results are reported here, the FOM has $N = 10,000$ degrees of freedom. It is verified that the full order system is stable: the maximum real part of the eigenvalues of \mathbf{A} is -0.0016 .

To generate the snapshots from which POD bases are constructed, the full order model (2) is solved using a backward Euler time integration scheme with an initial condition of $\mathbf{x}_N(0) = \mathbf{0}$ and an input corresponding to a periodic on/off switching, i.e.,

$$\mathbf{u}_P(t) = \begin{cases} 0.005 < t < 0.01, 0.015 < t < 0.02, \\ 1, & 0.03 < t < 0.035, \\ 0, & \text{otherwise} \end{cases} \quad (69)$$

A total of $K_{max} = 1000$ snapshots are collected, every $dt_{snap} = 5 \times 10^{-5}$ seconds, until time $t = 0.05$ seconds. From these snapshots, 5, 10, 20 and 30 mode ROMs are constructed using POD in the L^2 inner product, and POD in the Lyapunov inner

product. In solving the Lyapunov equation (63) for the Lyapunov inner product weighting matrix \mathbf{P} , the matrix \mathbf{Q} is taken to be the $N \times N$ identity matrix. The system (2) is reduced also using balanced truncation.

As for the ISS example, the first step is to study the stability of each ROM. Figure 8 shows the maximum real part of the ROM system matrices \mathbf{A}_M for each M considered. It is found that the L^2 ROM is unstable for each M , and becomes more unstable with increasing M . In contrast, the balanced truncation and POD Lyapunov inner product ROMs are stable for all M considered, as expected.

Next, the accuracy of each ROM is examined. Table 2 summarizes the errors (68) in the ROM solutions relative to the full order model solution for three runs of different lengths. As before, an entry of ‘-’ in the table indicates that the error overflowed due to an instability in the ROM.

The objective of the first run ($K_{max} = 1000$) is to study how well the POD ROMs can reproduce the snapshots from which they were constructed, and to compare these ROMs’ performance with the performance of ROMs constructed using balanced truncation. The reader can observe that the POD ROM constructed in the Lyapunov inner product is the most accurate. The POD L^2 ROM is both unstable as well as inaccurate (Figure 9(a)).

The second two runs ($K_{max} = 2000$ and $K_{max} = 5000$) are aimed to study the predictive capabilities of the ROMs for long-time simulations. The full order model is run until times 0.1 and 2.5 seconds respectively. As before, only snapshots up to time $t = 0.05$ seconds are used to construct the POD bases for the ROMs. In addition to the signal (69), the following inputs are applied in both the full order model and the ROM:

$$\mathbf{u}_P(t) = \begin{cases} 0.055 < t < 0.06, 0.065 < t < 0.07, \\ 0.08 < t < 0.085, 0.105 < t < 0.11, \\ 0.115 < t < 0.12, 0.13 < t < 0.135, \\ 0.205 < t < 0.21, 0.215 < t < 0.22, \\ 1, & 0.23 < t < 0.235, \\ 0, & \text{otherwise.} \end{cases} \quad (70)$$

The reader may observe by examining Table 2 and

Figure 9 that the balanced truncation ROMs are in general the most accurate. The POD ROMs constructed in the Lyapunov inner product nonetheless produce reasonable results (Figures 9(b)-(c)) and appear to be converging to the full order model solution with M -refinement (Table 2). The POD L^2 ROM result is not shown in Figures 9(b)-(c), as the solution produced by this ROM blows up around time $t = 0.02$ seconds.

Lastly, the level of computational resources required for computing the Lyapunov inner product and the level of computational resources required for performing model reduction via balanced truncation [29; 10] are compared. Table 3 gives the CPU times for the sum of the following operations in the balanced truncation [29; 10] algorithm as a function of N , the problem size: calculation of the observability Gramian, calculation of the controllability Gramian, and calculation of the balancing transformation (Appendix A.2). All computations are performed in serial using MATLAB’s linear algebra capabilities and MATLAB’s control toolbox [44], on a Linux workstation with 6 Intel Xeon 2.93 GHz CPUs. Both methods exhibit $O(N^3)$ scaling. Although the Lyapunov inner product computation is costly, as it requires the solution of a Lyapunov equation, it completes in 2-3 times less CPU time than the balanced truncation algorithm. This is because balanced truncation requires the solution of *two* Lyapunov equations for the observability and reachability Gramians, as well as the Cholesky and eigenvalue factorizations of these Gramians.

6. Summary and conclusions

The energy-stability preserving model reduction approach developed specifically for the equations of linearized compressible inviscid flow in [6; 7] is generalized: for ROMs constructed using the continuous projection approach, it is shown that a transformation of a generic PDE system of the hyperbolic or incompletely parabolic type leads to a stable formulation of the Galerkin ROM for this system. It is then shown that, for many linear PDE

Table 2: Relative errors (68) \mathcal{E}_{rel}^o in ROM output for electrostatically actuated beam problem

K_{max}	Method	M			
		5	10	20	30
1000	BT	6.29×10^{-2}	4.51×10^{-3}	6.93×10^{-5}	3.60×10^{-6}
	POD L^2	8.56×10^{-1}	6.62	—	—
	POD Lyapunov \mathbf{P}	2.05×10^{-3}	6.23×10^{-5}	2.09×10^{-8}	1.35×10^{-8}
2000	BT	5.84×10^{-2}	4.47×10^{-3}	6.29×10^{-5}	3.17×10^{-6}
	POD L^2	7.76	4.26×10^3	—	—
	POD Lyapunov \mathbf{P}	3.62×10^{-2}	1.12×10^{-2}	3.47×10^{-4}	4.13×10^{-5}
5000	BT	7.36×10^{-2}	4.77×10^{-3}	5.48×10^{-5}	2.77×10^{-6}
	POD L^2	4.40×10^3	—	—	—
	POD Lyapunov \mathbf{P}	1.80×10^{-1}	1.09×10^{-1}	2.03×10^{-2}	6.09×10^{-3}

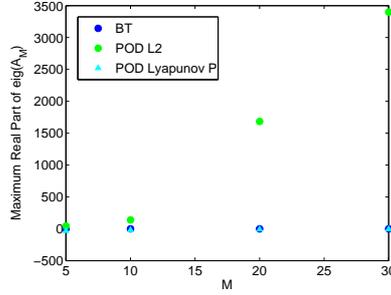


Figure 8: Maximum real part of eigenvalues of ROM system matrix \mathbf{A}_M for electrostatically actuated beam problem

Table 3: CPU Times (in seconds) for balanced truncation vs. Lyapunov inner product computations

Method	N			
	1250	2500	5000	10,000
Lyapunov Inner Product	5.08×10^1	4.60×10^2	4.02×10^3	6.09×10^4
Balanced Truncation	1.09×10^2	1.10×10^3	1.04×10^4	1.24×10^5

systems, the said transformation is induced by a special inner product, referred to as the “symmetry inner product”. If the Galerkin projection step of the model reduction procedure is performed in this inner product, the resulting ROM is guaranteed to satisfy certain stability bounds regardless of the reduced basis employed. It is demonstrated that a discrete counterpart of the symmetry inner product is the weighted L^2 inner product obtained by solving a Lyapunov equation, derived in [24] by Rowley *et al.* For completeness, this inner product, referred to as the “Lyapunov inner product”, is re-derived herein, and it is shown using the energy method that this inner product gives rise to stable ROMs constructed via discrete projection.

The performance of POD ROMs constructed using the symmetry and Lyapunov inner products are assessed on several numerical examples for which POD ROMs constructed in the L^2 inner product manifest instabilities.

The key properties of the symmetry inner product and Lyapunov inner product are summarized in Table 4. Both inner products are weighted L^2 inner products and have the same origin: they are induced by the Lyapunov function for the governing system of equations. The symmetry inner product is a continuous inner product derived for a specific PDE system of the form (24). Projection in this inner product requires access to the governing

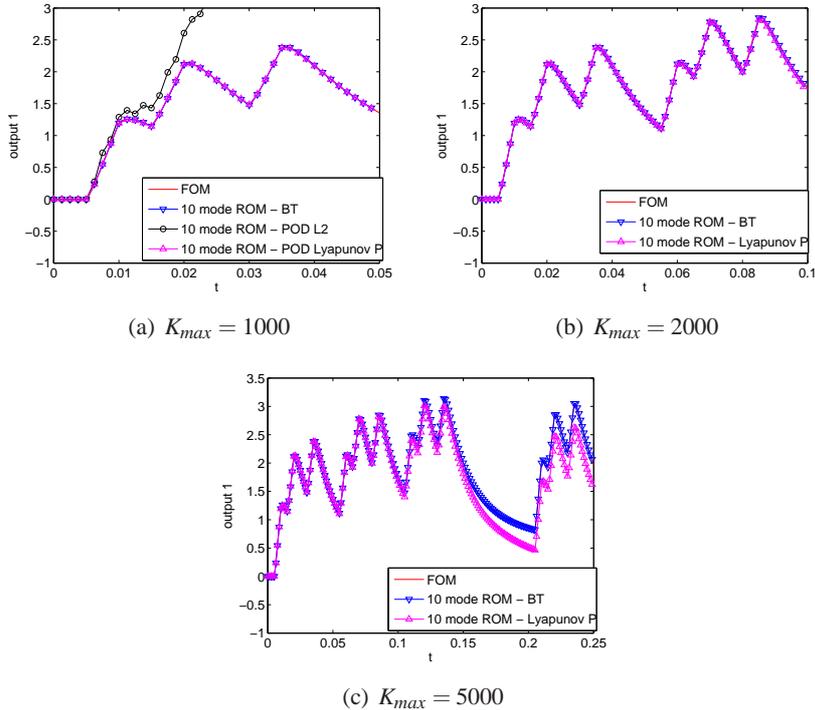


Figure 9: $y_{QM}(t)$ for $M = 10$ ROMs (FOM = full order model) for electrostatically actuated beam problem

PDEs, which gives rise to a projection algorithm that is embedded. The Lyapunov inner product is discrete, on the other hand, and operates on an LTI system of the form (2) arising from the discretization of a PDE of the form (1) in space using some numerical scheme, e.g., the finite element method. Projection in the Lyapunov inner product is therefore a black-box algorithm, as only the \mathbf{A} , \mathbf{B} and \mathbf{C} matrices in (2) are needed; in particular, access to the governing equations is *not* required. The symmetric positive definite matrix that defines the Lyapunov inner product can also be computed numerically in a black-box fashion by solving a Lyapunov equation. The existence of a solution to this Lyapunov equation is certain only if the full order system (2) is stable; hence the Lyapunov inner product is not defined for unstable systems. In contrast, the symmetry inner product *is* defined for unstable systems. In this case, a ROM constructed in this inner product will be energy-stable, by construction. However, it will not be time-stable, i.e., it may produce (physical) solutions that are unbounded as

$t \rightarrow \infty$. The discussion above may lead the reader to prefer the Lyapunov inner product to the symmetry inner product, as the former inner product can be computed in a black-box fashion for any stable linear system, and can be used to build a ROM for this system without accessing the PDEs. One of the biggest drawbacks of the Lyapunov inner product projection approach involves its large computational cost. To solve numerically the Lyapunov equation that defines this inner product, $\mathcal{O}(N^3)$ operations are required. Moreover, since the matrix that defines the Lyapunov inner product is typically dense (in contrast to the matrix defining the symmetry inner product, which is sparse), at least $\mathcal{O}(N^2)$ storage is required [11]. As a result, creating ROMs using the Lyapunov inner product may not be practical for systems of very large size. The Lyapunov inner product may nonetheless be preferable to balanced truncation, which requires the solution of two Lyapunov equations, and the storage of two Gramians, in addition to Cholesky and eigenvalue factorization of these Gramians. For

Table 4: Comparison of symmetry inner product and Lyapunov inner product

Symmetry Inner Product (39)	Lyapunov Inner Product (64)
Continuous	Discrete
For linear PDE system of the form $\dot{\mathbf{q}} + \mathbf{A}_i \frac{\partial \mathbf{q}}{\partial x_i} + \mathbf{K}_{ij} \frac{\partial^2 \mathbf{q}}{\partial x_i \partial x_j} + \mathbf{G}\mathbf{q} = \mathbf{f}$	For linear ODE system of the form $\dot{\mathbf{x}}_N = \mathbf{A}\mathbf{x}_N$
Defined for unstable systems but time-stability of ROM is not guaranteed	Undefined for unstable systems
Induced by Lyapunov function for the system	Induced by Lyapunov function for the system
Equation specific	Black-box
Derived analytically in closed form	Computed numerically by solving a Lyapunov equation
Sparse	Dense

large-scale unsteady problems, the symmetry inner product combined with the continuous projection approach is recommended by the authors, despite its more involved implementation.

Appendix

A.1. Gronwall's Lemma

Gronwall's lemma (also known as Gronwall's inequality) allows one to bound a function that is known to satisfy a certain differential or integral inequality by the solution of the corresponding differential or integral equation [55]. The differential form of this inequality is used herein:

$$\dot{\mathbf{x}}(t) \leq \beta(t)\mathbf{x}(t) \Rightarrow \mathbf{x}(T) \leq \mathbf{x}(0)e^{\int_0^T \beta(s)ds} \quad (71)$$

for $\beta \in L^2, t \geq 0, 0 \leq T \leq t$.

A.2. Balanced truncation algorithm for model reduction

The balanced truncation algorithm, first introduced by Moore [29], assumes a semi-discrete full order model of the form (2). The linear system (2) is first transformed into a balanced form that isolates observable and reachable (or controllable) modes. This is achieved by simultaneously diagonalizing

the reachability (or controllability) and observability Gramians. The reachability (or controllability) Gramian (Chapter 30 of [17])

$$\mathbf{P} \equiv \int_0^\infty e^{\mathbf{A}t} \mathbf{B} \mathbf{B}^T e^{\mathbf{A}^T t} dt, \quad (72)$$

is the unique symmetric (at least) positive semi-definite solution of the Lyapunov equation

$$\mathbf{A} \mathbf{P} + \mathbf{P} \mathbf{A}^T + \mathbf{B} \mathbf{B}^T = \mathbf{0}. \quad (73)$$

The observability Gramian (Chapter 30 of [17])

$$\mathbf{Q} \equiv \int_0^\infty e^{\mathbf{A}^T t} \mathbf{C}^T \mathbf{C} e^{\mathbf{A}t} dt, \quad (74)$$

is the unique symmetric (at least) positive semi-definite solution of the Lyapunov equation

$$\mathbf{A}^T \mathbf{Q} + \mathbf{Q} \mathbf{A} + \mathbf{C}^T \mathbf{C} = \mathbf{0}. \quad (75)$$

It will be assumed herein that the matrix \mathbf{A} defining the full order system (2) is stable, i.e., it has no eigenvalues with a positive real part. It will also be assumed (\mathbf{A}, \mathbf{C}) is observable and (\mathbf{A}, \mathbf{B}) is reachable (controllable). If this is true, the Lyapunov equations (73) and (75) will have positive definite solutions \mathbf{P} and \mathbf{Q} respectively (Chapter 6 of [18]). For a discussion of balanced truncation applied to unstable systems, the reader is referred to [22].

The balanced truncation algorithm is summarized below for the specific case of real system matrices⁶ \mathbf{A} , \mathbf{B} and \mathbf{C} . First, the reachability Gramian \mathbf{P} is obtained by solving the Lyapunov equation (73). Next, the observability Gramian \mathbf{Q} is obtained by solving the Lyapunov equation (75). The Cholesky factorization of \mathbf{P} is computed,

$$\mathbf{P} = \mathbf{U}\mathbf{U}^T. \quad (76)$$

followed by an eigenvalue decomposition of $\mathbf{U}^T\mathbf{Q}\mathbf{U}$:

$$\mathbf{U}^T\mathbf{Q}\mathbf{U} = \mathbf{K}\mathbf{\Sigma}^2\mathbf{K}^T. \quad (77)$$

The balancing transformation matrices:

$$\mathbf{T}_{bal} = \mathbf{\Sigma}^{1/2}\mathbf{K}^T\mathbf{U}^{-1}, \quad \mathbf{T}_{bal}^{-1} = \mathbf{U}\mathbf{K}\mathbf{\Sigma}^{-1/2}, \quad (78)$$

can now be computed⁷, where the entries of $\mathbf{\Sigma}$ are in decreasing order. The change of variables $\tilde{\mathbf{x}}_N(t) = \mathbf{T}_{bal}\mathbf{x}_N(t)$ is applied to the full-order LTI system (2) to yield:

$$\begin{aligned} \dot{\tilde{\mathbf{x}}}_N(t) &= \mathbf{T}_{bal}\mathbf{A}\mathbf{T}_{bal}^{-1}\tilde{\mathbf{x}}_N(t) + \mathbf{T}_{bal}\mathbf{B}\mathbf{u}_P(t), \\ \mathbf{y}_{QN}(t) &= \mathbf{C}\mathbf{T}_{bal}^{-1}\tilde{\mathbf{x}}_N(t). \end{aligned} \quad (79)$$

Next, the matrices $\tilde{\mathbf{A}} \equiv \mathbf{T}_{bal}\mathbf{A}\mathbf{T}_{bal}^{-1}$, $\tilde{\mathbf{B}} \equiv \mathbf{T}_{bal}\mathbf{B}$, $\tilde{\mathbf{C}} \equiv \mathbf{C}\mathbf{T}_{bal}^{-1}$ are partitioned as follows:

$$\begin{aligned} \tilde{\mathbf{A}} &= \begin{pmatrix} \tilde{\mathbf{A}}_{11} & \tilde{\mathbf{A}}_{12} \\ \tilde{\mathbf{A}}_{21} & \tilde{\mathbf{A}}_{22} \end{pmatrix}, \quad \tilde{\mathbf{B}} = \begin{pmatrix} \tilde{\mathbf{B}}_1 \\ \tilde{\mathbf{B}}_2 \end{pmatrix}, \\ \tilde{\mathbf{C}} &= (\tilde{\mathbf{C}}_1 \mid \tilde{\mathbf{C}}_2). \end{aligned} \quad (80)$$

Here, the blocks with subscript 1 correspond to the most observable and reachable states, and blocks with subscript 2 correspond to the least observable and reachable states. Finally, the reduced system for a ROM of size M is given by:

$$\begin{aligned} \dot{\mathbf{x}}_M(t) &= \mathbf{A}_M\mathbf{x}_M(t) + \mathbf{B}_M\mathbf{u}_P(t), \\ \mathbf{y}_{QM}(t) &= \mathbf{C}_M\mathbf{x}_M(t), \end{aligned} \quad (81)$$

⁶In the case these matrices are complex, the transpose operation T in the algorithm (and all analysis of this algorithm) should be replaced with a Hermitian transpose H .

⁷In practice, the transformation matrices (78) are typically computed as $\mathbf{T}_{bal} = \mathbf{V}^T\mathbf{Z}^T$, and $\mathbf{T}_{bal}^{-1} = \mathbf{U}\mathbf{W}$, where \mathbf{Z} is the Cholesky factor of the observability Gramian ($\mathbf{Q} = \mathbf{Z}\mathbf{Z}^T$), and \mathbf{W} is the left singular vector of $\mathbf{U}^T\mathbf{Z}$ ($\mathbf{U}^T\mathbf{Z} = \mathbf{W}\mathbf{\Sigma}\mathbf{V}^T$). This is due to numerical stability issues that could arise in computing $\mathbf{\Sigma}^{-1/2}$ in (78).

where $\mathbf{A}_M = \tilde{\mathbf{A}}_{11}$, $\mathbf{B}_M = \tilde{\mathbf{B}}_1$, $\mathbf{C}_M = \tilde{\mathbf{C}}_1$. The left and right reduced bases are given respectively by:

$$\mathbf{\Psi}_M = \mathbf{T}_{bal}^T(:, 1:M), \quad \mathbf{\Phi}_M = \mathbf{S}_{bal}(:, 1:M), \quad (82)$$

where $\mathbf{S}_{bal} \equiv \mathbf{T}_{bal}^{-1}$.

In effect, balanced truncation is a method for computing the test and trial bases $\mathbf{\Psi}_M$ and $\mathbf{\Phi}_M$ in (13). Given the test and trial bases defined in (82), the ROM system matrices (81) can be obtained from the formulas (14). The entries of the diagonal matrix $\mathbf{\Sigma}$ in (78) are known as the Hankel singular values of the system (2). Assuming a ROM of size M has been constructed using balanced truncation, the following error bound on the output can be shown [31]:

$$\|\mathbf{y}_{QN}(t) - \mathbf{y}_{QM}(t)\|_2 \leq 2 \sum_{i=M+1}^N \sigma_i \|\mathbf{u}_P(t)\|_2. \quad (83)$$

Generally, balanced truncation is viewed as the ‘‘gold standard’’ in model reduction. Although it is not optimal in the sense that there may be other ROMs with smaller error norms, the approach has *a priori* error bounds that are close to the lowest bounds achievable by any reduced order model [23]. Unfortunately, balanced truncation becomes computationally intractable for systems of very large dimension (e.g., of size $N \geq 10,000$), and hence is not practical for many systems of physical interest [24]. This is due to the high computational cost of solving the Lyapunov equations (73) and (75) for the reachability and observability Gramians ($\mathcal{O}(N^3)$ operations). The storage requirements of balanced truncation can be prohibitive as well. Even efficient iterative schemes developed for large sparse Lyapunov equations compute the solution to (73) and (75) in dense form, and hence require $\mathcal{O}(N^2)$ storage [11]. Unlike POD, balanced truncation delivers ROMs that preserve stability of a stable system (2) [29], however.

A.3. Lyapunov inner product associated with balanced truncation

In comparing the steps of the balanced truncation algorithm with the discussion in Section 5.1, the

reader may observe some similarities. In particular, both algorithms require the solution of a Lyapunov equation for a Gramian used to transform and reduce the system. Here, this connection is investigated further. In particular, it is shown that the balanced truncation algorithm (Appendix A.2) may be viewed as a projection algorithm in a special Lyapunov inner product.

Suppose the stable LTI system (2) has been reduced using the balanced truncation model reduction algorithm summarized in Appendix A.2. In order to uncover the inner product associated with balanced truncation, several transformations are required.

The first step is to substitute (78) into (82). Then, the following expressions for the left and right bases are obtained:

$$\Psi_M^T = \mathbf{T}_{bal}(1:M, :) = \Sigma^{1/2}(1:M, :)\mathbf{K}^T\mathbf{U}^{-1}, \quad (84)$$

$$\Phi_M = \mathbf{S}_{bal}(:, 1:M) = \mathbf{U}\mathbf{K}\Sigma^{-1/2}(:, 1:M). \quad (85)$$

Remark that (84) and (85) satisfy the following identity:

$$\Sigma^{-1}(1:M, 1:M)\Psi_M^T\mathbf{P} = \Phi_M^T, \quad (86)$$

where \mathbf{P} is the reachability Gramian (76). It follows that the ROM system matrices in (81) are:

$$\mathbf{A}_M = \Psi_M^T\mathbf{A}\Phi = \Psi_M^T\mathbf{A}\mathbf{P}^T\Psi_M\Sigma^{-1}(1:M, 1:M), \quad (87)$$

$$\mathbf{B}_M = \Psi_M^T\mathbf{B}, \quad (88)$$

$$\mathbf{C}_M = \mathbf{C}\Phi = \mathbf{C}\mathbf{P}^T\Psi_M\Sigma^{-1}(1:M, 1:M). \quad (89)$$

Defining

$$\mathbf{z}_M(t) \equiv \Sigma^{-1/2}(1:M, 1:M)\mathbf{x}_M(t), \quad (90)$$

and employing the symmetry property of the reachability Gramian ($\mathbf{P} = \mathbf{P}^T$), (81) becomes:

$$\begin{aligned} \dot{\mathbf{z}}_M(t) &= \hat{\Psi}_M^T\mathbf{A}\mathbf{P}\hat{\Psi}_M\mathbf{z}_M(t) + \hat{\Psi}_M^T\mathbf{B}\mathbf{u}_P(t), \\ \mathbf{y}_{QM}(t) &= \mathbf{C}\mathbf{P}\hat{\Psi}_M\mathbf{z}_M(t), \end{aligned} \quad (91)$$

where

$$\hat{\Psi}_M \equiv \Psi_M\Sigma^{-1/2}(1:M, 1:M). \quad (92)$$

It is clear that (91) defines a projection of the original LTI system (2) in an L^2 inner product weighted by the reachability Gramian matrix \mathbf{P} . This matrix defines a true inner product in the case when \mathbf{P} is symmetric positive-definite, which will hold if (\mathbf{A}, \mathbf{B}) is reachable (controllable)⁸.

A property of balanced truncation is that it preserves stability when applied to stable systems [10] (Appendix A.2). This result can be proven using the energy method. The proof is analogous to the proof of Theorem 5.1.1.

Acknowledgements

The research presented herein was funded by Sandia National Laboratories' Laboratory Directed Research and Development (LDRD) program. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

References

- [1] G. Serre, P. Lafon, X. Gloerfelt, C. Bailly, Reliable reduced-order models for time-dependent linearized Euler equations, *J. Comput. Phys.* **231**(15) (2012) 5176–5194.
- [2] W.J. Layton, Stable Galerkin methods for hyperbolic systems, *SIAM J. Numer. Anal.* **20**(3) (1983) 221–233.
- [3] F. Kwasniok, Empirical low-order models of barotropic flow, *J. Atmos. Sci.* **61**(2) (2004) 235–245.
- [4] M.D. Gunzburger, On the stability of Galerkin methods for initial-boundary value problems for hyperbolic systems, *Math. Comp.* **31** (139) (1977) 661–675.
- [5] D. Amsallem, C. Farhat, Stabilization of projection-based reduced order models, *Int. J. Numer. Meth. Engng.* **91** (4) (2012) 358–377.
- [6] M.F. Barone, I. Kalashnikova, D.J. Segalman, H. Thornquist, Stable Galerkin reduced order models for linearized compressible flow, *J. Comput. Phys.* **288** (2009) 1932–1946.

⁸Reachability (a.k.a. controllability) is a standard concept in control theory. The author is referred to [15] for a detailed discussion of reachability (controllability).

- [7] I. Kalashnikova, M.F. Barone, On the stability and convergence of a Galerkin reduced order model (ROM) for compressible flow with solid wall and far-field boundary treatment, *Int. J. Numer. Meth. Engng.* **83** (2010) 1345–1375.
- [8] B. Gustafsson, A. Sundstrom, Incompletely parabolic problems in fluid dynamics, *SIAM J. Appl. Math.* **35**(2) (1978) 343–357.
- [9] S. Abarbanel, D. Gottlieb, Optimal Time Splitting for Two- and Three-Dimensional Navier-Stokes Equations with Mixed Derivatives, *J. Comput. Phys.* **41** (1981) 1–33.
- [10] S. Gugercin, A.C. Antoulas, A survey of model reduction by balanced truncation and some new results, *Int. J. Control* **77**(8) (2004) 748–766.
- [11] S. Gugercin, J.-R. Li, Smith-Type Methods for Balanced Truncation of Large Sparse Systems, *Lecture Notes in Computational Science and Engineering* **45** (2005) 49–82.
- [12] B. Gustafsson, High order difference methods for time dependent PDE, Springer-Verlag, 2008.
- [13] G. Chen. Stability of Nonlinear Systems. *Encyclopedia of RF and Microwave Engineering*, Wiley, NY, 2004 (pp. 4881–4896).
- [14] B. Gustafsson, H.-O. Kreiss, J. Olinger, Time Dependent Problems and Difference Methods, Wiley-Interscience, 1995.
- [15] K.J. Astrom, R.M. Murray, Feedback systems: an introduction for scientists and engineers, Princeton University Press, 2008.
- [16] J.L. Lumley, Stochastic tools in turbulence , Academic Press: New York, 1971.
- [17] R.H. Bishop, The Mechatronics Handbook, CRC Press LLC, 2002.
- [18] H. Kimura, Chain-Scattering Approach to H-infinity Control, Springer, 1997.
- [19] P. Holmes, J.L. Lumley, G. Berkooz, Turbulence, Coherent Structures, Dynamical Systems and Symmetry, Cambridge University Press, 1996.
- [20] H.O. Kreiss, J. Lorenz, Initial-Boundary Value Problems and the Navier-Stokes Equations, Academic Press, Inc., 1989.
- [21] T. Bui-Thanh, K. Willcox, O. Ghattas, B. van Bloemen Waanders, Goal-oriented, model constrained optimization for reduction of large-scale systems, *J. Comp. Phys.* **224** (2007) 880–896.
- [22] P. Benner, M. Castillo, E.S. Quintana-Orti, G. Quintana-Orti, Parallel model reduction of large-scale unstable systems, *Advances in Parallel Computing* **13** (2004) 251–258.
- [23] C.W. Rowley, Model reduction for fluids using balanced proper orthogonal decomposition, *Int. J. Bif. Chaos* **15** (3) (2005) 997–1013.
- [24] C.W. Rowley, T. Colonius, R.M. Murray, Model reduction for compressible flows using POD and Galerkin projection, *Physica D* **189** (2004) 115–129.
- [25] L. Sirovich, Turbulence and the dynamics of coherent structures, part III: dynamics and scaling, *Q. Appl. Math.* **45** (3) (1987) 583–590.
- [26] N. Aubry, P. Holmes, J. Lumley, E. Stone, The dynamics of coherent structures in the wall region of a turbulent boundary layer, *J. Fluid Mech.* **192** (1988) 115–173.
- [27] K. Veroy, A.T. Patera, Certified real-time solution of the parametrized steady incompressible Navier-Stokes equations: rigorous reduced-bases *a posteriori* error bounds, *J. Num. Meth. Fluids* **47** (2005) 773–788.
- [28] G. Rozza. Reduced basis approximation and error bounds for potential flows in parametrized geometries. *Commun. Comput. Phys.* **9**(1) (2011) 1–48.
- [29] B. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Transactions on Automatic Control* **26** (1) (1981).
- [30] D. Amsallem, C. Farhat, On the stability of linearized reduced-order models: descriptor vs. non-descriptor form and application to fluid-structure interaction, AIAA Paper 2001-0926, *42nd AIAA Fluid Dynamics Conference & Exhibit*, New Orleans, LA (2012).
- [31] K. Willcox, J. Peraire, Balanced model reduction via the proper orthogonal decomposition, *AIAA Journal* **40** (11) (2002) 2323–2330.
- [32] E. Tadmor, Spectral Methods for Hyperbolic Problems, Lecture Notes Delivered at Ecole des Ondes, “*Méthodes numériques d’ordre élevé pour les ondes en régime transitoire*”, 1994.
- [33] M. Rathinam, L.R. Petzold, A new look at proper orthogonal decomposition, *SIAM J. Num. Anal.* **41** (5) (2003) 1893–1925.
- [34] K. Kunisch, S. Volkwein, Galerkin Proper Orthogonal Decomposition for a General Equation in Fluid Dynamics, *SIAM J. Num. Anal.* **40** (2) (2002) 492–515.
- [35] Z. Wang, I. Akhtar, J. Borggaard, Traian Iliescu, Proper orthogonal decomposition closure models for turbulent flows: A numerical comparison, *Comput. Methods Appl. Mech. Engrg.* (2012) 237–240.
- [36] A.C. Antoulas, D.C. Sorensen, S. Gugercin, A survey of model reduction methods for large-scale systems, *Contemporary Mathematics* **280** (2001) 193–219.
- [37] D. Funaro, D. Gottlieb, Convergence results for pseudospectral approximations of hyperbolic systems by a penalty-type boundary treatment, *Mathematics of Computation* **196** (57) (1991) 585–596.
- [38] A. Barbagallo, D. Sipp, P.J. Schmid, Closed-loop control of an open cavity flow using reduced-order models, *J. Fluid Mech.* **641** (2009) 1–50.
- [39] B.N. Bond, L. Daniel, Guaranteed stable projection-based model reduction for indefinite and unstable linear systems, *Proceedings of the 2008 IEEE/ACM International Conference on Computer-Aided Design* (2008).
- [40] M.J. Balajewicz, E.H. Dowell, B.R. Noack, Low-dimensional modelling of high-Reynolds-number shear flows incorporating constraints from the Navier-Stokes

- equation, *J. Fluid Mech.* **729** (2013) 285–308.
- [41] C.W. Rowley, I. Mezic, S. Baheri, P. Schlatter, D.S. Henningson, Reduced-order models for flow control: balanced models and Koopman modes, *Seventh IU-TAM Symposium on Laminar-Turbulent Transition*, June 2009.
- [42] Y. Chahlaoui, P. Van Dooren, Benchmark examples for model reduction of linear time invariant systems, <http://www.icm.tu-bs.de/NICONET/benchmodred.html>, Mar. 2013.
- [43] T.R. Smith, Low-dimensional models of plane Couette flow using the proper orthogonal decomposition, Ph.D. thesis, Princeton University, 2003.
- [44] The MathWorks, Inc., Control Systems Toolbox User's Guide, 1992–1998.
- [45] J. Lienemann, E.B. Rudnyi, J.G. Korvink, MST MEMS model order reduction: Requirements and benchmarks, *Linear Algebra Appl.* **415**(2-3) (2006) 469–498.
- [46] W. Weaver, Jr., S.P. Timoshenko, D.H. Young, Vibration problems in engineering, *Wiley, 5th Ed.*, 1990.
- [47] Oberwolfach benchmark collection, <http://portal.uni-freiburg.de/imteksimulation/downloads/benchmark/>, 2005.
- [48] D. Gottlieb, S.A. Orszag, *Numerical Analysis of Spectral Methods*, SIAM, 1977.
- [49] M.A. Heroux, R.A. Bartlett, V.E. Howle, R.J. Hoekstra, J.J. Hu, T.G. Kolda, R.B. Lehoucq, K.R. Long, R.P. Pawlowski, E.T. Phipps, A.G. Salinger, H.K. Thornquist, R.S. Tuminaro, J.M. Willenbring, A. Williams, K.S. Stanley, An overview of the Trilinos project, *ACM Trans. Math. Softw.* **31** (3) (2005).
- [50] B. Kirk, J.W. Peterson, R.H. Stogner, G.F. Carey, libMesh: A C++ library for parallel adaptive mesh refinement/coarsening simulations, *Eng. Comput.* **22** (3–4) (2006) 237254.
- [51] V. Sankaran, S. Menon. LES of Scalar Mixing in Supersonic Shear Layers, *Proceedings of the Combustion Institute*, **30** (2) 2835–2842 (2004).
- [52] F. Genin, S. Menon, Studies of Shock/Turbulent Shear Layer Interaction Using Large- Eddy Simulation, *Computers and Fluids*, **39** 800–819 (2010).
- [53] F. Genin, S. Menon, Dynamics of Sonic Jet Injection into Supersonic Crossflow, *J. Turbulence* **11**(4) 1–30 (2010).
- [54] W.M. Haddad, S.G. Nersesov. Stability and control of large-scale dynamical systems: A Vector Dissipative Systems Approach, Princeton University Press, 2011.
- [55] T.H. Gronwall, Note on the derivatives with respect to a parameter of the solutions of a system of differential equations, *Ann. of Math.* **20** (2) 292–296 (1919).
- [56] **I. Kalashnikova, B.G. van Bloemen Waanders, S. Arunajatesan, M.F. Barone, Stabilization of Projection-Based Reduced Order Models for Linear Time-Invariant Systems via Optimization-Based Eigenvalue Reassignment, *Comput. Meth. Appl.***

Mech. Engng. **272** 251–270 (2014).