SANDIA REPORT

SAND2013-4063
Unlimited Release
Printed May 2013

Supersedes SAND2013-4063
Dated May 2013

Construction of Energy-Stable Galerkin Reduced Order Models

Irina Kalashnikova, Matthew F. Barone, Srinivasan Arunajatesan, Bart G. van Bloemen Waanders

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online ordering: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2013-4063
Unlimited Release
Printed May 2013

Supersedes SAND2013-4063
dated May 2013

Construction of Energy-Stable Galerkin Reduced Order Models

Irina Kalashnikova
Numerical Analysis & Applications Department
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-1320

Matthew F. Barone and Srinivasan Arunajatesan
Aerosciences Department
Sandia National laboratories
P.O. Box 5800
Albuquerque, NM 87185-0825

Bart G. van Bloemen Waanders
Numerical Analysis & Applications Department
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-1318

Abstract

This report aims to unify several approaches for building stable projection-based reduced order models (ROMs). Attention is focused on linear time-invariant (LTI) systems. The model reduction procedure consists of two steps: the computation of a reduced basis, and the projection of the governing partial differential equations (PDEs) onto this reduced basis. Two kinds of reduced bases are considered: the proper orthogonal decomposition (POD) basis and the balanced truncation basis. The projection step of the model reduction can be done in two ways: via continuous projection or via discrete projection. First, an approach for building energy-stable Galerkin ROMs for

linear hyperbolic or incompletely parabolic systems of PDEs using continuous projection is proposed. The idea is to apply to the set of PDEs a transformation induced by the Lyapunov function for the system, and to build the ROM in the transformed variables. The resulting ROM will be energy-stable for any choice of reduced basis. It is shown that, for many PDE systems, the desired transformation is induced by a special weighted L^2 inner product, termed the "symmetry inner product". Attention is then turned to building energy-stable ROMs via discrete projection. A discrete counterpart of the continuous symmetry inner product, a weighted L^2 inner product termed the "Lyapunov inner product", is derived. 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 system of PDEs in space. It is shown that a ROM constructed via discrete projection using the Lyapunov inner product will be energy-stable for any choice of reduced basis. Connections between the Lyapunov inner product and the inner product induced by the balanced truncation algorithm are made. Comparisons are also made between the symmetry inner product and the Lyapunov inner product. The performance of ROMs constructed using these inner products is evaluated on several benchmark test cases.

Contents

1 Introduction 7
2 Projection-Based Model Reduction 10
2.1 Projection 11
2.2 Reduced Bases 13
3 Stability Definitions 19
3.1 Energy-Stability 19
3.2 Time-Stability 20
3.3 Lyapunov Stability 21
4 Stable Model Reduction for LTI Systems via Continuous Projection 22
4.1 A Stabilizing Transformation 22
4.2 Stability-Preserving "Symmetry Inner Product" and Petrov-Galerkin Connection 25
4.3 Lyapunov Function Connection 26
4.4 Examples 27
4.5 Stability-Preserving Discrete Implementation 31
4.6 Numerical Experiment 32
5 Stable Model Reduction for LTI Systems via Discrete Projection 38
5.1 Stability-Preserving Lyapunov Inner Product 38
5.2 Lyapunov Inner Product Associated with Balanced Truncation 40
5.3 Numerical Experiments 42
6 Summary and Conclusions 50
References 52

Appendix

Figures

1 Sparsity structure of representative $\mathbf{W}$ matrix (98) 32
2 Time history of modal amplitudes for inviscid pressure pulse problem 34
3 Time history of modal amplitudes for inviscid pressure pulse problem for longer time horizon 34
4 Pressure field at time $t = 4.5 \times 10^{-4}$ seconds 35
5 Pressure field at time $t = 2.95 \times 10^{-3}$ seconds 36
6 Pressure field at time $t = 7.95 \times 10^{-3}$ seconds 37
7 Sparsity structure of representative $\mathbf{P}$ matrix for a given sparse $\mathbf{A}$ matrix . . . 40
8 Maximum real part of eigenvalues of ROM system matrix $\mathbf{A}_M$ for ISS problem 43
9 $\mathbf{y}_{QM}(t)$ for $M = 10$ ROMs (FOM = full order model) for ISS Problem 45
10 $\mathbf{y}_{QM}(t)$ for $M = 40$ Lyapunov POD and Balanced Truncation ROMs, $K_{max} = 100,000$ (FOM = full order model) for ISS Problem 45
11 Maximum real part of eigenvalues of ROM system matrix $\mathbf{A}_M$ for beam problem 47
12 $\mathbf{y}_{QM}(t)$ for $M = 10$ ROMs (FOM = full order model) for Beam Problem 49

Tables

1 Relative Errors (122) $\mathcal{E}_{rel}^{o}$ in ROM Outputs for ISS Problem 44
2 Relative Errors (122) $\mathcal{E}_{rel}^{o}$ in ROM Outputs for Beam Problem 48
3 CPU Times (in seconds) for Balanced Truncation vs. Lyapunov Inner Product Computations 49

1 Introduction

Despite improved algorithms and the availability of massively parallel computing resources, simulations of the fidelity commonly employed in system modeling for today's science and engineering applications are in practice often too computationally expensive for use in a design or analysis setting. This situation has pushed researchers to develop reduced order models (ROMs): models constructed from a high-fidelity simulation that retain the essential physics and dynamics of a high-fidelity model, but have a a much lower computational cost and can thus be run in real or near-real time for use in applications that require on-the-spot decision making, optimization, 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.
- 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 the accuracy of the ROM. It is well-known that the model reduction method known as balanced truncation [36, 18] has a rigorous stability guarantee. 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 [39]). Less costly model reduction approaches such as the balanced proper orthogonal decomposition (BPOD) method [48, 38], and the proper orthogonal decomposition (POD) method [42, 37, 24] lack, in general, an *a priori* stability guarantee. In [3], 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. The stability of a POD or BPOD ROM is commonly evaluated *a posteriori*: the ROM is constructed, used to predict some dynamical behavior, and deemed a success if the solutions generated by the ROM are numerically stable and accurately reproduce the expected behavior. There is some risk inherent in this sort of analysis, as the ROM could introduce non-physical instabilities into the approximation.

The importance of obtaining stable ROMs has been recognized in recent years by a number of authors. In [39], Rowley *et al.* show that Galerkin projection preserves the stability of an equilibrium point at the origin if an "energy-based" inner product is employed. In [9, 27, 26], Barone *et al.* demonstrate that a symmetry transformation leads to a stable formulation for a Galerkin ROM for the linearized compressible Euler equations [9, 27] and non-linear compressible Navier-Stokes equations [26] with solid wall and far-field boundary conditions. In [41], Serre *et al.* propose applying the stabilizing projection developed by Barone *et al.* in

[9, 27] 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 energy of the physical 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 will minimally modify the ROM. In [4], Amsallem *et al.* propose a method for stabilizing projection-based linear ROMs through the solution of a small-scale convex optimization problem. In [12], 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. In [7], 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 [47], 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).

The primary objective of the present work is to unify various approaches that fall into the first "class" of ROM stabilization approaches described above (those derived to have an *a priori* stability guarantee) using the energy method [23] and the concept of "energy-stability". The work is motivated by the observation that many of these approaches, e.g., those presented in [39, 9, 27, 26, 41], have certain similarities. In particular:

- All these methods require a transformation of the governing equations prior to constructing the ROM.
- The application of this transformation is equivalent to performing the Galerkin projection of the model reduction in a special weighted inner product, the so-called "energy inner product".
- For linear problems, Galerkin projection in the "energy inner product" is equivalent to a Petrov-Galerkin projection (a projection in which the test and trial reduced bases differ) in the L^2 inner product.

The observations made above motivate the following questions:

- The energy inner products derived in [39, 9, 27, 26, 41] are specific to the equations of compressible flow. Is it possible to derive the energy inner product for a general partial differential equation (PDE)?
- The energy inner products derived in [39, 9, 27, 26, 41] assume the Galerkin projection step of the model reduction is performed at the level of the continuous PDEs. Is it possible to compute numerically a discrete form of the energy inner product?

- What is the connection between balanced trunction [36, 18], a model reduction technique with a stability guarantee, and ROMs constructed in the aforementioned stability-preserving energy inner products?

The present work aims to address these questions. To this effect, the remainder of this report is organized as follows. Projection-based model reduction is overviewed in Section 2. The notions of stability that are employed in this report (energy-stability, time-stability, Lyapunov stability) are defined in Section 3. Section 4 focuses on the construction of energy-stable ROMs for linear systems of PDEs using continuous projection. 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 (in particular, the POD basis). It is then shown that, for many PDE systems, the desired transformation is induced by a special weighted L^2 inner product, termed the "symmetry inner product". Examples of this inner product are given for several systems of physical interest (the wave equation, the linearized shallow water equations, the linearized compressible Euler equations, and the linearized compressible Navier-Stokes equations). A discrete counterpart of the continuous symmetry inner product, a weighted L^2 inner product termed the "Lyapunov inner product", is derived in Section 4. The weighting matrix that defines this 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. The Galerkin projection of the LTI full order system in this inner product gives rise to a ROM with a stability guarantee, again for *any* choice of reduced basis. Connections between the Lyapunov inner product and the inner product induced by balanced truncation are made. Conclusions are offered in Section 6.

2 Projection-Based Model Reduction

In this section, several approaches to building projection-based reduced order models are reviewed. Attention is restricted to linear time-invariant (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 [6].

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

$$\dot{\mathbf{x}}(t) = \mathcal{L}(\mathbf{x}(t)) + \mathcal{L}_c(\mathbf{u}(t)), \quad \mathbf{y}(t) = \mathcal{L}_o(\mathbf{x}(t)), \text{ in } \Omega. \tag{1}$$

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 control variables. $\mathbf{y} \in \mathbb{R}^q$ is the measured signal or output. Ω is an open bounded domain. The '$\cdot$' symbol denotes diffentiation with respect to time, i.e., $\dot{\mathbf{x}} \equiv \frac{\partial \mathbf{x}}{\partial t}$. The operator $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^n$ is a smooth linear spatial-differential operator, and $\mathcal{L}_c : \mathbb{R}^p \to \mathbb{R}^n$ and $\mathcal{L}_o : \mathbb{R}^n \to \mathbb{R}^q$ are smooth linear mappings.

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

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

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 a function of time t).

The general approach to projection-based model reduction consists of two steps:

Step 1: Calculation of trial and test reduced bases, each of order M, with M "small".
Step 2: Projection of the governing system ((1) or (2)) onto the test reduced basis.

The result of this procedure is a "small" dynamical system that accurately describes the dynamics of the full order system for some set of conditions. The two steps comprising the projection-based ROM procedure are detailed in the following subsections. There are two approaches for performing step 2 of the model reduction: continuous and discrete projection, described in Section 2.1. There exist a number of approaches for calculating the reduced basis modes (step 1 of the model reduction), e.g., POD [42, 37, 24], balanced POD [48, 38], balanced truncation [36, 18], goal-oriented bases [13], generalized eigenmodes [8], Koopman modes [40]. Attention is restricted here to two kinds of reduced bases: POD and balanced truncation (Section 2.2).

2.1 Projection

Model Reduction via Continuous Projection

In the continuous projection approach [9, 27], the continuous system of PDEs (1) is projected onto a continuous test reduced basis $\{\boldsymbol{\psi}_i\}_{i=1}^M \in \mathbb{R}^n$ in a continuous inner product, denoted generically (for now) by $(\cdot,\cdot)$. For example, $(\cdot,\cdot)$ could denote the usual L_2 inner product, i.e.,

$$\left(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}\right) = \int_\Omega \mathbf{x}^{(1)T}\mathbf{x}^{(2)} d\Omega. \tag{3}$$

First, the solution to (1) is approximated as

$$\mathbf{x}(t) \approx \sum_{i=1}^{M} x_{M,i}(t)\boldsymbol{\phi}_i, \tag{4}$$

where $x_{M,i}(t)$ are the unknown ROM coefficients or modal amplitudes, to be determined in solving the ROM. The reduced trial basis functions $\boldsymbol{\phi}_i$ (as well as the reduced test basis functions $\boldsymbol{\psi}_i$) are a function of space but not time.

Substituting (4) into (1), the following is obtained

$$\sum_{i=1}^{M} \dot{x}_{M,i}(t)\boldsymbol{\phi}_i = \mathcal{L}\left(\sum_{i=1}^{M} x_{M,i}(t)\boldsymbol{\phi}_i\right) + \mathcal{L}_c(\mathbf{u}), \quad \mathbf{y}_{QM}(t) = \mathcal{L}_o\left(\sum_{i=1}^{M} x_{M,i}(t)\boldsymbol{\phi}_i\right), \tag{5}$$

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

Next, a test reduced basis $\{\boldsymbol{\psi}_i\}_{i=1}^M \in \mathbb{R}^n$ is introduced, and the system of PDEs (5) is projected onto the test reduced basis modes $\boldsymbol{\psi}_j$ for $j = 1, 2, ..., M$ in the inner product $(\cdot,\cdot)$ to yield

$$\sum_{i=1}^{M} \dot{x}_{M,i}(t)\left(\boldsymbol{\psi}_j, \boldsymbol{\phi}_i\right) = \left(\boldsymbol{\psi}_j, \mathcal{L}\left(\sum_{i=1}^{M} x_{M,i}(t)\boldsymbol{\phi}_i\right)\right) + \left(\boldsymbol{\psi}_j, \mathcal{L}_c(\mathbf{u})\right), \quad \mathbf{y}_{QM}(t) = \mathcal{L}_o\left(\sum_{i=1}^{M} x_{M,i}(t)\boldsymbol{\phi}_i\right), \tag{6}$$

for $j = 1, 2, ..., M$. Typically, the trial and test reduced bases $\boldsymbol{\phi}_i$ and $\boldsymbol{\psi}_i$ are chosen to be orthonormal in the inner product $(\cdot,\cdot)$, so that $(\boldsymbol{\psi}_j, \boldsymbol{\phi}_i) = \delta_{ij}$, where δ_{ij} denotes the Krönecker delta function. Invoking this property, as well as the linearity property of the operators $\mathcal{L}$ and $\mathcal{L}_o$, (6) simplifies to

$$\dot{x}_{M,j}(t) = \sum_{i=1}^{M} x_{M,i}(t)\left(\boldsymbol{\psi}_j, \mathcal{L}(\boldsymbol{\phi}_i)\right) + \left(\boldsymbol{\psi}_j, \mathcal{L}_c(\mathbf{u})\right), \quad \mathbf{y}_{QM}(t) = \sum_{i=1}^{M} x_{M,i}(t)\mathcal{L}_o(\boldsymbol{\phi}_i), \tag{7}$$

for $j = 1, 2, ..., M$. (7) is a set of M time-dependent ordinary differential equations (ODEs) for the modal amplitudes $x_{M,i}(t)$ in (4).

In the case that the test and trial reduced basis vectors differ, the projection (6) is referred to as a *Petrov-Galerkin projection*. Otherwise, if $\boldsymbol{\psi}_i = \boldsymbol{\phi}_i$ for $i = 1, 2, ..., M$, the projection (6) is referred to as a *Galerkin projection*.

Model Reduction via Discrete Projection

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

$$\left(\mathbf{x}_N^{(1)}, \mathbf{x}_N^{(2)}\right)_{\mathbf{P}} = \mathbf{x}_N^{(1)T}\mathbf{P}\mathbf{x}_N^{(2)}, \tag{8}$$

where $\mathbf{P} \in \mathbb{R}^{N\times N}$ is a symmetric positive-definite matrix. Let $\mathbf{\Phi}_M \in \mathbb{R}^{N\times M}$ and $\mathbf{\Psi}_M \in \mathbb{R}^{N\times M}$ denote the trial and test reduced bases for (2), respectively. Assume these matrices have full column rank, and are orthonormal in the inner product (8), so that $\mathbf{\Psi}_M^T\mathbf{P}\mathbf{\Phi}_M = \mathbf{I}_M$, 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

$$\mathbf{x}_N(t) \approx \mathbf{\Phi}_M\mathbf{x}_M(t), \tag{9}$$

where $\mathbf{x}_M(t) \in \mathbb{R}^M$ is the ROM solution (to be determined). As in the continuous projection approach, the reduced bases $\mathbf{\Phi}_M$ and $\mathbf{\Psi}_M$ are not a function of time. Substituting (9) into (2), and projecting this system onto the test reduced basis, the following is obtained:

$$\dot{\mathbf{x}}_M(t) = \mathbf{\Psi}_M^T\mathbf{P}\mathbf{A}\mathbf{\Phi}_M\mathbf{x}_M(t) + \mathbf{\Psi}_M^T\mathbf{P}\mathbf{B}\mathbf{u}_P(t), \quad \mathbf{y}_{QM}(t) = \mathbf{C}\mathbf{\Phi}_M\mathbf{x}_M(t), \tag{10}$$

where $\mathbf{y}_{QM}$ is a reduced approximation of the output. (11) is an $M \times M$ LTI system of the form

$$\dot{\mathbf{x}}_M(t) = \mathbf{A}_M\mathbf{x}_M(t) + \mathbf{B}_M\mathbf{u}_P(t), \quad \mathbf{y}_{QM}(t) = \mathbf{C}_M\mathbf{x}_M(t), \tag{11}$$

where

$$\mathbf{A}_M = \mathbf{\Psi}_M^T\mathbf{P}\mathbf{A}\mathbf{\Phi}_M, \quad \mathbf{B}_M = \mathbf{\Psi}_M^T\mathbf{P}\mathbf{B}, \quad \mathbf{C}_M = \mathbf{C}\mathbf{\Phi}_M. \tag{12}$$

Again, in the case that $\mathbf{\Psi}_M \neq \mathbf{\Phi}_M$, the projection (11) is referred to as a *Petrov-Galerkin projection*. Otherwise, if $\mathbf{\Psi}_M = \mathbf{\Phi}_M$, the projection (11) is referred to as a *Galerkin projection*.

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. Certain properties of the numerical scheme used to solve the full equations may be inherited by the ROM as well. The discrete approach is black-box, at least for linear systems (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 and/or access to the governing equations is not required. 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 (7) using a numerical quadrature (Section 4.5). Although the continuous approach is not black-box, as it requires access to the governing PDEs, its similarity to spectral numerical approximation methods allows the use of analysis techniques employed by the spectral methods community [17, 27].

Note that, regardless of which projection approach is used to build the ROM, the ROM dynamical system will have the form (11), as (7) has this form when written as a matrix problem. The solution to the ROM is obtained by advancing (7) forward in time using a time-integration scheme. Since the system considered here is linear, the projection terms in (7) 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.

2.2 Reduced Bases

Attention is now turned to the computation of the reduced bases employed in the ROM. Two approaches for computing reduced bases are summarized herein: the proper orthogonal decomposition (POD) and balanced truncation.

Proper Orthogonal Decomposition (POD)

The proper orthogonal decomposition, or POD, is a widely used approach for computing efficient bases for dynamical systems. Discussed in detail in Lumley [34] and Holmes *et al.* [24], POD is a mathematical procedure that, given an ensemble of data and an inner product, denoted generically by $(\cdot,\cdot)$, constructs a basis for that ensemble that 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 ($M << K$) subspace spanned by the set $\{\boldsymbol{\phi}_i\}$ 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 [9, 24, 30, 35] that the solution to the POD optimization problem reduces to the eigenvalue problem

$$\mathcal{R}\boldsymbol{\phi} = \lambda\boldsymbol{\phi}, \tag{13}$$

where

$$\mathcal{R} \equiv \langle \mathbf{x}^k \otimes \mathbf{x}^k \rangle, \tag{14}$$

is a self-adjoint and positive semi-definite operator. If it is assumed that $\mathcal{R}$ is compact, then there exists a countable set of non-negative eigenvalues λ_i with associated eigenfunctions $\boldsymbol{\phi}_i$. It can be shown [24, 34] that the set of M eigenfunctions, or POD modes, $\{\boldsymbol{\phi}_i : i = 1,\dots,M\}$ corresponding to the M largest eigenvalues of $\mathcal{R}$ is precisely the desired basis.

In practice, the POD basis is typically obtained through a singular value decomposition (SVD) of the snapshot matrix rather than by solving the eigenvalue problem (13). The POD procedure is summarized in Algorithm 1 below.

Algorithm 1 Proper Orthogonal Decomposition (POD)

Step 1: Collect K snapshots of the solution vector $\{\mathbf{x}^k\}_{k=1}^K \in \mathbb{R}^N$. Place these snapshots into the columns of a matrix $\mathbf{X}$ defined by

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}^1 & \dots \mathbf{x}^K \end{pmatrix} \in \mathbb{R}^{N \times K}. \tag{15}$$

Step 2: Select an inner product to build the reduced basis in, e.g., the inner product (8).
Step 3: Compute the SVD of the matrix

$$\frac{1}{K}\mathbf{X}^T\mathbf{P}\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^T. \tag{16}$$

Step 4: Let

$$\mathbf{\Phi}_M = \begin{pmatrix} \boldsymbol{\phi}_1, & \cdots, & \boldsymbol{\phi}_M \end{pmatrix} = \mathbf{X}\mathbf{U}(:, 1:M) \tag{17}$$

be the POD reduced basis of size M.
Step 5: Orthonormalize the POD basis computed in Step 4:

$$\boldsymbol{\phi}_i = \frac{\boldsymbol{\phi}_i}{\boldsymbol{\phi}_i^T\mathbf{P}\boldsymbol{\phi}_i}. \tag{18}$$

for $i = 1$ to M.

The following error formula can be shown for the POD [24, 30]:

$$\frac{1}{M}\sum_{i=1}^{M}\left|\left|\mathbf{x}^i - \sum_{j=1}^{M}\left(\mathbf{x}^i, \boldsymbol{\phi}_j\right)_{\mathbf{P}}\boldsymbol{\phi}_j\right|\right|_{\mathbf{P}}^2 = \sum_{k=M+1}^{N}\lambda_k, \tag{19}$$

where $\lambda_1 \geq ... \geq \lambda_N > 0$ are the positive eigenvalues of the operator $\mathcal{R}$ (14).

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

$$E_{POD}(M) \geq \text{tol} \tag{20}$$

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

$$E_{POD}(M) \equiv \frac{\sum_{i=1}^{M}\lambda_i}{\sum_{i=1}^{N}\lambda_i}. \tag{21}$$

Balanced Truncation

Another approach for constructing reduced bases in building projection-based ROMs is balanced truncation, first introduced by Moore [36]. The balanced truncation algorithm 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, defined below.

Definition 2.4.1 (Chapter 30 of [11]): The *reachability (or controllability) Gramian*

$$\mathbf{P} \equiv \int_0^\infty e^{\mathbf{A}t}\mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T t} dt \tag{22}$$

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

$$\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = \mathbf{0}. \tag{23}$$

Definition 2.4.2 (Chapter 30 of [11]): The *observability Gramian*

$$\mathbf{Q} \equiv \int_0^\infty e^{\mathbf{A}^T t}\mathbf{C}^T\mathbf{C} e^{\mathbf{A}t} dt \tag{24}$$

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

$$\mathbf{A}^T\mathbf{Q} + \mathbf{Q}\mathbf{A} + \mathbf{C}^T\mathbf{C} = \mathbf{0}. \tag{25}$$

The following lemmas state some important properties of the reachability and observability Gramians.

Lemma 2.4.3 (Chapter 6 of [28]): Assume that $(\mathbf{A}, \mathbf{B})$ is reachable (controllable). Then the Lyapunov equation (23) has a positive definite solution $\mathbf{P}$ if and only if $\mathbf{A}$ is stable (i.e., does not have any eigenvalues with a positive real part).

Lemma 2.4.4 (Chapter 6 of [28]): Assume that $(\mathbf{A}, \mathbf{C})$ is observable. Then the Lyapunov equation (25) has a positive definite solution $\mathbf{Q}$ if and only if $\mathbf{A}$ is stable (i.e., does not have any eigenvalues with a positive real part).

In the present work, it will be assumed 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). For a discussion of balanced truncation applied to unstable systems for which the conditions of Lemmas 2.4.3 and 2.4.4 do not hold, the reader is referred to [10]. The balanced truncation algorithm is summarized in Algorithm 2 below[1].

[1]By Lemmas 2.4.3 and 2.4.4, the $\mathbf{P}$ and $\mathbf{Q}$ matrices (solutions to (26) and (27) respectively) exist. Moreover, these matrices are both symmetric and at least positive semi-definite. Hence the Cholesky factorization (28) exists. Note that in Algorithm 2 and all subsequent analysis of this algorithm, it has been assumed that $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are real matrices. In the case these matrices are complex, the transpose operation ^{T} in Algorithm 2 (and all subsequent analysis of this algorithm) should be replaced with a Hermitian transpose ^{H}.

Algorithm 2 Model Reduction via Balanced Truncation

Step 1: Solve for the reachability Gramian $\mathbf{P}$ by solving the Lyapunov equation:

$$\mathbf{AP} + \mathbf{PA}^T + \mathbf{BB}^T = \mathbf{0}. \tag{26}$$

Step 2: Solve for the observability Gramian $\mathbf{Q}$ by solving the Lyapunov equation:

$$\mathbf{A}^T\mathbf{Q} + \mathbf{QA} + \mathbf{C}^T\mathbf{C} = \mathbf{0}. \tag{27}$$

Step 3: Compute the Cholesky factorization of $\mathbf{P}$:

$$\mathbf{P} = \mathbf{UU}^T. \tag{28}$$

Step 4: Compute the eigenvalue decomposition of $\mathbf{U}^T\mathbf{QU}$:

$$\mathbf{U}^T\mathbf{QU} = \mathbf{K}\boldsymbol{\Sigma}^2\mathbf{K}^T. \tag{29}$$

Step 5: Compute the balancing transformation matrices:

$$\mathbf{T}_{bal} = \boldsymbol{\Sigma}^{1/2}\mathbf{K}^T\mathbf{U}^{-1}, \quad \mathbf{T}_{bal}^{-1} = \mathbf{UK}\boldsymbol{\Sigma}^{-1/2}, \tag{30}$$

where the entries of $\boldsymbol{\Sigma}$ are in decreasing order.

Step 6: Apply the change of variables $\tilde{\mathbf{x}}_N(t) = \mathbf{T}_{bal}\mathbf{x}_N(t)$ to the full-order LTI system (2) to yield:

$$\begin{array}{ll} \dot{\tilde{\mathbf{x}}}_N(t) & = \mathbf{T}_{bal}\mathbf{A}\mathbf{T}_{bal}^{-1}\tilde{\mathbf{x}}_N(t) + \mathbf{T}_{bal}\mathbf{B}\mathbf{u}_P(t), \\ \mathbf{y}_{QN}(t) & = \mathbf{C}\mathbf{T}_{bal}^{-1}\tilde{\mathbf{x}}_N(t). \end{array} \tag{31}$$

Step 7: Partition $\tilde{\mathbf{A}} \equiv \mathbf{T}_{bal}\mathbf{A}\mathbf{T}_{bal}^{-1}$, $\tilde{\mathbf{B}} \equiv \mathbf{T}_{bal}\mathbf{B}$, $\tilde{\mathbf{C}} \equiv \mathbf{C}\mathbf{T}_{bal}^{-1}$ as follows:

$$\tilde{\mathbf{A}} = \left(\begin{array}{c|c} \tilde{\mathbf{A}}_{11} & \tilde{\mathbf{A}}_{12} \\ \hline \tilde{\mathbf{A}}_{21} & \tilde{\mathbf{A}}_{22} \end{array}\right), \quad \tilde{\mathbf{B}} = \left(\begin{array}{c} \tilde{\mathbf{B}}_1 \\ \tilde{\mathbf{B}}_2 \end{array}\right), \quad \tilde{\mathbf{C}} = \left(\begin{array}{c|c} \tilde{\mathbf{C}}_1 & \tilde{\mathbf{C}}_2 \end{array}\right), \tag{32}$$

where 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.

Step 8: Return the ROM system for a ROM of size M, given by:

$$\begin{array}{ll} \dot{\mathbf{x}}_M(t) & = \mathbf{A}_M\mathbf{x}_M(t) + \mathbf{B}_M\mathbf{u}_P(t), \\ \mathbf{y}_{QM}(t) & = \mathbf{C}_M\mathbf{x}_M(t), \end{array} \tag{33}$$

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:

$$\boldsymbol{\Psi}_M = \mathbf{T}_{bal}^T(:, 1:M), \quad \boldsymbol{\Phi}_M = \mathbf{S}_{bal}(:, 1:M), \tag{34}$$

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

In practice, the transformation matrices (30) are typically computed as:

$$\mathbf{T}_{bal} = \mathbf{V}^T\mathbf{Z}^T, \quad \mathbf{T}_{bal}^{-1} = \mathbf{U}\mathbf{W}, \tag{35}$$

where $\mathbf{Z}$ is the Cholesky factor of the observability Gramian

$$\mathbf{Q} = \mathbf{Z}\mathbf{Z}^T, \tag{36}$$

and $\mathbf{W}$ is the left singular vector of $\mathbf{U}^T\mathbf{Z}$, that is,

$$\mathbf{U}^T\mathbf{Z} = \mathbf{W}\mathbf{\Sigma}\mathbf{V}^T. \tag{37}$$

This is due to numerical stability issues that could arise in compuing $\mathbf{\Sigma}^{-1/2}$ in (30).

In effect, balanced truncation is a method for computing the test and trial bases $\mathbf{\Psi}_M$ and $\mathbf{\Phi}_M$ in (11). Given the test and trial bases defined in (34), the ROM system matrices (33) can be computed from the formulas (12).

The reader can verify that the reachability and observability Gramians satisfy the following property:

$$\mathbf{T}_{bal}\mathbf{P}\mathbf{T}_{bal}^T = \mathbf{T}_{bal}^{-T}\mathbf{Q}\mathbf{T}_{bal}^{-1} = \mathbf{\Sigma} = \text{diag}\{\sigma_1, \ldots, \sigma_N\}. \tag{38}$$

The entries of the diagonal matrix $\mathbf{\Sigma}$ are known as the *Hankel singular values* of the system (2). Assuming a ROM of size M has been constructed using balanced truncation (Algorithm 2 above), the following error bound on the output can be shown [48]:

$$||\mathbf{y}_{QN}(t) - \mathbf{y}_{QM}(t)||_2 \leq 2 \sum_{i=M+1}^{N} \sigma_i ||\mathbf{u}_P(t)||_2. \tag{39}$$

As with POD, the size of the reduced basis used in model reduction via balanced truncation is typically determined from an energy criterion. That is, M is selected to be the minimum integer such that

$$E_{BT}(M) \geq \text{tol} \tag{40}$$

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

$$E_{BT}(M) \equiv \frac{\sum_{i=1}^{M} \sigma_i}{\sum_{i=1}^{N} \sigma_i}. \tag{41}$$

POD vs. Balanced Truncation

Physically, POD modes maximize the average energy in the projection of the snapshot data onto the subspace spanned by these modes. Although POD modes can be very effective at approximating a given dataset, they may not describe all the dynamics in a particular dataset. In particular, these modes in general do not contain low-energy features of the dataset, features which may be critically important to the system dynamics [38]. Adding more POD modes to the reduced basis (increasing M) may actually decrease the accuracy

of the ROM [39]. Moreover, POD ROMs usually lack robustness with respect to parameter changes: it is in general unknown how well a POD ROM will predict the dynamics of a system with parameters different than those in the snapshot set used to generate the POD basis. Similarly, it is unknown how well the ROM will predict the system dynamics at times beyond the final snapshot collection time [4]. A final limitation of POD Galerkin ROMs is that they lack in general a stability guarantee. That is, a POD ROM (11) constructed for a stable LTI system (2) may produce solutions that are unbounded as $t \to \infty$ [43, 9, 13, 27].

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 [38]. 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 [39]. This is due to the high computational cost of solving the Lyapunov equations (26) and (27) 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 (26) and (27) in dense form, and hence require $\mathcal{O}(N^2)$ storage [19]. Unlike POD, balanced truncation delivers ROMs that preserve stability of a stable system (2) [36], however.

3 Stability Definitions

As stated in the Introduction, the objective of this work is to present and unify some model reduction techniques that have a stability guarantee. Before beginning this task, some general definitions of stability are reviewed.

3.1 Energy-Stability

The definition of stability considered in the present work is known as "energy-stability" [23, 21, 39]. The concept of energy-stability originated in the literature involving the numerical analysis of spectral discretizations to time-dependent PDEs [23, 21]. It has also appeared in the Galerkin finite element method literature, e.g., [20, 32], 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.* [38] and Kwasniok [31]. In [38], Rowley *et al.* introduced a family of "energy-based" inner products for the purpose of constructing stable Galerkin ROMs for fluid problems. In [31], 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 conserve turbulent kinetic energy or turbulent enstrophy.

Consider, without loss of generality, the following scalar initial value problem, known as a Cauchy problem:

$$\begin{array}{ll} \frac{\partial u}{\partial t} & = \mathcal{L}u, \quad \mathbf{x} \in \mathbb{R}^n, \quad t \geq 0 \\ u(\mathbf{x}, 0) & = f(\mathbf{x}). \end{array} \tag{42}$$

Here, $\mathcal{L}$ denotes a linear differential operator with constant coefficients. 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 $w(x) \in L^2$:

$$(w, \mathcal{L}w) \leq \alpha(w, w), \tag{43}$$

where $\alpha \in \mathbb{R}$. The following theorem (quoted from [29]) states the conditions under which the Cauchy problem (42) is well-posed.

Theorem 3.1.1 [29]: The Cauchy problem (42) 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 (42), denoted here by u_N, and satisfying

$$\left(\frac{\partial u_N}{\partial t}, \phi\right) = (\mathcal{L}u_N, \phi), \tag{44}$$

for all ϕ sufficiently smooth, and suppose $\mathcal{L}$ is semi-bounded with respect to $(\cdot,\cdot)$. Setting $\phi = u_N$ in (44) leads to the following energy estimate for the Galerkin approximation:

$$\frac{dE_N}{dt} \leq 2\alpha E_N, \tag{45}$$

where

$$E_N \equiv \frac{1}{2}||u_N||^2 \tag{46}$$

denotes the energy of the Galerkin approximation u_N, and $||\cdot||$ is the norm induced by the inner product $(\cdot,\cdot)$. Applying Gronwall's lemma to (49) gives the inequality

$$||u_N(\mathbf{x},t)|| \leq e^{\frac{1}{2}\alpha t}||u_N(\mathbf{x},0)||. \tag{47}$$

The result (47) says that the energy of the numerical solution to (44) is bounded in a way that is consistent with the behavior of the energy of the exact solution to the original differential equation (42), i.e., it is *energy-stable*.

This definition can be extended to a ROM LTI system of the form (11).

Definition 3.1.2 (Energy-Stability [21]): A ROM LTI system (11) is called energy-stable if

$$E_M(t) \leq e^{\alpha t}E_M(0), \tag{48}$$

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

$$E_M \equiv \frac{1}{2}||\mathbf{x}_M||^2 \tag{49}$$

is the system energy of the ROM numerical solution $\mathbf{x}_M$ to (11), and $||\cdot||$ is a norm equivalent to the L^2 norm.

In general, a ROM LTI system (11) 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 (48) holds for the ROM LTI system through a careful selection of the trial and test bases $\mathbf{\Phi}_M$ and $\mathbf{\Psi}_M$ and/or the inner product in which the projection step of the model reduction is performed (Sections 4 and 5).

3.2 Time-Stability

Having defined energy-stability, some discussion of how this concept relates to other common definitions of stability is in order. A common definition of stability is "time-stability". A numerical solution is said to be time-stable if it remains bounded as $t \to \infty$. The following is a more precise definition of time-stability.

Definition 3.2.1 (Time-Stability [21]): A ROM LTI system (11) 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

$$\frac{dE_N}{dt} \leq 0. \tag{50}$$

It is straightforward to demonstrate that a scheme that is time-stable is energy-stable. Suppose an LTI ROM (11) is time-stable, so (50) holds. Applying Gronwall's lemma to this inequality, $E_N(t) \leq E_N(0)$. Thus, (48) 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 that these governing PDEs support instabilities. In this case, an energy-stable ROM may possess physical unbounded solutions as $t \to \infty$, as (it can be argued) it should, if these unbounded solutions are physical.

3.3 Lyapunov Stability

The concept of energy-stability may be related to another concept of stability, namely Lyapunov stability.

Consider a system:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^n, \tag{51}$$

where $\mathbf{f} \in \mathbb{R}^n$ is a given function, subject to some initial condition $\mathbf{x}(0) = \mathbf{x}_0$. Suppose $\mathbf{x} = \mathbf{0}$ be an equilibrium point of (51). The following theorem, known as the Lyapunov stability theorem [6] characterizes the stability of this point.

Theorem 3.3.1 (Lyapunov Stability Theorem): 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 (51). Let $B_r = B_r(\mathbf{0})$ be a ball of radius r around the origin. 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 $\mathbf{x} = \mathbf{0}$ is locally stable in the sense of Lyapunov.

The function V defined in Theorem 3.3.1 above is known as the *Lyapunov function* for the system (51). Observe that the numerical energy E_N defined in (49) satisfies the definition of a Lyapunov function (Theorem 3.3.1) if (50) holds. Thus, if an LTI ROM (2) is energy-stable with $\alpha = 0$ (Definition 3.1.2), then the ROM is stable in the sense of Lyapunov.

A corollary of Theorem 3.3.1 is the following.

Corollary 3.3.2 [16]: An LTI system (2) is stable in the sense of Lyapunov if and only if all the eigenvalues of $\mathbf{A}$ have real parts less than or equal to 0, and those with real parts equal to 0 are non-repeated.

Corollary 3.3.2 is commonly used to check numerically (*a posteriori*) the stability of an LTI system (2) or a ROM (11) constructed for an LTI system.

4 Stable Model Reduction for LTI Systems via Continuous Projection

In some recent journal articles by the authors [9, 27], a method for constructing energy-stable ROMs for linearized compressible inviscid flow using POD and the continuous projection method was proposed. In this section, the approach is generalized to PDE systems of the form:

$$\dot{\mathbf{q}} + \mathbf{A}_i \mathbf{q}_{,i} - \mathbf{K}_{ij} \mathbf{q}_{,ij} + \mathbf{C}\mathbf{q} = \mathbf{F}. \tag{52}$$

In (52), $\mathbf{q} \in \mathbb{R}^n$ denotes a vector of unknowns, $\mathbf{F} \in \mathbb{R}^n$ is a source term, $\mathbf{A}_i$, $\mathbf{K}_{ij}$ and $\mathbf{C}$ are $n \times n$ matrices, where $1 \leq i, j \leq d$, with d denoting the number of spatial dimensions. The matrices $\mathbf{A}_i$, $\mathbf{K}_{ij}$ and $\mathbf{C}$ could be a function of space, but they are assumed to be steady (not a function of time t). The notation $_{,i}$ denotes differentiation in space with respect to the i^{th} spatial direction, i.e., $\mathbf{q}_{,i} \equiv \frac{\partial \mathbf{q}}{\partial x_i}$, and the so-called Eisenstein notation (implied summation on repeated indices) has been employed. Most conservation laws, as well as many PDEs of physical interest, can be written in the form (52). If $\mathbf{K}_{ij} = \mathbf{0}\ \forall i, j$, (52) is known as a hyperbolic system [22]. An example of a system of this form is the linearized compressible Euler system. Otherwise, if $\mathbf{K}_{ij} \neq \mathbf{0}$, (52) is known as an incompletely parabolic system [22]. A canonical example of such a system is the linearized compressible Navier-Stokes system.

In Section 4.1, a change of variables for the system (52) is derived such that the L^2 inner product is the energy inner product for the system in these new variables. It is then shown that an energy-based inner product, referred to as the "symmetry inner product" in [9, 27], induces the desired transformation (Section 4.2). It is also shown that a Galerkin projection of (52) in the symmetry inner product may be viewed as a Petrov-Galerkin projection of the original equations in the L^2 inner product. An approach for deriving the stabilizing transformation using Lyapunov functions representing the total energy of the system (52) is outlined in Section 4.3. Examples of the symmetry inner product for several PDEs that can be written as (52) are given in Section 4.4. A stability-preserving discrete implementation of the projection of (52) in the symmetry inner product is outlined in Section 4.5. The stability properties of a POD ROM constructed using the continuous projection approach and the symmetry inner product are studied on a numerical example in Section 4.6.

4.1 A Stabilizing Transformation

Suppose there exists a transformation

$$\begin{array}{rccl} T: & \mathbb{R}^n & \rightarrow & \mathbb{R}^n, \\ & \mathbf{q} & \rightarrow & \mathbf{v}, \end{array} \tag{53}$$

such that in the new variables $\mathbf{v}$, the system (52) has the form

$$\dot{\mathbf{v}} + \mathbf{A}_i^S \mathbf{v}_{,i} - \mathbf{K}_{ij}^S \mathbf{v}_{,ij} + \mathbf{C}^S \mathbf{v} = \mathbf{F}^S, \tag{54}$$

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:

$$\mathbf{K}^S \equiv \begin{pmatrix} \mathbf{K}_{11}^S & \mathbf{K}_{12}^S & \mathbf{K}_{13}^S \\ \mathbf{K}_{21}^S & \mathbf{K}_{22}^S & \mathbf{K}_{23}^S \\ \mathbf{K}_{31}^S & \mathbf{K}_{32}^S & \mathbf{K}_{33}^S \end{pmatrix} \tag{55}$$

  is positive semi-definite.

Theorem 4.1.1: Suppose a ROM for (54) is constructed using continuous Galerkin projection in the $L_2(\Omega)$ inner product. Suppose the matrices in (54) satisfy Properties 1–3 above. Assume 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 way[2]. Let $\mathbf{v}_M$ denote the ROM solution to (54). Then the ROM is energy-stable with energy estimate

$$||\mathbf{v}_M(\cdot, T)||_2 \leq e^{\frac{1}{2}\beta T}||\mathbf{v}_M(\cdot, 0)||_2, \tag{56}$$

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

$$\mathbf{B} \equiv \frac{\partial \mathbf{A}_i^S}{\partial x_i} + \frac{\partial \mathbf{K}_{ij}^S}{\partial x_i \partial x_j} - 2\mathbf{C}^S. \tag{57}$$

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

Proof. To prove energy-stability of (54), it is necessary to bound the energy of the ROM solution to (54) with $\mathbf{F}^S = \mathbf{0}$. First, note that:

$$\mathbf{K}_{ij}^S \frac{\partial^2 \mathbf{v}_M}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i}\left(\mathbf{K}_{ij}^S \frac{\partial \mathbf{v}_M}{\partial x_j}\right) - \left(\frac{\partial \mathbf{K}_{ij}^S}{\partial x_i}\frac{\partial \mathbf{v}_M}{\partial x_j}\right). \tag{58}$$

Now:

$$\begin{aligned} \frac{dE_M}{dt} &= \frac{1}{2}\frac{d}{dt}||\mathbf{v}_M||_2^2 \\ &= \frac{1}{2}\frac{d}{dt}\left(\mathbf{v}_M, \mathbf{v}_M\right) \\ &= \left(\mathbf{v}_M, \frac{\partial \mathbf{v}_M}{\partial t}\right) \\ &= \left(\mathbf{v}_M, -\mathbf{A}_i^S \frac{\partial \mathbf{v}_M}{\partial x_i} + \mathbf{K}_{ij}^S \frac{\partial^2 \mathbf{v}_M}{\partial x_i \partial x_j} - \mathbf{C}^S \mathbf{v}_M\right) \\ &= -\int_\Omega \mathbf{v}_M^T \mathbf{A}_i^S \frac{\partial \mathbf{v}}{\partial x_i} \partial\Omega + \int_\Omega \mathbf{v}_M^T \mathbf{K}_{ij}^S \frac{\partial^2 \mathbf{v}_M}{\partial x_i \partial x_j} \partial\Omega - \int_\Omega \mathbf{v}_M^T \mathbf{C}^S \mathbf{v}_M \partial\Omega. \end{aligned} \tag{59}$$

Each of the terms in (59) will be bounded separately. First,

$$\begin{aligned} -\int_\Omega \mathbf{v}_M^T \mathbf{A}_i^S \frac{\partial \mathbf{v}_M}{\partial x_i} \partial\Omega &= -\frac{1}{2}\int_\Omega \frac{\partial}{\partial x_i}\left(\mathbf{v}_M^T \mathbf{A}_i^S \mathbf{v}_M\right) d\Omega + \frac{1}{2}\int_\Omega \mathbf{v}_M^T \frac{\partial \mathbf{A}_i^S}{\partial x_i}\mathbf{v} d\Omega \\ &= -\frac{1}{2}\int_{\partial\Omega} \mathbf{v}_M^T \mathbf{A}_i^S n_i \mathbf{v}_M dS + \frac{1}{2}\int_\Omega \mathbf{v}_M^T \frac{\partial \mathbf{A}_i^S}{\partial x_i}\mathbf{v}_M d\Omega. \end{aligned} \tag{60}$$

[2]The reader is referred to [27] 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.

In (60),the property that each of the matrices $\mathbf{A}_i^S$ is symmetric has been employed (Property 1).

Next,

$$\int_\Omega \mathbf{v}_M^T \mathbf{K}_{ij}^S \frac{\partial^2 \mathbf{v}_M}{\partial x_i \partial x_j} \partial\Omega \;\; = \int_\Omega \mathbf{v}_M^T \frac{\partial}{\partial x_i}\left(\mathbf{K}_{ij}^S \frac{\partial \mathbf{v}_M}{\partial x_j}\right) d\Omega - \int_\Omega \mathbf{v}_M^T \frac{\partial \mathbf{K}_{ij}^S}{\partial x_i} \frac{\partial \mathbf{v}_M}{\partial x_j} \partial\Omega. \tag{61}$$

Again, each of the two terms in (61) will be bounded separately.

$$\begin{array}{l} \int_\Omega \mathbf{v}_M^T \frac{\partial}{\partial x_i}\left(\mathbf{K}_{ij}^S \frac{\partial \mathbf{v}_M}{\partial x_j}\right) d\Omega \;\; = -\int_\Omega \frac{\partial \mathbf{v}_M}{\partial x_i}^T \mathbf{K}_{ij}^S \frac{\partial \mathbf{v}_M}{\partial x_j} + \int_{\partial\Omega} \mathbf{v}_M^T \mathbf{K}_{ij}^S \frac{\partial \mathbf{v}_M}{\partial x_j} n_i dS \\ \leq \int_{\partial\Omega} \mathbf{v}_M^T \mathbf{K}_{ij}^S \frac{\partial \mathbf{v}_M}{\partial x_j} n_i dS, \end{array} \tag{62}$$

provided the matrix (66) is positive semi-definite (Property 3).

Now for the second term in (61):

$$\begin{array}{rl} -\int_\Omega \mathbf{v}_M^T \frac{\partial \mathbf{K}_{ij}^S}{\partial x_i} \frac{\partial \mathbf{v}_M}{\partial x_j} \partial\Omega \;\; & = -\frac{1}{2}\int_\Omega \frac{\partial}{\partial x_j}\left(\mathbf{v}_M^T \frac{\partial \mathbf{K}_{ij}^S}{\partial x_i} \mathbf{v}_M\right) d\Omega + \frac{1}{2}\int_\Omega \mathbf{v}_M^T \frac{\partial \mathbf{K}_{ij}^S}{\partial x_i \partial x_j} \mathbf{v}_M d\Omega \\ & = -\frac{1}{2}\int_{\partial\Omega} \mathbf{v}_M^T \frac{\partial \mathbf{K}_{ij}^S}{\partial x_i} n_j \mathbf{v}_M dS + \frac{1}{2}\int_\Omega \mathbf{v}_M^T \frac{\partial \mathbf{K}_{ij}^S}{\partial x_i \partial x_j} \mathbf{v}_M d\Omega. \end{array} \tag{63}$$

In (63), the property that the $\mathbf{K}_{ij}^S$ matrices and therefore their derivatives are symmetric has been employed (Property 2).

Finally, (60) and (61) are substituted into (59). 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:

$$\begin{array}{rl} \frac{1}{2}\frac{d}{dt}||\mathbf{v}_M||_2^2 \;\; & \leq \frac{1}{2}\int_\Omega \mathbf{v}_M^T \left(\frac{\partial \mathbf{A}_i^S}{\partial x_i}\right) \mathbf{v}_M d\Omega + \frac{1}{2}\int_\Omega \mathbf{v}_M^T \frac{\partial \mathbf{K}_{ij}^S}{\partial x_i \partial x_j} \mathbf{v}_M d\Omega - \int_\Omega \mathbf{v}_M^T \mathbf{C}^S \mathbf{v}_M \partial\Omega \\ & = \frac{1}{2}\int_\Omega \mathbf{v}_M^T \mathbf{B} \mathbf{v}_M d\Omega, \end{array} \tag{64}$$

where $\mathbf{B}$ is given by (69). Applying Gronwall's inequality to (64), it is found that:

$$||\mathbf{v}_M(\cdot, T)||_2 \leq e^{\frac{1}{2}\beta T} ||\mathbf{v}_M(\cdot, 0)||_2, \tag{65}$$

where β is an upper bound on the eigenvalues of the matrix $\mathbf{B}$ (69).

□

Note that, if $\mathbf{C} = \mathbf{0}$ in (52) and the $\mathbf{A}_i$ and $\mathbf{K}_{ij}$ matrices are spatially-constant, it follows that $\beta = 0$ in (64). In this case, if the ROM for (52) 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 $\mathbf{C} = \mathbf{0}$ 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 (52) is that they are symmetrizable [23, 9, 27]; that is, it is possible to derive a symmetric positive-definite matrix $\mathbf{H}$ such that:

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

$$\mathbf{K}^H \equiv \begin{pmatrix} \mathbf{HK}_{11} & \mathbf{HK}_{12} & \mathbf{HK}_{13} \\ \mathbf{HK}_{21} & \mathbf{HK}_{22} & \mathbf{HK}_{23} \\ \mathbf{HK}_{31} & \mathbf{HK}_{32} & \mathbf{HK}_{33} \end{pmatrix} \tag{66}$$

  is positive semi-definite.

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

$$\left(\mathbf{q}^{(1)}, \mathbf{q}^{(2)}\right)_{(\mathbf{H},\Omega)} \equiv \int_\Omega \mathbf{q}^{(1)T} \mathbf{H} \mathbf{q}^{(2)} d\Omega. \tag{67}$$

Following the terminology introduced in [9, 27], the inner product (67) 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 (52) is constructed using continuous Galerkin projection in the symmetry inner product (67). Assume Properties 1*-3* hold. Assume 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 (52). Then the ROM is energy-stable with energy estimate

$$||\mathbf{q}_M(\cdot, T)||_{(\mathbf{H},\Omega)} \leq e^{\frac{1}{2}\beta T} ||\mathbf{q}_M(\cdot, 0)||_{(\mathbf{H},\Omega)}, \tag{68}$$

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

$$\mathbf{B} \equiv \frac{\partial(\mathbf{HA}_i)}{\partial x_i} + \frac{\partial(\mathbf{HK}_{ij})}{\partial x_i \partial x_j} - 2\mathbf{HC}. \tag{69}$$

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

Proof. Analogous to the proof of Theorem 4.1.1.

□

Again, in the case that $\mathbf{C} = \mathbf{0}$ 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 (67) 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 (52) in the symmetry inner product (67) is equivalent to a Petrov-Galerkin projection. Let $\boldsymbol{\phi}_i$ for $i = 1, ..., M$ denote the trial reduced basis vector for the solution $\mathbf{q}$. Performing a Galerkin projection of the equations (52) onto the modes $\boldsymbol{\phi}_i$ gives

$$\int_\Omega \boldsymbol{\phi}_i^T \mathbf{H} \left(\dot{\mathbf{q}} + \mathbf{A}_i \mathbf{q}_{,i} + \mathbf{K}_{ij} \mathbf{q}_{,ij} + \mathbf{C}\mathbf{q} \right) d\Omega = \int_\Omega \boldsymbol{\phi}_i^T \mathbf{H}\mathbf{F} d\Omega, \tag{70}$$

for $i = 1, ..., M$. (70) is equivalent to a Petrov-Galerkin projection of the equations (52) in the regular L_2 inner product

$$\int_\Omega \boldsymbol{\psi}_i^T \left(\dot{\mathbf{q}} + \mathbf{A}_i \mathbf{q}_{,i} + \mathbf{K}_{ij} \mathbf{q}_{,ij} + \mathbf{C}\mathbf{q} \right) d\Omega = \int_\Omega \boldsymbol{\psi}_i^T \mathbf{F} d\Omega, \tag{71}$$

where the test reduced basis functions are given by $\boldsymbol{\psi}_i = \mathbf{H}\boldsymbol{\phi}_i$, for all $i = 1, ..., M$.

4.3 Lyapunov Function Connection

The stabilizing transformation described in Section 4.1 can be found using Lyapunov function theory. By the Lyapunov Stability Theorem (Theorem 3.3.1), if an equilibrium point of a system is stable, there exists a non-negative function that is always decreasing along the system trajectories – the Lyapunov function $V = V(\mathbf{q})$ [6]. In the case the governing system conserves energy or is dissipative, the total energy of the system is a Lyapunov function $V(\mathbf{q})$ for the system [6, 38]. Hence, a transformation is sought such that

$$E_T = V(\mathbf{q}) = ||\mathbf{v}||_E^2, \tag{72}$$

where E_T denotes the total energy of the system, and $|| \cdot ||_E$ is some norm equivalent to the L^2 norm (commonly referred to as the "energy norm" [38, 9, 27]). Consider, for example, the compressible Euler equations in two dimensions (2D):

$$\begin{aligned} \frac{\partial \rho}{\partial t} + u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y} + \rho\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) &= 0, \\ \rho\left(\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial v}{\partial y}\right) + \frac{\partial p}{\partial x} &= 0, \\ \rho\left(\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y}\right) + \frac{\partial p}{\partial y} &= 0, \\ \rho\left(\frac{\partial e}{\partial t} + u\frac{\partial e}{\partial x} + v\frac{\partial e}{\partial y}\right) + \rho e\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \frac{\partial (up)}{\partial x} + \frac{\partial (vp)}{\partial y} &= 0, \end{aligned} \tag{73}$$

in some open bounded domain $\Omega \in \mathbb{R}^2$. Here, ρ denotes the fluid density, u and v are the fluid velocities, e is the internal energy per unit mass of the fluid, and γ is the ratio of specific heats. The total energy of the fluid system (73) is given by

$$E_T = \int_\Omega \left\{ \rho e + \frac{1}{2}\rho(u^2 + v^2) \right\} d\Omega \tag{74}$$

Now, define the following transformation from the variables $\mathbf{q}$ to the variables $\mathbf{v}$:

$$\mathbf{q} \equiv \begin{pmatrix} \rho \\ u \\ v \\ e \end{pmatrix} \rightarrow \begin{pmatrix} a \\ u \\ v \\ c \end{pmatrix} \equiv \mathbf{v}, \tag{75}$$

where

$$a^2 \equiv \rho, \quad c^2 \equiv \gamma(\gamma-1)e, \tag{76}$$

as well as the following inner product:

$$\left(\mathbf{v}^{(1)}, \mathbf{v}^{(2)}\right)_E \equiv \int_\Omega \left(\frac{1}{\gamma(\gamma-1)} a^{(1)} c^{(1)} a^{(2)} c^{(2)} + \frac{1}{2} a^{(1)} a^{(2)} \left[u^{(1)} u^{(2)} + v^{(1)} v^{(2)} \right] \right) d\Omega. \tag{77}$$

The norm induced by the inner product (77) is:

$$\begin{aligned} ||\mathbf{v}||_E^2 &\equiv (\mathbf{v}, \mathbf{v})_E \\ &= \int_\Omega \left(\tfrac{1}{\gamma(\gamma-1)} a^2 c^2 + \tfrac{1}{2} a^2 \left[u^2 + v^2 \right] \right) d\Omega \\ &= \int_\Omega \left(\rho e + \tfrac{1}{2} \rho [u^2 + v^2] \right) d\Omega \\ &= E_T. \end{aligned} \tag{78}$$

E_T is a Lyapunov function for the system (73), satisfying the conditions given in Theorem 3.3.1. In particular, $\frac{dE_t}{dt} \leq 0$. Hence, if a ROM for (73) is constructed in the $\mathbf{v}$ variables using the inner product (77), this ROM will be time-stable, energy-stable and stable in the sense of Lyapunov.

It was shown in Section 4.2 that the symmetry inner product (67) is the energy inner product for a linear system of PDEs of the form (52).

4.4 Examples

Example 1: Wave Equation

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

$$\ddot{u} = a^2 u_{,xx} \tag{79}$$

where $a \in \mathbb{R}$ denotes the wave speed, and $\ddot{u} \equiv \frac{\partial^2 u}{\partial t^2}$. (79) is a canonical PDE of the hyperbolic type. Remark that (79) can be written as a first order system

$$\dot{\mathbf{q}} = \mathbf{A}\mathbf{q}_{,x}, \tag{80}$$

where

$$\mathbf{q} = \begin{pmatrix} \dot{u} \\ u_{,x} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 0 & a^2 \\ 1 & 0 \end{pmatrix}. \tag{81}$$

By Corollary 4.2.1, a matrix that symmetrizes $\mathbf{A}$ is sought. Remark that if

$$\mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & a^2 \end{pmatrix}, \tag{82}$$

the matrix $\mathbf{HA}$ is symmetric [44]. It follows that to ensure energy-stability, a ROM for (79) should be constructed by projecting the system (80) onto the reduced basis modes in the symmetry inner product (67) with $\mathbf{H}$ given by (82).

Example 2: Linearized Shallow Water Equations

Consider the linearized form of the shallow water equations:

$$\dot{\mathbf{q}}' + \mathbf{A}_i \mathbf{q}'_{,i} + \mathbf{C}\mathbf{q}' = \mathbf{0}. \tag{83}$$

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.,

$$\mathbf{q}(\mathbf{x}, t) + \bar{\mathbf{q}}(\mathbf{x}) + \mathbf{q}'(\mathbf{x}, t) \tag{84}$$

and linearizing the full shallow water equations around the steady mean state $\bar{\mathbf{q}}$. If $\mathbf{q}^T = \begin{pmatrix} u, & v, & w, & \phi \end{pmatrix}$, then the convective flux matrices in the hyperbolic system (83) in three-dimensions (3D) are given by:

$$\mathbf{A}_1 = \begin{pmatrix} \bar{u} & 0 & 0 & 1 \\ 0 & \bar{u} & 0 & 0 \\ 0 & 0 & \bar{u} & 0 \\ \bar{\phi} & 0 & 0 & \bar{u} \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} \bar{v} & 0 & 0 & 0 \\ 0 & \bar{v} & 0 & 1 \\ 0 & 0 & \bar{v} & 0 \\ 0 & \bar{\phi} & 0 & \bar{v} \end{pmatrix}, \quad \mathbf{A}_3 = \begin{pmatrix} \bar{w} & 0 & 0 & 0 \\ 0 & \bar{w} & 0 & 0 \\ 0 & 0 & \bar{w} & 1 \\ 0 & 0 & \bar{\phi} & \bar{w} \end{pmatrix}, \tag{85}$$

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 [44]. Remark that each of the convective flux matrices (85) can be symmetrized by the matrix

$$\mathbf{H} = \begin{pmatrix} \bar{\phi} & 0 & 0 & 0 \\ 0 & \bar{\phi} & 0 & 0 \\ 0 & 0 & \bar{\phi} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{86}$$

From Corollary 4.2.1, the equations (83) should be projected onto the reduced basis modes in the symmetry inner product (67) with $\mathbf{H}$ given by (86) to guarantee an energy-stable ROM.

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 (84) and linearizing these equations around the steady mean state $\bar{\mathbf{q}}$. If $\mathbf{q}^T = \begin{pmatrix} u, & v, & w, & \zeta, & p \end{pmatrix}$, 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 (83). In 3D, the convective flux matrices $\mathbf{A}_i$ in the linearized compressilbe Euler hyperbolic system (83) are given by:

$$\mathbf{A}_1 = \begin{pmatrix} \bar{u} & 0 & 0 & 0 & \bar{\zeta} \\ 0 & \bar{u} & 0 & 0 & 0 \\ 0 & 0 & \bar{u} & 0 & 0 \\ -\bar{\zeta} & 0 & 0 & \bar{u} & 0 \\ \gamma\bar{p} & 0 & 0 & 0 & \bar{u} \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} \bar{v} & 0 & 0 & 0 & 0 \\ 0 & \bar{v} & 0 & 0 & \bar{\zeta} \\ 0 & 0 & \bar{v} & 0 & 0 \\ 0 & -\bar{\zeta} & 0 & \bar{v} & 0 \\ 0 & \gamma\bar{p} & 0 & 0 & \bar{v} \end{pmatrix}, \quad \mathbf{A}_3 = \begin{pmatrix} \bar{w} & 0 & 0 & 0 & 0 \\ 0 & \bar{w} & 0 & 0 & 0 \\ 0 & 0 & \bar{w} & 0 & \bar{\zeta} \\ 0 & 0 & -\bar{\zeta} & \bar{w} & 0 \\ 0 & 0 & \gamma\bar{p} & 0 & \bar{w} \end{pmatrix}. \tag{87}$$

Here, $\gamma = C_P/C_V$ is the ratio of specific heats. The reader may verify that if the linearized compressible Euler system (83) is pre-multiplied by the following symmetric positive definite matrix:

$$\mathbf{H} = \begin{pmatrix} \bar{\rho} & 0 & 0 & 0 & 0 \\ 0 & \bar{\rho} & 0 & 0 & 0 \\ 0 & 0 & \bar{\rho} & 0 & 0 \\ 0 & 0 & 0 & \alpha^2\gamma\bar{\rho}^2\bar{p} & \bar{\rho}\alpha^2 \\ 0 & 0 & 0 & \bar{\rho}\alpha^2 & \frac{(1+\alpha^2)}{\gamma\bar{p}} \end{pmatrix}, \tag{88}$$

where α is a real, non-zero parameter to yield the system, the convective flux matrices $\mathbf{HA}_i$ are all symmetric [9, 27]. It follows that if a ROM for the linearized compressible Euler equations is constructed in the symmetry inner product (67) with $\mathbf{H}$ given by (88), the resulting ROM will be energy-stable.

Example 4: Linearized Compressible Navier-Stokes Equations

Consider the 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 (84), and a linearization around the steady mean is performed. If the viscous work terms are neglected from the equations[3] (appropriate, for example, in a low Mach number regime), the result is a linear incompletely parabolic system of the form (52). If $\mathbf{q}^T = \begin{pmatrix} u, & v, & w, & T, & \rho \end{pmatrix}$, where T and ρ denote the fluid temperature

[3]A survey of the literature reveals that the viscous work terms are invariably neglected from the linearized compressible Navier-Stokes equations by authors studying energy-stability of these equations [23, 2]. 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 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.

and density respectively, the convective and viscous flux matrices that appear in (52) are given by:

$$\mathbf{A}_1 = \begin{pmatrix} \bar{u} & \bar{\rho} & 0 & R & \frac{R\bar{T}}{\bar{\rho}} \\ 0 & \bar{u} & 0 & 0 & 0 \\ 0 & 0 & \bar{u} & 0 & 0 \\ \bar{T}(\gamma-1) & 0 & 0 & \bar{u} & 0 \\ \bar{\rho} & 0 & 0 & 0 & \bar{u} \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} \bar{v} & 0 & 0 & 0 & 0 \\ 0 & \bar{v} & 0 & R & \frac{R\bar{T}}{\bar{\rho}} \\ 0 & 0 & \bar{v} & 0 & 0 \\ 0 & \bar{T}(\gamma-1) & 0 & \bar{v} & 0 \\ 0 & \bar{\rho} & 0 & 0 & \bar{v} \end{pmatrix}, \tag{89}$$

$$\mathbf{A}_3 = \begin{pmatrix} \bar{w} & 0 & 0 & 0 & 0 \\ 0 & \bar{w} & 0 & 0 & 0 \\ 0 & 0 & \bar{w} & R & \frac{R\bar{T}}{\bar{\rho}} \\ 0 & 0 & \bar{T}(\gamma-1) & \bar{w} & 0 \\ 0 & 0 & 0 & \bar{\rho} & \bar{w} \end{pmatrix}, \quad \mathbf{K}_{11} = \frac{1}{\bar{\rho}Re} \begin{pmatrix} 2\mu+\lambda & 0 & 0 & 0 & 0 \\ 0 & \mu & 0 & 0 & 0 \\ 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & \frac{\gamma\kappa}{Pr} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{90}$$

$$\mathbf{K}_{12} \equiv \frac{1}{2\bar{\rho}Re} \begin{pmatrix} 0 & \lambda+\mu & 0 & 0 & 0 \\ \mu+\lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{K}_{13} \equiv \frac{1}{2\bar{\rho}Re} \begin{pmatrix} 0 & 0 & \lambda+\mu & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \mu+\lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{91}$$

$$\mathbf{K}_{21} \equiv \frac{1}{2\bar{\rho}Re} \begin{pmatrix} 0 & \mu+\lambda & 0 & 0 & 0 \\ \lambda+\mu & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{K}_{22} \equiv \frac{1}{\bar{\rho}Re} \begin{pmatrix} \mu & 0 & 0 & 0 & 0 \\ 0 & 2\mu+\lambda & 0 & 0 & 0 \\ 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & \frac{\gamma\kappa}{Pr} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{92}$$

$$\mathbf{K}_{23} \equiv \frac{1}{2\bar{\rho}Re} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda+\mu & 0 & 0 \\ 0 & \mu+\lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{K}_{31} \equiv \frac{1}{2\bar{\rho}Re} \begin{pmatrix} 0 & 0 & \mu+\lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \lambda+\mu & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{93}$$

$$\mathbf{K}_{32} \equiv \frac{1}{2\bar{\rho}Re} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \mu+\lambda & 0 & 0 \\ 0 & \lambda+\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{K}_{33} \equiv \frac{1}{\bar{\rho}Re} \begin{pmatrix} \mu & 0 & 0 & 0 & 0 \\ 0 & \mu & 0 & 0 & 0 \\ 0 & 0 & 2\mu+\lambda & 0 & 0 \\ 0 & 0 & 0 & \frac{\gamma\kappa}{Pr} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{94}$$

Here μ and λ are the so-called Lamé constants, Re is the Reynolds number, Pr is the Prandtl number, R is the universal gas constant, and γ is the ratio of specific heats. The reader can verify that if the system (52) is pre-multiplied by the symmetric positive definite matrix

given by

$$\mathbf{H} \equiv \begin{pmatrix} \bar{\rho} & 0 & 0 & 0 & 0 \\ 0 & \bar{\rho} & 0 & 0 & 0 \\ 0 & 0 & \rho & 0 & 0 \\ 0 & 0 & 0 & \frac{\bar{\rho}R}{\bar{T}(\gamma-1)} & 0 \\ 0 & 0 & 0 & 0 & \frac{R\bar{T}}{\bar{\rho}} \end{pmatrix}, \tag{95}$$

the "symmetrized" convective flux matrices $\mathbf{HA}_i$ and diffusive flux matrices $\mathbf{HK}_{ij}$ satisfy Properties 1*–3* in Section 4.2. It follows that if a ROM for the linearized compressible Navier-Stokes equations (52) is constructed in the symmetry inner product (67) with $\mathbf{H}$ given by (95), the resulting ROM will be energy-stable.

Note that the symmetry transformations in the examples above are not unique. For example, in [2], Abarbanel *et al.* exhibit a transformation of the form (54) for the linearized compressible Navier-Stokes equations written in the primitive variables $\mathbf{q}^T = \begin{pmatrix} \rho, & u, & v, & w, & p \end{pmatrix}$.

4.5 Stability-Preserving Discrete Implementation

The stability analysis of the preceding subsections has assumed that the integrals resulting from the projection of the governing equations onto the reduced basis modes are evaluated exactly in continuous form. At first glance, it appears there may be a problem translating this continuous result to the discrete setting. This apparent difficulty is reminiscent of a similar problem that appears in spectral methods, where spectral projections need to be computed exactly. The problem is resolved in a similar way as in the spectral method context, namely through the use of high-precision numerical quadrature. 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 all continuous inner products arising from the continuous Galerkin projection of the equations onto the POD modes.

More specifically, 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 nn nodes. Then, the finite element representation of the vector $\mathbf{q}$ in (52) in each element Ω_e is:

$$\mathbf{q}_e^h = \sum_{i=1}^{nn} N_i(\mathbf{x})\mathbf{q}(\mathbf{x}), \quad \mathbf{x} \in \Omega^e. \tag{96}$$

For examples 1–4 above, it is necessary to compute numerically integrals of the form:

$$\left(\mathbf{q}^{(1)}, \mathbf{q}^{(2)}\right)_{(\mathbf{H},\Omega)} = \int_\Omega \mathbf{q}^{(1)T}\mathbf{H}\mathbf{q}^{(2)} d\Omega. \tag{97}$$

Suppose, without loss of generality, that the finite element shape functions are chosen to be bilinear, so $nn = 4$. 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 (97) can be written as

$$\left(\mathbf{q}^{(1)}, \mathbf{q}^{(2)}\right)_{(\mathbf{H},\Omega)} = \mathbf{q}^{h(1)T} \mathbf{W} \mathbf{q}^{h(2)}, \tag{98}$$

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

$$w_{kl} = \sum_{e=1}^{n_{kl}^{el}} \mathbf{H}_e^h \sum_{j=1}^{4} N_{k_e}(\mathbf{x}_{j_e}) N_{l_e}(\mathbf{x}_{j_e}) \omega_{j_e}. \tag{99}$$

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. The sparsity structure of a representative $\mathbf{W}$ matrix (98) for a problem with four degrees of freedom per node (such as Example 2 in 3D, or Examples 3-4 in 2D) is shown in Figure 1.

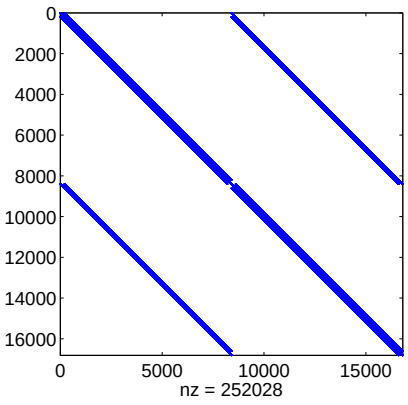


Figure 1. Sparsity structure of representative $\mathbf{W}$ matrix (98)

The introduction of C^0 finite elements, namely bilinears, requires a relaxation of the smoothness requirements on $\mathbf{q}$, $\mathbf{H}$, $\mathbf{A}_i$, and $\mathbf{K}_{ij}$. The projection integrals are then to be interpreted in the sense of distributions. Higher order finite element representations of the POD modes and snapshots are possible. If these are to be employed, the order of the quadrature rule must be increased to ensure that no error is introduced into the numerical computation of the relevant inner products.

4.6 Numerical Experiment

The test case considered here 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 4.4). For this problem, the base flow is uniform, with the following values:

$$\begin{gathered}\bar{p} = 101,325 \text{ Pa} \\ \bar{T} = 300 \text{ K} \\ \bar{\rho} = \frac{\bar{p}}{R\bar{T}} = 1.17 \text{ kg/m}^3 \\ \bar{u}_1 = \bar{u}_2 = 0.0 \text{ m/s} \\ \bar{c} = 348.0 \text{ m/s}.\end{gathered} \tag{100}$$

In (100), $\bar{c} \equiv \sqrt{\gamma R\bar{T}}$ is the mean speed of sound. The problem is initialized with a pressure pulse in the middle of the domain:

$$\begin{gathered}p'(\mathbf{x};0) = 141.9e^{-10(x^2+y^2)}, \\ \rho'(\mathbf{x};0) = \frac{p'(\mathbf{x};0)}{R\bar{T}}, \\ T'(\mathbf{x};0) = 0, \\ u_1'(\mathbf{x};0) = u_2'(\mathbf{x};0) = 0.\end{gathered} \tag{101}$$

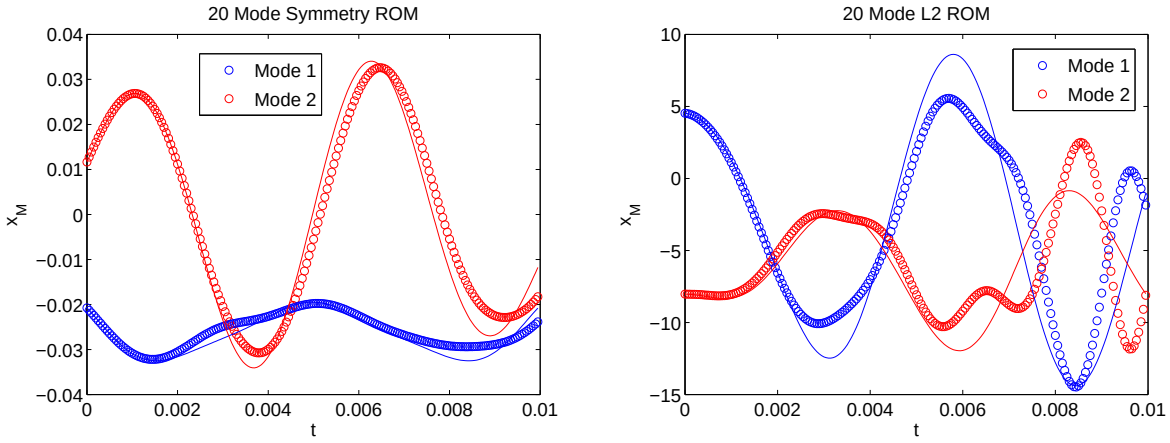
In terms of the mean values, the amplitude of the initial pressure pulse (101) is $0.001\bar{\rho}\bar{c}^2$. As both the high-fidelity code as well as the ROM code are 3D codes, a 2D mesh of the domain Ω is converted to a 3D mesh by extruding the 2D mesh in the z-direction by one element. The computational 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 the $z =$ constant boundaries in the high-fidelity code. The high-fidelity CFD simulation from which the ROM is generated is performed until time $T = 0.01$ seconds. During this simulation, the initial pressure pulse (101) 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 employed to construct a 20 mode POD basis. Two different procedures were used to generate a fluid ROM for this problem: the POD/Galerkin method with the symmetry inner product (67) with $\mathbf{H}$ given by (88), and the POD/Galerkin method with the classical L^2 inner product. Using both the symmetry and the L^2 inner product, the POD modes captured essentially 100% of the snapshot energy. Since the base flow for this example is uniform (100), $\mathbf{C} = \mathbf{0}$ and $\mathbf{A}_i$ and $\mathbf{K}_{ij}$ are spatially-constant in (52), meaning an energy-stable ROM is expected to be time-stable and stable in the sense of Lyapunov.

Figure 2 shows a time history of the first two ROM modal amplitudes (circles) compared to the projection of the full 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:

$$x_{M,i}(t) \quad \text{vs.} \quad \left(\mathbf{q}'_{\text{CFD}}, \boldsymbol{\phi}_i\right)_{(\mathbf{H},\Omega)}, \quad i = 1, 2, \tag{102}$$

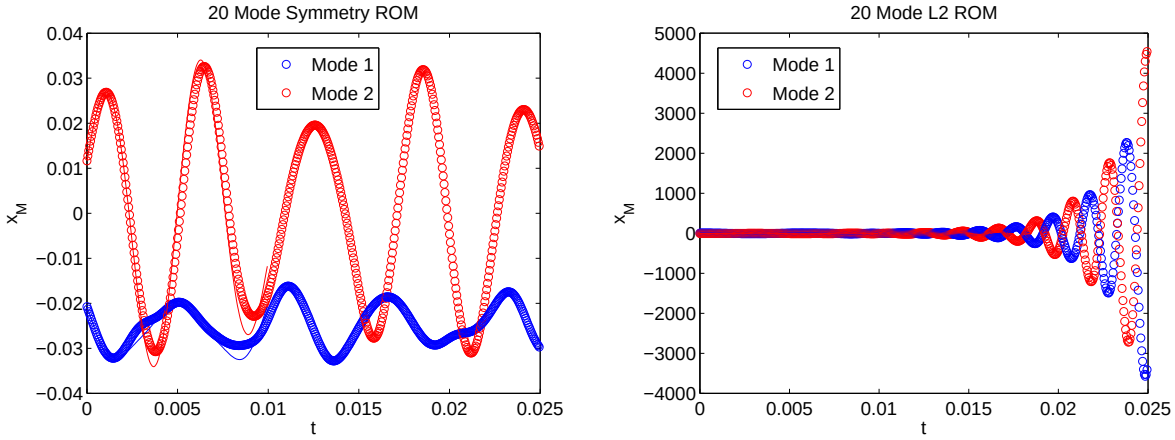
where $\mathbf{q}'_{\text{CFD}}$ is the high-fidelity CFD solution from which the ROMs were constructed. The reader may observe agreement between the symmetry ROM and the full simulation (Figure 2(a)) for the time interval considered. In contrast, agreement between the L^2 ROM and the

full simulation is reasonable only until approximately $t = 0.005$ seconds (Figure 2(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 3), the instability in the L^2 ROM is apparent.



(a) 20 mode symmetry ROM **(b)** 20 mode L^2 ROM

Figure 2. Time history of modal amplitudes for inviscid pressure pulse problem

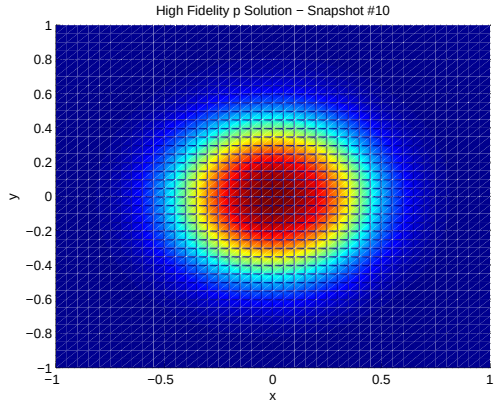


(a) 20 mode symmetry ROM **(b)** 20 mode L^2 ROM

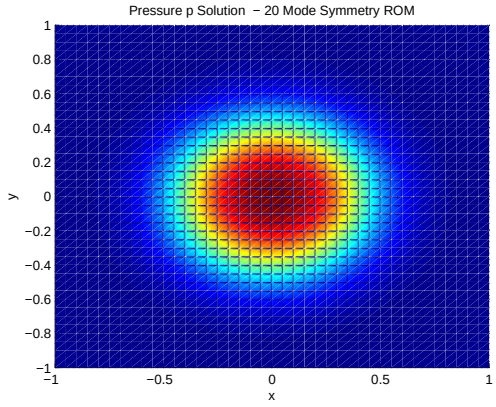
Figure 3. Time history of modal amplitudes for inviscid pressure pulse problem for longer time horizon

Figures 4–6 compare the CFD pressure field (a) with the field reconstructed from the symmetry (b) and L^2 (c) ROM solutions at times $t = 4.5 \times 10^{-4}$, 2.95×10^{-3} 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 4). At the later times, $t = 2.95 \times 10^{-3}$ and 7.95×10^{-3} seconds, there is a good qualitative agreement between the high-fidelity solution and the symmetry ROM solution (Figures 5–6(a), (b)). The same cannot be said of

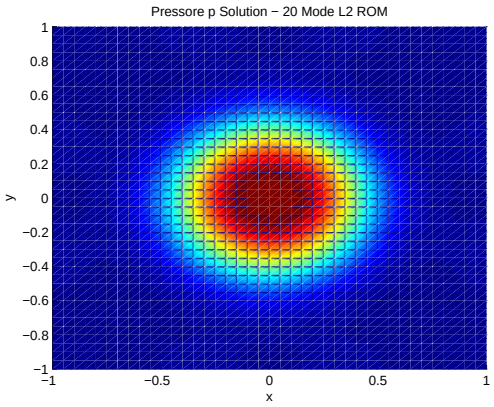
the L^2 ROM solution at these later times, however. It is apparent from Figure 6(c) that the L^2 ROM solution has blown up by $t = 7.95 \times 10^{-3}$ seconds, which confirms the instability of the 20 mode L^2 ROM suggested in Figure 2.



(a) CFD

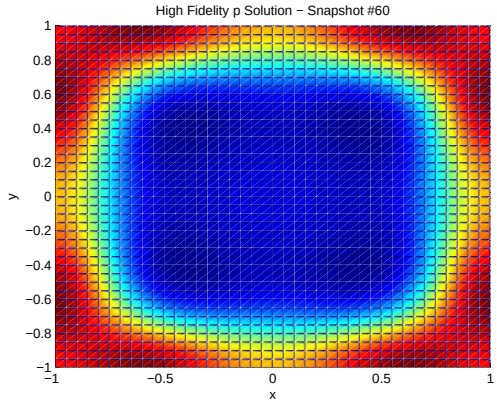


(b) 20 mode symmetry ROM

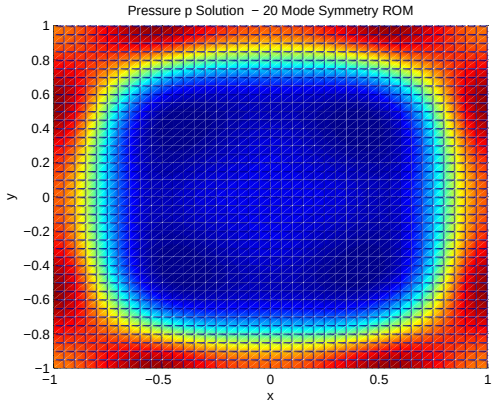


(c) 20 mode L^2 ROM

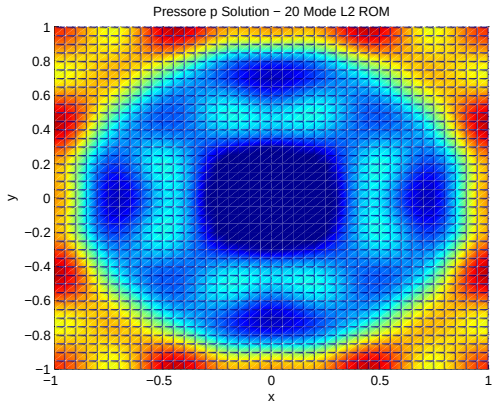
Figure 4. Pressure field at time $t = 4.5 \times 10^{-4}$ seconds



(a) CFD

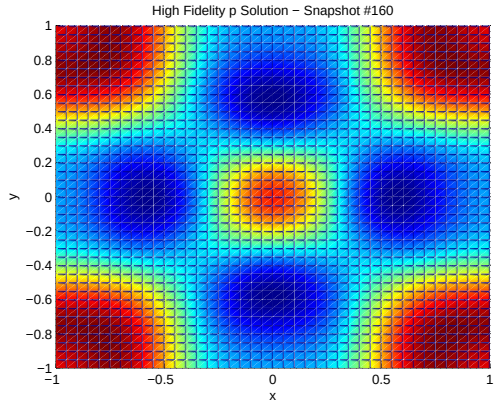


(b) 20 mode symmetry ROM

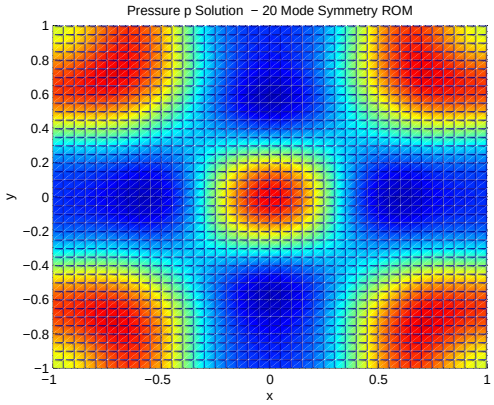


(c) 20 mode L^2 ROM

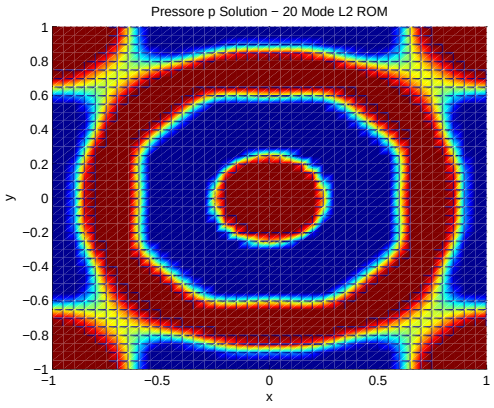
Figure 5. Pressure field at time $t = 2.95 \times 10^{-3}$ seconds



(a) CFD



(b) 20 mode symmetry ROM



(c) 20 mode L^2 ROM

Figure 6. Pressure field at time $t = 7.95 \times 10^{-3}$ seconds

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. This formulation raises the following question: is it possible to extend the approach developed in Section 4 to ROMs constructed using the discrete projection approach outlined in Section 2.1?

It turns out that the answer to this question is yes. In Section 5.1, a discrete counterpart of the continuous symmetry inner product (Section 4.2), termed the "Lyapunov inner product" is derived. It is shown that if a ROM for the LTI system (2) is constructed in the Lyapunov inner product, the ROM will be time-stable and therefore stable in the sense of Lyapunov as well as energy-stable. Unlike the equation-specific symmetry inner product described in Section (4.2), the Lyapunov inner product can be computed numerically in a black-box fashion by solving a Lyapunov equation. Section 5.2 demonstrates that the balanced truncation approach to model reduction may be viewed as a Galerkin projection in a particular Lyapunov inner product. This inner product is derived from the balanced truncation algorithm (Algorithm 2) and energy-stability of balanced truncation is proven using the energy method. The performance of the two approaches in addition to the classical Galerkin/POD method is evaluated on two benchmark test cases in Section 5.3.

An *a posteriori* literature review reveals that the Lyapunov inner product has been studied by several authors. Among the first presentations of this inner product (to the authors' knowledge) appeared in [39] by Rowley *et al.* The inner product was mentioned in some recent works by Amsallem *et al.* [3] and Serre *et al.* [41]. To the authors' knowledge, a numerical study of the properties and performance of POD ROMs constructed in the Lyapunov inner product is lacking in the literature at the present time.

5.1 Stability-Preserving Lyapunov Inner Product

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

$$\dot{\mathbf{x}}_N(t) = \mathbf{A}\mathbf{x}_N(t) \tag{103}$$

(Theorem 3.4). In particular,

$$V(\mathbf{x}_N) = \mathbf{x}_N^T \mathbf{P} \mathbf{x}_N, \tag{104}$$

is a Lyapunov function for (103), where $\mathbf{P}$ is the solution of the following Lyapunov equation:

$$\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} = -\mathbf{Q}. \tag{105}$$

Here, $\mathbf{Q}$ is some positive-definite matrix [6]. A positive definite solution $\mathbf{P}$ to (105) 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 (105) can be obtained, for instance, using the `lyap` function in the MATLAB control toolbox [45]:

```
P = lyap(A', Q, [], speye(N, N)).
```

Assume the system (103) is stable and a positive-definite symmetric $\mathbf{P}$ has been computed from (105). Since $\mathbf{P}$ is symmetric positive-definite, the following

$$\left(\mathbf{x}_N^{(1)}, \mathbf{x}_N^{(2)}\right)_{\mathbf{P}} \equiv \mathbf{x}_N^{(1)T} \mathbf{P} \mathbf{x}_N^{(2)}, \tag{106}$$

defines an inner product. By the discussion of Section 4.3, (106), the energy inner product induced by the Lyapunov function for (103), is the energy inner product for this system. Let $\mathbf{\Phi}_M$ be a reduced basis of size M, so that

$$\mathbf{x}_N(t) \approx \mathbf{\Phi}_M \mathbf{x}_M(t), \tag{107}$$

where $\mathbf{x}_M(t)$ denotes the ROM solution.

Theorem 5.1.1: Assume the linear full order system (103) is stable. Suppose a ROM for (103) is constructed via a Galerkin projection in the $(\cdot,\cdot)_{\mathbf{P}}$ inner product (106), to yield the following reduced linear system:

$$\dot{\mathbf{x}}_M = \mathbf{\Phi}_M^T \mathbf{P} \mathbf{A} \mathbf{\Phi}_M \mathbf{x}_M, \tag{108}$$

where it has been assumed that the basis $\mathbf{\Phi}_M$ has been constructed to be orthonormal in the $(\cdot,\cdot)_{\mathbf{P}}$ inner product, i.e., $\mathbf{\Phi}_M^T \mathbf{P} \mathbf{\Phi}_M = \mathbf{I}_M$ where $\mathbf{I}_M$ denotes the $M \times M$ identity matrix. Then, the ROM (108) is energy-stable, time-stable and stable in the sense of Lyapunov.

Proof. It is shown that the energy $E_M \equiv \frac{1}{2}||\mathbf{x}_M||_2^2$ of the ROM system (108) is non-increasing:

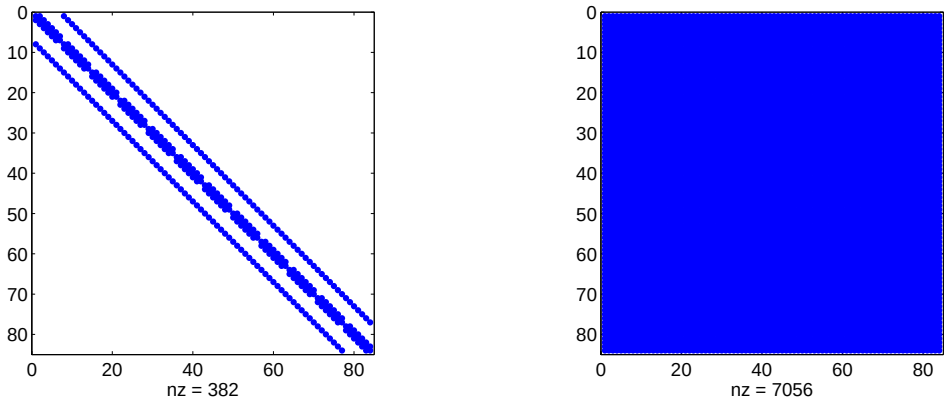
$$\begin{aligned}
\tfrac{dE_M}{dt} &= \tfrac{1}{2}\tfrac{d}{dt}\left(\mathbf{x}_M, \mathbf{x}_M\right)_2 \\
&= \mathbf{x}_M^T \dot{\mathbf{x}}_M \\
&= \mathbf{x}_M^T \mathbf{\Phi}_M^T \mathbf{P} \mathbf{A} \mathbf{\Phi}_M \mathbf{x}_M \\
&= \mathbf{x}_M^T \mathbf{\Phi}_M^T \left(\tfrac{1}{2}\mathbf{P}\mathbf{A} + \tfrac{1}{2}\mathbf{P}^T\mathbf{A}\right) \mathbf{\Phi}_M \mathbf{x}_M \\
&= \mathbf{x}_M^T \mathbf{\Phi}_M^T \left(\tfrac{1}{2}\mathbf{P}\mathbf{A} + \tfrac{1}{2}\mathbf{A}^T\mathbf{P}\right) \mathbf{\Phi}_M \mathbf{x}_M \\
&= -\tfrac{1}{2}\mathbf{x}_M^T \mathbf{\Phi}_M^T \mathbf{Q} \mathbf{\Phi}_M \mathbf{x}_M \\
&< 0,
\end{aligned} \tag{109}$$

since $\mathbf{Q} > \mathbf{0}$. It follows that (108) is time-stable, stable in the sense of Lyapunov and energy-stable (Section 3).

□

The Lyapunov inner product (106) is the discrete counterpart of the continuous symmetry inner product (67). 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 (106) is related in some way to the matrix $\mathbf{W}$ (98) that

arises when performing a continuous projection in the symmetry inner product. In general, the answer is no. In particular, $\mathbf{W}$ is by construction a sparse matrix (Figure 1), whereas $\mathbf{P}$ may be dense even if $\mathbf{A}$ is sparse. This is clear from Figures 7 (a) and (b), which show (respectively) the sparsity pattern of a sample $\mathbf{A}$ matrix[4], and its corresponding $\mathbf{P}$ matrix.



(a) A **(b) P**

Figure 7. Sparsity structure of representative $\mathbf{P}$ matrix for a given sparse $\mathbf{A}$ matrix

As observed earlier for the symmetry inner product, it is clear from (108) that the Galerkin projection of the system (103) in the Lyapunov inner product (106) can be viewed as a Petrov-Galerkin projection of this system in the regular L^2 inner product, with the test reduced basis given by $\mathbf{\Psi}_M = \mathbf{P}\mathbf{\Phi}_M$, where $\mathbf{\Phi}_M$ is the trial reduced basis.

5.2 Lyapunov Inner Product Associated with Balanced Truncation

In comparing the steps of the balanced truncation algorithm (Algorithm 2) 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 (Algorithm 2) may be viewed as a projection algorithm in a special Lyapunov inner product.

Substituting (30) into (34), the following expressions for the left and right bases are obtained:

$$\mathbf{\Psi}_M^T = \mathbf{T}_{bal}(1:M,:) = \mathbf{\Sigma}^{1/2}(1:M,:)\mathbf{K}^T\mathbf{U}^{-1}, \tag{110}$$

$$\mathbf{\Phi}_M = \mathbf{S}_{bal}(:,1:M) = \mathbf{U}\mathbf{K}\mathbf{\Sigma}^{-1/2}(:,1:M). \tag{111}$$

[4]The $\mathbf{A}$ matrix whose sparsity pattern is shown in Figure 7 is the "pde example" in the SLICOT model reduction benchmark repository [15]. For a physical description of this problem, the reader is referred to [14].

Remark that (110) and (111) satisfy the following identity:

$$\begin{aligned}\boldsymbol{\Sigma}^{-1}(1:M,1:M)\boldsymbol{\Psi}_M^T\mathbf{P} &= \boldsymbol{\Sigma}^{-1}(1:M,1:M)\boldsymbol{\Sigma}^{1/2}(1:M,:)\mathbf{K}^T\mathbf{U}^{-1}\mathbf{U}\mathbf{U}^T \\ &= \boldsymbol{\Sigma}^{-1/2}(1:M,:)\mathbf{K}^T\mathbf{U}^T \\ &= [\mathbf{U}\mathbf{K}\boldsymbol{\Sigma}^{-1/2}(1:M,:)]^T \\ &= \boldsymbol{\Phi}_M^T,\end{aligned} \tag{112}$$

where $\mathbf{P}$ is the reachability Gramian (28). It follows that the ROM system matrices in (33) are:

$$\mathbf{A}_M = \boldsymbol{\Psi}_M^T\mathbf{A}\boldsymbol{\Phi} = \boldsymbol{\Psi}_M^T\mathbf{A}\mathbf{P}^T\boldsymbol{\Psi}_M\boldsymbol{\Sigma}^{-1}(1:M,1:M), \tag{113}$$

$$\mathbf{B}_M = \boldsymbol{\Psi}_M^T\mathbf{B}, \tag{114}$$

$$\mathbf{C}_M = \mathbf{C}\boldsymbol{\Phi} = \mathbf{C}\mathbf{P}^T\boldsymbol{\Psi}_M\boldsymbol{\Sigma}^{-1}(1:M,1:M). \tag{115}$$

The system (33) becomes:

$$\begin{aligned}\dot{\mathbf{x}}_M(t) &= \boldsymbol{\Psi}_M^T\mathbf{A}\mathbf{P}^T\boldsymbol{\Psi}_M\boldsymbol{\Sigma}^{-1}(1:M,1:M)\mathbf{x}_M(t) + \boldsymbol{\Psi}_M^T\mathbf{B}\mathbf{u}_P(t), \\ \mathbf{y}_{QM}(t) &= \mathbf{C}\mathbf{P}^T\boldsymbol{\Psi}_M\boldsymbol{\Sigma}^{-1}(1:M,1:M)\mathbf{x}_M(t).\end{aligned} \tag{116}$$

One further transformation is required to exhibit the inner product associated with the balanced truncation algorithm. Define:

$$\mathbf{z}_M(t) \equiv \boldsymbol{\Sigma}^{-1/2}(1:M,1:M)\mathbf{x}_M(t). \tag{117}$$

With this transformation, (116) becomes:

$$\begin{aligned}\dot{\mathbf{z}}_M(t) &= \hat{\boldsymbol{\Psi}}_M^T\mathbf{A}\mathbf{P}^T\hat{\boldsymbol{\Psi}}_M\mathbf{z}_M(t) + \hat{\boldsymbol{\Psi}}_M^T\mathbf{B}\mathbf{u}_P(t), \\ \mathbf{y}_{QM}(t) &= \mathbf{C}\mathbf{P}^T\hat{\boldsymbol{\Psi}}_M\mathbf{z}_M(t),\end{aligned} \tag{118}$$

where

$$\hat{\boldsymbol{\Psi}}_M \equiv \boldsymbol{\Psi}_M\boldsymbol{\Sigma}^{-1/2}. \tag{119}$$

Using the symmetry property of the reachability Gramian ($\mathbf{P} = \mathbf{P}^T$), (118) is equivalent to

$$\begin{aligned}\dot{\mathbf{z}}_M(t) &= \hat{\boldsymbol{\Psi}}_M^T\mathbf{A}\mathbf{P}\hat{\boldsymbol{\Psi}}_M\mathbf{z}_M(t) + \hat{\boldsymbol{\Psi}}_M^T\mathbf{B}\mathbf{u}_P(t) \\ \mathbf{y}_{QM}(t) &= \mathbf{C}\mathbf{P}\hat{\boldsymbol{\Psi}}_M\mathbf{z}_M(t).\end{aligned} \tag{120}$$

It is clear that (120) 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, by Lemma 2.3.4, if $(\mathbf{A}, \mathbf{B})$ is reachable (controllable).

A property of balanced truncation is that, when applied to stable systems, balanced truncation preserves stability [18] (Section 2.2). This result can be proven using the energy method.

Theorem 5.2.1: Assume the linear full order system (2) is stable. Suppose a ROM for (2) is constructed via balanced truncation to yield the reduced system (120). Then, the ROM (120) is energy-stable, time-stable and stable in the sense of Lyapunov.

Proof. It is shown that the energy $E_M \equiv \frac{1}{2}||\mathbf{z}_M||_2^2$ of the homogeneous version of the ROM system (120) is non-increasing:

$$\begin{aligned}
\tfrac{dE_M}{dt} &= \tfrac{1}{2}\tfrac{d}{dt}\left(\mathbf{z}_M, \mathbf{z}_M\right)_2^2 \\
&= \mathbf{z}_M^T \dot{\mathbf{z}}_M \\
&= \mathbf{z}_M^T \hat{\mathbf{\Psi}}_M^T \mathbf{A}\mathbf{P}\hat{\mathbf{\Psi}}_M \mathbf{z}_M \\
&= \mathbf{z}_M^T \hat{\mathbf{\Psi}}_M^T \left(\tfrac{1}{2}\mathbf{A}\mathbf{P} + \tfrac{1}{2}\mathbf{A}\mathbf{P}^T\right) \hat{\mathbf{\Psi}}_M \mathbf{z}_M \\
&= \mathbf{z}_M^T \hat{\mathbf{\Psi}}_M^T \left(\tfrac{1}{2}\mathbf{A}\mathbf{P} + \tfrac{1}{2}\mathbf{P}\mathbf{A}^T\right) \hat{\mathbf{\Psi}}_M \mathbf{z}_M \\
&= -\tfrac{1}{2}\mathbf{z}_M^T \mathbf{B}\mathbf{B}^T \mathbf{z}_M \\
&\leq 0,
\end{aligned} \tag{121}$$

since the eigenvalue of $\mathbf{B}\mathbf{B}^T$ are the singular values of $\mathbf{B}$ which are by definition ≥ 0. It follows that a ROM constructed using balanced truncation is energy-stable, time-stable and stable in the sense of Lyapunov (Section 3).

□

5.3 Numerical Experiments

International Space Station (ISS)

The first numerical example considered here involves a structural model of component 1r (Russian service module) of the International Space Station (ISS) [5]. 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 [15]. It is verified that the 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 $K = 2000$ 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 (106). The matrix $\mathbf{P}$ defining the inner product (106) 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.

First, the eigenvalues of each ROM matrix $\mathbf{A}_M$ for each M are computed to determine stability using Corollary 3.3.2. The maximum real part of the eigenvalues of these ROM system matrices is plotted in Figure 8 as a function of M. The reader can observe that the

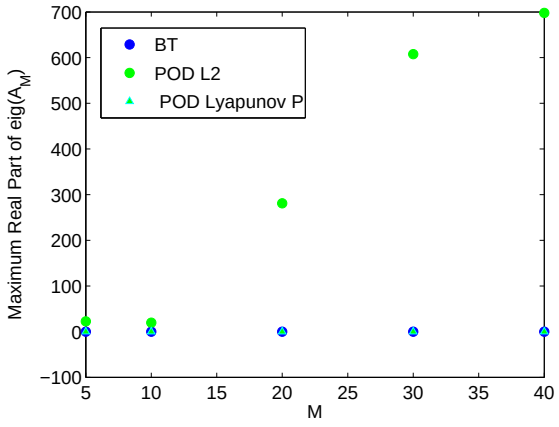


Figure 8. Maximum real part of eigenvalues of ROM system matrix $\mathbf{A}_M$ for ISS problem

Lyapunov inner product POD ROMs and balanced truncation ROMs are stable for all M considered – all 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, i.e.,

$$\mathcal{E}^o_{rel} = \frac{\sum_{i=1}^{K_{max}} ||\mathbf{y}_{QN}(t_i) - \mathbf{y}_{QM}(t_i)||_2}{\sum_{i=1}^{K_{max}} ||\mathbf{y}_{QN}(t_i)||_2}, \tag{122}$$

is computed. Here K_{max} is the integer such that $T_{max} = K_{max} dt_{snap}$. The relative errors (122) in the output for ROMs of different size run up to different values of T_{max} are summarized in Table 1. In the case a ROM went unstable and (122) 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 constructed, as exactly $K = 2000$ snapshots (taken up to $t = 0.1$ seconds) were used to generate these bases. The following conclusions can be drawn from the results given below:

- Although the L^2 POD ROM is unstable for all values of M considered (Figure 8), this ROM still produces a reasonable ROM for $M = 5$ and $M = 10$ (Figure 9(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).

For 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 ROMs are run for a much longer time horizon than the run used to generate the POD bases employed in building the ROMs. The

following observations are noteworthy:

- For $K_{max} = 5000$, The L^2 POD ROM manifests 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 9(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 9(c). Hence, the instability identified in the earlier eigenvalue analysis (Figure 8) manifests itself if the L^2 POD ROM 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 convergence with M-refinement.

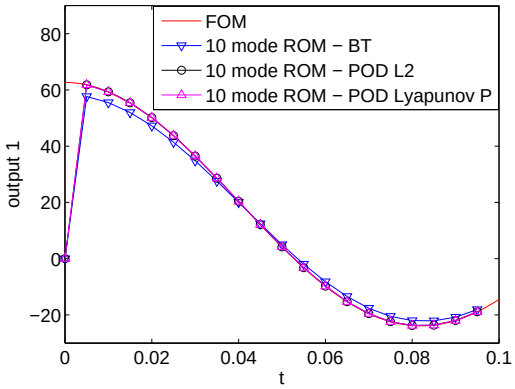
Table 1. Relative Errors (122) $\mathcal{E}^o_{rel}$ in ROM Outputs for ISS Problem

		M				
K_{max}	Method	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}

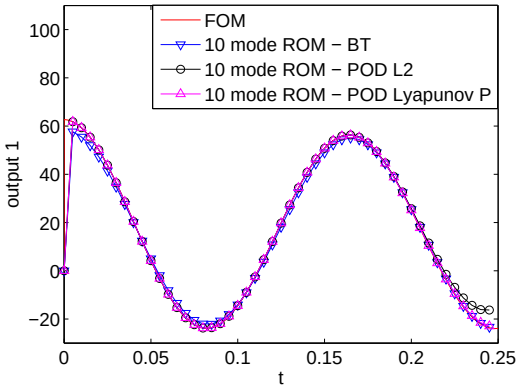
As a final test, the Lyapunov POD ROM and balanced truncation ROMs with $M = 40$ modes are run for a very long time, until $T_{max} = 5$ seconds ($K_{max} = 100,000$). As before, the ROM bases were constructed from only for the first $K = 2000$ snapshots (until time $t = 0.1$ seconds) of the solution. The output computed by the ROMs is plotted in Figure 9. The L^2 POD ROM is not shown as it goes unstable at $t \approx 0.3$ seconds. The balanced truncation ROM agrees very well with the full order solution. Performance of the Lyapunov POD ROM is reasonable given that the basis employed with this ROM knows nothing about the solution for $t > 0.1$ seconds.

Electrostatically Actuated Beam

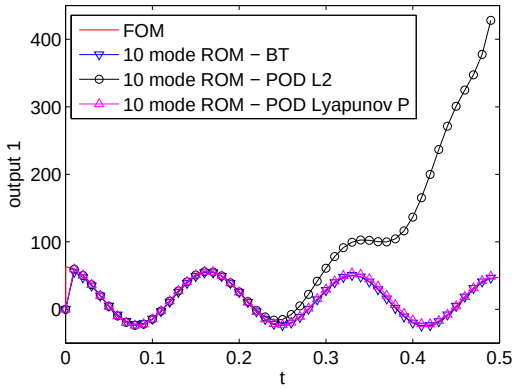
The second numerical example is that of an electrostatically actuated beam. One application for this model are microelectromechanical systems (MEMS) devices such as electromechanical radio frequency (RF) filters [33]. Given a simple enough shape, these devices can be modeled as 1D beams embedded in two or three dimensional space. The beam considered



(a) $K_{max} = 2000$



(b) $K_{max} = 5000$



(c) $K_{max} = 10,000$

Figure 9. $\mathbf{y}_{QM}(t)$ for $M = 10$ ROMs (FOM = full order model) for ISS Problem

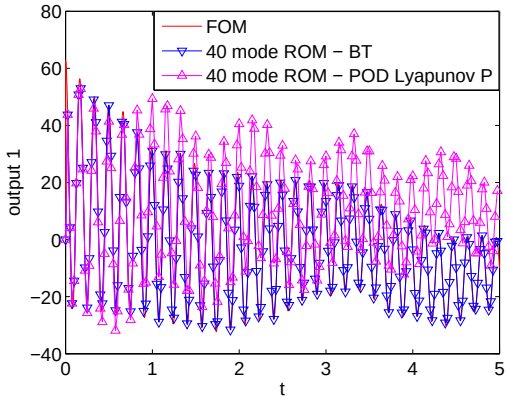


Figure 10. $\mathbf{y}_{QM}(t)$ for $M = 40$ Lyapunov POD and Balanced Truncation ROMs, $K_{max} = 100,000$ (FOM = full order model) for ISS Problem

here is supported on both sides, and has two degrees of freedom: the deflection perpendicular to the beam (the flexural displacement), and the rotation in the deformation plane (the flexural rotation). The equations of motion are determined from a Lagrangian formulation. 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, 33]. The result of this discretization is a second order linear semi-discrete system of the form:

$$\begin{aligned} \mathbf{M}\ddot{\mathbf{x}}_N + \mathbf{E}\dot{\mathbf{x}}_N + \mathbf{K}\mathbf{x}_N &= \mathbf{B}\mathbf{u}_P \\ \mathbf{y}_{QN} &= \mathbf{C}\mathbf{x}_N, \end{aligned} \tag{123}$$

where $\ddot{\mathbf{x}}_N \equiv \frac{\partial^2 \mathbf{x}_N}{\partial t^2}$. The input matrix $\mathbf{B}$ corresponds to a loading of the middle node of the domain, and $\mathbf{y}_{QN}$ is the flexural displacement at the middle node of the domain. The damping matrix $\mathbf{E}$ is taken to be a linear combination of the mass matrix $\mathbf{M}$ and the stiffness matrix $\mathbf{K}$:

$$\mathbf{E} = c_M\mathbf{M} + c_K\mathbf{K}, \tag{124}$$

with $c_M = 10^2$ and $c_K = 10^{-2}$. Letting $\tilde{\mathbf{x}}_N \equiv \dot{\mathbf{x}}_N$, the second order system (123) can be written as the following first order system:

$$\begin{aligned} \begin{pmatrix} \mathbf{E} & \mathbf{M} \\ \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \dot{\mathbf{x}}_N \\ \dot{\tilde{\mathbf{x}}}_N \end{pmatrix} + \begin{pmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{x}_N \\ \tilde{\mathbf{x}}_N \end{pmatrix} &= \begin{pmatrix} \mathbf{B} \\ \mathbf{0} \end{pmatrix} \mathbf{u}_P \\ \mathbf{y}_{QN} &= \begin{pmatrix} \mathbf{C} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x}_N \\ \tilde{\mathbf{x}}_N \end{pmatrix}, \end{aligned} \tag{125}$$

or

$$\begin{aligned} \dot{\mathbf{z}}_{2N} &= \mathbf{A}\mathbf{z}_{2N} + \tilde{\mathbf{B}}\mathbf{u}_P \\ \mathbf{y}_{QN} &= \tilde{\mathbf{C}}\mathbf{z}_{2N}, \end{aligned} \tag{126}$$

where $\mathbf{z}_{2N}^T \equiv \begin{pmatrix} \mathbf{x}_N \\ \tilde{\mathbf{x}}_N \end{pmatrix}$ and

$$\mathbf{A} \equiv \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{E} \end{pmatrix}, \quad \tilde{\mathbf{B}} \equiv \begin{pmatrix} \mathbf{M}^{-1}\mathbf{B} \\ \mathbf{0} \end{pmatrix}, \quad \tilde{\mathbf{C}} \equiv \begin{pmatrix} \mathbf{C} & \mathbf{0} \end{pmatrix}. \tag{127}$$

The matrices $\mathbf{M}$ and $\mathbf{K}$ in (123) are downloaded from the Oberwolfach model reduction benchmark collection [1]. 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, $N = 5000$, so (126) has 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 (126) is solved using a backward Euler time integration scheme with an initial condition of $\mathbf{z}_{2N}(0) = \mathbf{0}$ and an input corresponding to a periodic on/off switching, i.e.,

$$\mathbf{u}_P = \begin{cases} 1, & 0.005 < t < 0.01, 0.015 < t < 0.02, 0.03 < t < 0.035 \\ 0, & \text{otherwise} \end{cases} \tag{128}$$

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 balanced truncation, POD in the L^2 inner product, and POD in the Lyapunov inner product. In solving the Lyapunov equation (105) for the Lyapunov inner product weighting matrix $\mathbf{P}$, the matrix $\mathbf{Q}$ is taken to be the $2N \times 2N$ identity matrix.

As for the ISS example, the first step is to study the stability of each ROM. Figure 11 shows the maximum real part of each ROM system matrix $\mathbf{A}_M$ 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.

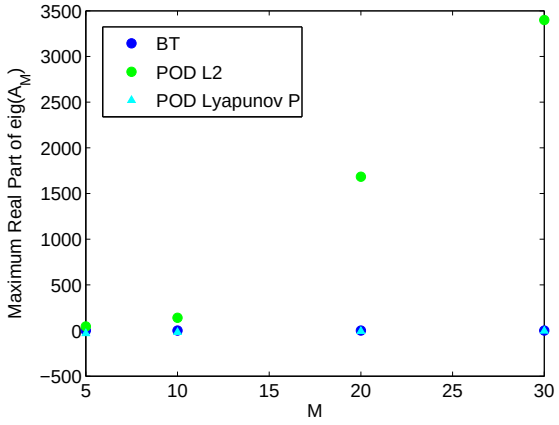


Figure 11. Maximum real part of eigenvalues of ROM system matrix $\mathbf{A}_M$ for beam problem

Next, the accuracy of each ROM is examined. Table 5.3 summarizes the errors (122) in the ROM solutions relative to the full order model solution for three runs of different lengths. 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 12(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 time 2.5 seconds. 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 (128), the following inputs are applied

in both the full order model and the ROM:

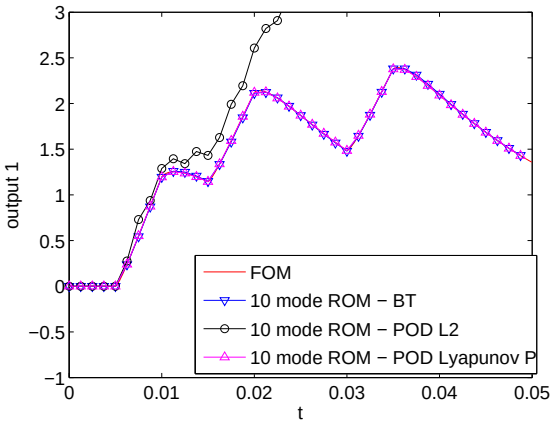
$$\mathbf{u}_P = \begin{cases} & 0.055 < t < 0.06, 0.065 < t < 0.07, 0.08 < t < 0.085, \\ & 0.105 < t < 0.11, 0.115 < t < 0.12, 0.13 < t < 0.135, \\ 1, & 0.205 < t < 0.21, 0.215 < t < 0.22, 0.23 < t < 0.235, \\ 0, & \text{otherwise.} \end{cases} \tag{129}$$

The reader may observe by examining Table 5.3 and Figure 12 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 12(b)-(c)) and appear to be converging to the full order model solution with M-refinement (Table 5.3). The POD L^2 ROM result is not shown in Figures 12(b)-(c), as the solution produced by this ROM blows up around time $t = 0.02$ seconds.

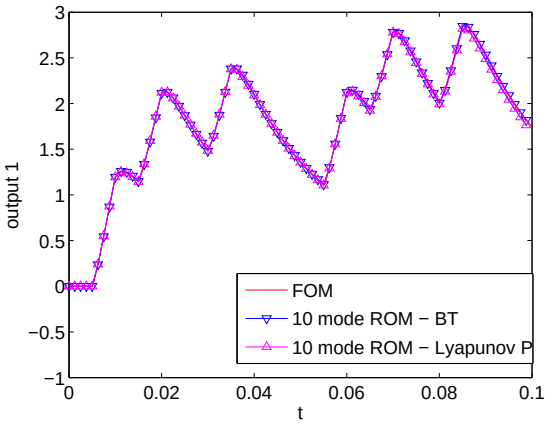
Table 2. Relative Errors (122) $\mathcal{E}^o_{rel}$ in ROM Outputs for Beam Problem

		M			
K_{max}	Method	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}

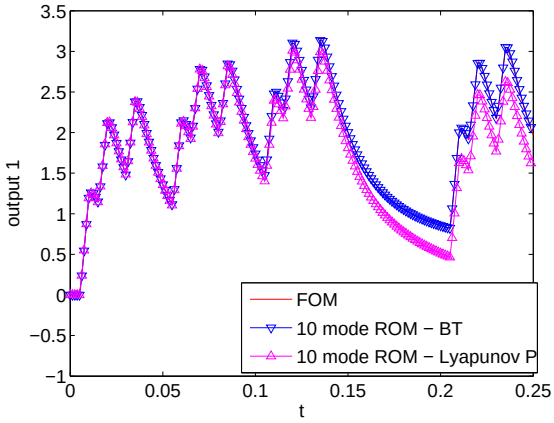
Lastly, the level of computational resources required for computing the Lyapunov inner product and the the level of computational resources required for performing model reduction via balanced truncation are compared. Table 3 gives the CPU times for Steps 1-5 of the balanced truncation algorithm (Algorithm 2) and the CPU times for solving the Lyapunov equation (105) as a function of $2N$, the problem size. All computations are performed in serial using MATLAB's linear algebra capabilities and MATLAB's control toolbox, on a Linux workstation with 6 Intel Xeon 2.93 GHz CPUs. Although the Lyapunov inner product computation is costly, as it requires the solution of a Lyapunov equation, it requires 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.



(a) $K_{max} = 1000$



(b) $K_{max} = 2000$



(c) $K_{max} = 5000$

Figure 12. $\mathbf{y}_{QM}(t)$ for $M = 10$ ROMs (FOM = full order model) for Beam Problem

Table 3. CPU Times (in seconds) for Balanced Truncation vs. Lyapunov Inner Product Computations

	$2N$			
Method	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

6 Summary and Conclusions

In this report, several approaches for building projection-based ROMs with an *a priori* stability guarantee are presented and unified using the notion of energy-stability. 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 systems, the said transformation is induced by a special weighted L^2 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. A discrete counterpart of the symmetry inner product, referred to as the "Lyapunov inner product", is derived, and it is demonstrated that a ROM constructed for an LTI system via discrete projection in this inner product has an *a priori* stability guarantee, again regardless of the choice of reduced basis. Connections between the Lyapunov inner product and the inner product induced by the balanced truncation approach to model reduction are made. 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 properites of the symmetry inner product and Lyapunov inner product are summarized in the table below.

Symmetry Inner Product (67)	**Lyapunov Inner Product** (106)
Continuous	Discrete
For linear PDE system of the form $\dot{\mathbf{q}} + \mathbf{A}_i\mathbf{q}_{,i} + \mathbf{K}_{ij}\mathbf{q}_{,ji} = \mathbf{F}$	For linear ODE system of the form $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$
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

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 (52). Projection in this inner product requires access to the governing PDEs, which gives rise to a projection algorithm that is embedded. In contrast, the Lyapunov inner product is discrete, 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 $t \to \infty$.

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 [19]. As a result, creating ROMs using the Laypunov 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 factorizations of these Gramians.

For the reasons described above, for large-scale unsteady problems, the symmetry inner product combined with the continuous projection approach is recommended by the authors, despite its more intrusive implementation.

References

[1] Oberwolfach benchmark collection, 2005.

[2] S. Abarbanel and D. Gottlieb. Optimal time splitting for two- and three-dimensional navier-stokes equations with mixed derivatives. *J. Comput. Phys.*, 41:1–33, 1981.

[3] D. Amsallem and C. Farhat. On the stability of projection-based 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.

[4] D. Amsallem and C. Farhat. Stabilization of projection-based reduced order models. *Int. J. Numer. Meth. Engng.*, 91(4):358–377, 2012.

[5] A.C. Antoulas, D.C. Sorensen, and S. Gugercin. A survey of model reduction methods for large-scale systems. *Contemporary Mathematics*, (280), 2001.

[6] K.J. Astrom and R.M. Murray. *Feedback systems: an introduction for scientists and engineers*. Princeton University Press, 2008.

[7] M.J. Balajewicz, E.H. Dowell, and B.R. Noack. A novel model order reduction approach for navier-stokes equations at high reynolds number. *J. Fluid Mech.* (under review), 2012.

[8] A. Barbagallo, D. Sipp, and P.J. Schmid. Closed-loop control of an open cavity flow using reduced-order models. *J. Fluid Mech.*, 641:1–50, 2009.

[9] M.F. Barone, I. Kalashnikova, D.J. Segalman, and H. Thornquist. Stable galerkin reducerd order models for linearized compressible flow. *J. Comput. Phys.*, 288:1932–1946, 2009.

[10] P. Benner, M. Castillo, E.S. Quintana-Orti, and G. Quintana-Orgi. Parallel model reduction of large-scale unstable systems. *Advanced in Parallel Computing*, 13:251–258, 2004.

[11] R.H. Bishop. *The Mechatronics Handbook*. CRC Press LLC, 2002.

[12] L. Daniel B.N. Bond. 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.

[13] T. Bui-Thanh, K. Willcox, O. Ghattas, and B. van Bloemen Waanders. Goal-oriented, model constrained optimization for reduction of large-scale systems. *J. Comp. Phys.*, 224:880–896, 2007.

[14] Y. Chahlaoui and P. Van Dooren. A collection of benchmark examples for model reduction of linear time invariant dynamical systems. SLICOT Working Note 2002-2,

February 2002.

[15] Y. Chahlaoui and P. Van Dooren. Benchmark examples for model reduction of linear time invariant systems, 2013.

[16] G. Chen. *Stability of Nonlinear Systems*. Wiley, 2004.

[17] D. Funaro and D. Gottlieb. Convergence results for pseudospectral approximations of hyperbolic systems by a penalty-type boundary treatment. *Mathematics of Computation*, 196(57):585–596, 1991.

[18] S. Gugercin and A.C. Antoulas. A survey of model reduction by balanced truncation and some new results. *Int. J. Control*, 77(8):748–766, 2004.

[19] S. Gugercin and J.-R. Li. Smith-type mehtods for balanced truncation of large sparse systems. *Lecture Notes in Computational Science and Engineering*, 45:49–82, 2005.

[20] M.D. Gunzburger. On the stability of galerkin methods for initial-boundary value problems for hyperbolic systems. *Math. Comp.*, 31(139):661–675, 1977.

[21] B. Gustaffson. *High order difference methods for time dependent PDE*. Springer-Verlag, 2008.

[22] B. Gustaffson, H.-O. Kreiss, and J. Oliger. *Time Dependent Problems and Difference Methods*. Wiley-Interscience, 1995.

[23] B. Gustafsson and A. Sundstrom. Incompletely parabolic problems in fluid dynamics. *SIAM J. Appl. Math.*, 35(2):343–357, 1978.

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

[25] A.T. Patera K. Veroy. Certified real-time solution of the parametrized steady incompressible navier-stokes equations: rigorous reduced-bases *a posteriori* error bounds. *J. Num. Meth. Fluids*, 47:773–788, 2005.

[26] I. Kalashnikova and S. Arunajatesan. A stable galerkin reduced order model for compressible flow. *WCCM-2012-19407, 10th World Congress on Computational Mechanics (WCCM)*, Sao Paulo, Brazil, 2012.

[27] I. Kalashnikova and 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.*, 83:1345–1375, 2010.

[28] H. Kimura. *Chain-Scattering Approach to H-infinity Control*. Springer, 1997.

[29] H.O. Kreiss and J. Lorenz. *Initial-Boundary Value Problems and the Navier-Stokes Equations*. Academic Press, Inc., 1989.

[30] K. Kunisch and S. Volkwein. Galerkin proper orthogonal decomposition for a general equation in fluid dynamics. *SIAM J. Num. Anal.*, 40(2):492–515, 2002.

[31] F. Kwasniok. Empirical low-order models of barotropic flow. *J. Atmos. Sci*, 61(2):235–245, 2004.

[32] W.J. Layton. Stable galerkin methods for hyperbolic systems. *SIAM J. Numer. Anal.*, 20(3):221–233, 1983.

[33] J. Lienemann, E.B. Rudnyi, and J.G. Korvink. Mst mems model order reduction: Requirements and benchmarks. *Linear Algebra Appl.*, 415(2-3):469–498, 2006.

[34] J.L. Lumley. *Stochastic tools in turbulence*. Academic Press, New York, 1971.

[35] L.R. Petzold M. Rathinam. A new look at proper orthogonal decomposition. *SIAM J. Num. Anal.*, 41(5):1893–1925, 2003.

[36] B. Moore. Principal component analysis in linear systems: Controllability, observability and model reduction. *IEEE Transactions on Automatic Control*, 26(1), 1981.

[37] J. Lumley E. Stone N. Aubry, P. Holmes. The dynamics of coherent structures in the wall region of a turbulent boundary layer. *J. Fluid Mech.*, 192:115–173, 1988.

[38] C.W. Rowley. Model reduction for fluids using balanced proper orthogonal decomposition. *Int. J. Bif. Chaos*, 15(3):997–1013, 2005.

[39] C.W. Rowley, T. Colonius, and R.M. Murray. Model reduction for compressible flows using pod and galerkin projection. *Physica D*, 189:115–129, 2004.

[40] C.W. Rowley, I. Mezic, S. Baheri, P. Schlatter, and D.S. Henningson. Reduced-order models for flow control: balanced models and koopman modes. Seventh IUTAM Symposium on Laminar-Turbulent Transition, 2009.

[41] G. Serre, P. Lafon, X. Gloerfelt, and C. Bailly. Reliable reduced-order models for time-dependent linearized euler equations. *J. Comput. Phys.*, 231(15):5176–5194, 2012.

[42] L. Sirovich. Turbulence and the dynamics of coherent structures, part iii: dynamics and scaling. *Q. Appl. Math.*, 45(3):583–590, 1987.

[43] T.R. Smith. Low-dimensional models of plane couette flow using the proper orthogonal decomposition. Ph.D. thesis, Princeton University, 2003.

[44] E. Tadmor. Spectral methods for hyperbolic problems. Lecture Notes Delivered at Ecole des Ondes, *"Méthods numériques d'order élevé pour les ondes en régime transitoire"*, 1994.

[45] Inc. The MathWorks. Control systems toolbox user's guide.

[46] Jr. W. Weaver, S.P. Timoshenko, and D.H. Young. *Vibration problems in engineering.*

Wiley, 5th Ed., 1990.

[47] Z. Wang, I. Akhtar, J. Borggaard, and Traian Iliescu. Proper orthogonal decomposition closure models for turbulent flows: A numerical comparison. *Comput. Methods Appl. Mech. Engrg.*, pages 237–240, 2012.

[48] K. Willcox and J. Peraire. Balanced model reduction via the proper orthogonal decomposition. *AIAA Journal*, 40(11):2323–2330, 2002.

[49] K. Zhou. *Robust and optimal control*. Prentice Hall, 1996.

DISTRIBUTION:

1	MS 0825	Matthew Barone, 1515
1	MS 0825	Srinivasan Arunajatesan, 1515
1	MS 0825	Jeffrey Payne, 1515
1	MS 0828	Kenneth Hu, 1544
1	MS 1070	Matthew Brake, 1526
1	MS 1318	Bart van Bloemen Waanders, 1442
1	MS 1318	Andrew Salinger, 1442
1	MS 1320	S. Scott Collis, 1440
1	MS 1320	Michael Parks, 1442
1	MS 1320	Richard Lehoucq, 1444
1	MS 8259	Daniel Segalman, 9042
1	MS 9159	Kevin Carlberg, 8954
1	MS 9159	Julien Cortial, 8954
1	MS 9159	Martin Drohmann, 8954
1	MS 0899	Technical Library, 9536 (electronic copy)

Sandia National Laboratories