Canonical and noncanonical Hamiltonian operator inference

Anthony Gruber[a],[*], Irina Tezaur[b]

[a] *Center for Computing Research, Sandia National Laboratories, 1450 Innovation Parkway SE, Albuquerque, 87123, NM, USA*
[b] *Quantitative Modeling and Software Engineering, Sandia National Laboratories, 7011 East Avenue, Livermore, 94550, CA, USA*

Abstract

A method for the nonintrusive and structure-preserving model reduction of canonical and noncanonical Hamiltonian systems is presented. Based on the idea of operator inference, this technique is provably convergent and reduces to a straightforward linear solve given snapshot data and gray-box knowledge of the system Hamiltonian. Examples involving several hyperbolic partial differential equations show that the proposed method yields reduced models which, in addition to being accurate and stable with respect to the addition of basis modes, preserve conserved quantities well outside the range of their training data.
© 2023 Elsevier B.V. All rights reserved.

1. Introduction

In recent years, Digital Twins (DTs) have emerged as a new paradigm in the field of modeling and simulation. A DT is a computational model of a physical asset, such as a component, system or process, that evolves continuously in real or near-real time, so as to persistently represent the ever-changing structure and behavior of the underlying physical asset. In order for DTs to achieve their full potential as enablers of beyond-forward analyses such as optimal experimental design (OED), control and uncertainty quantification (UQ), it is essential that these computational models are: (1) capable of incorporating real-time data as it becomes available, (2) computationally efficient enough to provide predictions in real or near-real time, and (3) equipped with rigorous mathematical convergence, stability and accuracy guarantees.

Particularly helpful in establishing the above criteria is making appropriate use of well-studied mathematical structure inherent in the underlying partial differential equations (PDEs) when such structure is available. In the case that the system modeled obeys a variational principle, there are centuries of knowledge regarding dynamical properties (e.g., conservation laws) which can be leveraged to produce accurate and realistic simulations. Of particular interest at the present time are Hamiltonian systems, which form compact models of reversible, potentially

---

* Corresponding author.
  E-mail addresses: adgrube@sandia.gov (A. Gruber), ikalash@sandia.gov (I. Tezaur).

chaotic dynamics. Since many common systems relevant to digital twins have a Hamiltonian form (e.g., Hénon–Heiles, *n*-body motion, idealized MHD, solid dynamics), it is becoming increasingly necessary to have useful ways of building relatively cheap Hamiltonian surrogates which can be used to inform a high-quality digital representation.

Projection-based model order reduction (PMOR) is a promising strategy for reducing the computational cost of high-fidelity simulations, making projection-based reduced order models (ROMs) ideal candidates for constructing DTs. The key idea in PMOR is to learn a low-dimensional trial subspace by performing a data compression on a set of snapshots collected from a high-fidelity simulation or physical experiment, and to restrict the state variables to reside in this subspace. This effectively projects the high-fidelity dynamics into a much smaller function space, which must be carefully imbued with sufficient information for accurate reconstruction of the high-fidelity solutions. Traditionally, affine (or linear) approaches have been employed for constructing the low-dimensional trial subspace in which the ROM solution is sought, e.g., Proper Orthogonal Decomposition (POD) [1,2], Dynamic Mode Decomposition (DMD) [3,4], balanced POD (BPOD) [5,6], balanced truncation [7,8], and the reduced basis method (RBM) [9,10]. While all such methods have their own strengths and weaknesses, without loss of generality, this work restricts attention to the POD approach for calculating reduced bases due to its prevalence, flexibility, and simplicity. Beyond linear techniques, it is interesting to note that, in recent years, the idea of employing trial subspaces defined by nonlinear manifolds has started to be explored by a growing number of authors; see, e.g., [11–17] and references therein for nonlinear manifold approaches based on convolutional autoencoders, and [18,19] for quadratic manifold approaches. Nonlinear approximation approaches have the advantage of mitigating the so-called Kolmogorov *n*-width barrier [20], which reduces the efficacy and efficiency of linear manifold ROMs for convection-dominated problems.[1] However, they are often more difficult to train and can exhibit poor convergence behavior when compared with their linear counterparts [11,14].

Once a low-dimensional trial subspace has been constructed, the mathematical operators defining a ROM are obtained through a projection of the corresponding full order model (FOM) operators onto the reduced subspace. Performing this projection step is in general a very intrusive process, as it requires access to the FOM code used to generate the snapshot data. This intrusive nature of the projection step in PMOR limits the class of problems to which the approach can be applied, precluding the application of PMOR to FOMs that are given as a black-box. A promising approach for overcoming this limitation is data-driven Operator Inference (OpInf) (e.g., [21–23]), which aims to construct projection-based ROMs in a nonintrusive way. OpInf is motivated by the observation that projection preserves algebraic structure, that is, if the semi-discretized FOM has polynomial nonlinearities, a projection-based ROM for this system will also have polynomial nonlinearities of the same degree. Once the functional, algebraic structure of the FOM (and hence the ROM) is determined, OpInf works by replacing the intrusive projection step that is typically used to determine the ROM operators with a least-squares problem that infers these operators directly in a black-box fashion using available snapshot data (c.f. Section 2.3).

It is well known that projection-based ROMs constructed using either intrusive or non-intrusive techniques will generally not automatically inherit key mathematical properties of the PDEs from which they are derived. Since these properties are often well-understood to be responsible for the involved physics, this is a major defect which can harm the predictive performance of ROMs, limiting their utility in practical cases of interest. To remedy this difficulty, a variety of methodologies have been proposed which focus around preserving different mathematical structures often seen in application settings. Here, we summarize the literature on this subject for several common properties whose numerical preservation is critical to a wide range of applications: energy-/entropy- stability, conservation law preservation, and variational structure preservation (most notably involving Hamiltonian or Lagrangian structure, and including the focus of this paper).

It is worth noting that the majority of structure-preserving PMOR approaches in the literature focus on intrusive ROMs rather than non-intrusive OpInf models. The present work is a step towards filling this gap for the specific case of Hamiltonian systems. In order to distinguish our approach from other related work, we provide a succinct overview of existing OpInf methods below, after our overview of commonly-preserved structures/properties.

Energy- and entropy-stability. The bulk of the literature on energy- and entropy-stability preserving PMOR approaches focuses on the specific case of compressible flow. It is well-known that projection-based ROMs for compressible flow constructed via Galerkin projection in the L^2 inner product lack an *a priori* stability guarantee [24–26]. This problem can be circumvented for traditional intrusive ROMs through a variable transformation or

---

[1] As discussed in Section 6, extending the approach proposed herein to nonlinear manifold bases will be the subject of future work.

by changing the inner product in which the projection is done, yielding energy-stable [25–28] or entropy-stable [29–31] approaches. Alternate approaches for mitigating the problem which are less intrusive and possible to apply in conjunction with OpInf model reduction include subspace rotation [32] and eigenvalue reassignment [33,34]. An interesting and very recent pre-print that considers incompressible flow is the work of Klein and Sanderse [35], which develops a novel kinetic energy and momentum conserving hyper-reduction method for projection-based ROMs for the incompressible Navier–Stokes equations.

Conservation law preservation. A second problem arising in PMOR for fluid mechanics applications, and, more broadly, conservative systems of PDEs, is lack of conservation: ROMs constructed from conservative models are not guaranteed to maintain the underlying model's conservation laws. The following three references for mitigating this problem for intrusive projection-based ROMs are noteworthy. In [36], Carlberg et al. present a methodology for constructing conservative compressible flow ROMs by modifying the minimization problem defining the Least Squares Petrov–Galerkin (LSPG) [37] to include local or global conservation law constraints. The formulation in [36] is extended to the case of incompressible flow in [38], yielding a method that is both mass- and kinetic energy-conserving, and thus nonlinearly stable. An alternate way to create an incompressible flow ROM with mass and energy conservation is presented in [39]. In this work, the authors demonstrate that these properties can be attained at the ROM level through a careful selection of the boundary condition treatment and finite element space underlying the ROM. The preservation of conservation laws in OpInf remains an open problem, although progress has been made on problems with a variational form through [40,41], and this work.

Variational structure preservation. Another mathematical property mentioned previously and exhibited by a wide range of physical systems (e.g., solid dynamics, the shallow water equations, etc.), including the ones considered in the present work, is variational structure. This includes systems amenable to the standard Hamiltonian and Lagrangian formalisms, as well as the more general formalisms of, e.g., Euler–Poincaré, Lie–Poisson and metriplecticity. The advantages of biasing towards this structure are clear to see; for example, since the Hamiltonian can be considered a representation of the energy of a system, a Hamiltonian structure-preserving discretization will automatically obey at least one conservation law. Several Hamiltonian (or Lagrangian) structure-preserving approaches have been developed in recent years for the specific case of solid dynamics. In [42], it is shown that performing a Galerkin projection of the second-order-in-time Euler–Lagrange equations defining a canonical solid dynamics problem preserves Lagrangian structure, provided no hyper-reduction is employed. As discussed in [43], traditional hyper-reduction approaches such as collocation, the Discrete Empirical Interpolation Method (DEIM) [44] and gappy POD [45] destroy the Lagrangian structure of the ROM. In [43], two Lagrangian structure-preserving approaches for performing hyper-reduction on these systems, termed reduced basis sparsification (RBS) and matrix gappy POD, are proposed. Both approaches are of the "approximate-then-project" flavor, meaning they apply hyper-reduction to the nonlinear terms in the governing equations prior to projecting these terms onto a reduced basis. An alternate "project-then-approximate" approach for preserving Lagrangian structure in ROMs for nonlinear solid dynamics applications is the Energy-Conserving Sampling and Weighting (ECSW) method of Farhat et al. [46]. In this method, the nonlinear projected function is approximated using a set of points and weights, the latter set of which are obtained by solving a non-negative least-squares optimization problem. Interestingly, there has recently been some headway into OpInf techniques for PMOR on solid mechanical systems as well. In [41], the authors develop a gray-box method for learning the linear parts of Lagrangian systems in a way that respects the symmetric positive definite nature of the governing operators.

A broader class of symplecticity-preserving PMOR methods focus on directly reducing the Hamiltonian first-order-in-time system (1). As discussed in [42], performing a Galerkin projection of these equations onto a set of reduced basis vectors will generally not preserve the Hamiltonian/symplectic structure of the system. Several works, e.g., [47,48], propose to remedy this through Proper Symplectic Decomposition (PSD) and symplectic Galerkin projection. In [47], Peng et al. propose three algorithms for calculating the PSD, based on the cotangent lift, complex SVD and nonlinear programming. These algorithms effectively generate reduced bases such that projection onto the subspaces spanned by these bases will maintain symplecticity. Further, a version of DEIM for Hamiltonian systems, termed Symplectic DEIM (SDEIM), is developed for maintaining skew-symmetry (but not necessarily symplecticity or Hamiltonian structure) when performing hyper-reduction. An approach based on a globally optimal symplectic reduced basis in the sense of the PSD is derived in [49]. Here, it is shown the POD of a canonizable Hamiltonian system is automatically symplectic, from which the authors deduce optimality of the PSD. In [50], PSD is extended

to create a greedy approach for symplectic basis generation. The approach is advertised as more cost-effective than traditional POD and PSD, and exhibits exponentially-fast convergence. The follow-on work [51] presents a reduced dissipative Hamiltonian (RDH) method as a structure-preserving model reduction approach for Hamiltonian systems with dissipation. Unlike other approaches, the proposed approach enables the reduced system to be integrated using a symplectic integrator. The recent work [48], based on a lot of the same ideas as [50,51], demonstrates that linear symplectic maps can be used to guarantee that the reduced models inherit the geometric formulation from the full dynamics. The approach evolves the approximating symplectic reduced space in time along a trajectory locally constrained on the tangent space of the high-dimensional dynamics. The recent pre-print [52] presents a different DEIM-based hyper-reduction method for nonlinear parametric dynamical systems characterized by gradient fields such as Hamiltonian and port-Hamiltonian systems and gradient flows. The authors decompose the nonlinear part of the Hamiltonian into a sum of d terms, each characterized by a sparse dependence on the system state, and obtain a hyper-reduced approximation of the Jacobian by applying DEIM to the derived function. The resulting hyper-reduced model retains the gradient structure, and possesses *a priori* error estimates showing that the hyper-reduced model converges to the reduced model and the Hamiltonian is asymptotically preserved.

It is also possible to derive Hamiltonian structure-preserving ROMs using the classical POD reduced basis. In [53], a least-squares system is solved to ensure skew-symmetry of the POD-Galerkin system corresponding to the governing Hamiltonian form. An *a priori* error estimate for the resulting POD/Galerkin ROM is developed, but hyper-reduction is not considered, rendering the approach inefficient. In [54], Sockwell presents a Hamiltonian structure-preserving approach that is most closely related to the approach in [53] and that possesses similar error estimates; however, the technique in [54] is derived in a Hilbert space and takes advantage of the Hamiltonian framework in order to abstract the technique to a wide variety of weighted inner-product spaces. This method is shown to preserve linear Casimir invariants, and is demonstrated in the context of the rotating shallow water equations, commonly used in ocean modeling on the sphere. In addition to intrusive PMOR approaches, there is recent work in developing non-intrusive OpInf PMOR approaches that preserve Lagrangian [55] as well as Hamiltonian structure [16,40], and ROMs with nonlinear manifold (e.g., convolutional autoencoder) bases [56]. Notably, [16,40] are the only works to the authors' knowledge in which a Hamiltonian structure-preserving nonintrusive OpInf PMOR methodology is developed, although this method is limited to canonical Hamiltonian systems with a block-diagonal gradient structure.

Beyond Hamiltonian systems, it is worth mentioning some current references focusing on structure-preserving model reduction for port-Hamiltonian and metriplectic systems, e.g., [57,58], which are extensions of the Hamiltonian formalism to systems with dissipation. Metriplectic dynamical systems separate dynamics into terms that are "energy-preserving" and "dissipative", represented by a noncanonical Poisson structure and a degenerate Riemannian metric structure, respectively. To the best of our knowledge, [58] is the first paper to develop a structure-preserving (intrusive) ROM for PDEs with metriplectic structure. Conversely, the work [57] presents three techniques for constructing reduced bases for port-Hamiltonian systems: one based on POD, one based on $\mathcal{H}_2/\mathcal{H}_\infty$-derived optimized bases (which can be calculated without any snapshots), and one that is a mixture of the two. Interestingly, the approach in [57] is based on Petrov–Galerkin projection, rather than Galerkin projection.

Operator Inference. Data-driven operator inference originated in the seminal work of Peherstorfer and Willcox [21], which demonstrates that reduced operators in a projection-based ROM can be inferred non-intrusively (i.e., without access to the corresponding FOM operators or code) through the numerical solution of an optimization problem, given a set of FOM snapshots. An acknowledged deficiency of the original OpInf formulation is that it is only applicable to PDEs that contain low-order polynomial nonlinearities. As demonstrated in subsequent works [59–61], this shortcoming can be circumvented for many physical systems by using a technique known as "lifting", which defines a transformation of the state variables into auxiliary variables that make the governing PDEs linear or quadratic. The resulting approach, termed "Lift and Learn" [59] has been applied to a wide range of problems, including fluid mechanics and combustion [59,62–64], additive manufacturing [60], magnetohydrodynamics (MHD) [65], and solid mechanics [66].

During the past 1–2 years, researchers have begun to extend operator inference in several important directions. In [23], non-intrusive operator inference is extended to problems with non-polynomial nonlinearities given in analytic form, in a way that does require the definition of a lifting transformation. In several recent works, the group of Kramer et al. has developed OpInf methodologies that preserve Hamiltonian (or symplectic) [16,41] and Lagrangian [55] structure, to ensure energy-conserving ROMs. Note that, in the Hamiltonian case, all OpInf work

to date has been restricted to purely canonical systems (c.f. Section 2.1). A primary contribution of this work is the ability to treat both canonical and noncanonical systems of interest.

Other recent works have focused on improving the efficiency and robustness of the optimization problem underlying OpInf. It is well-known that this optimization problem generally requires regularization, and the results can be extremely sensitive to the choice of regularization parameters. Several researchers have begun to look at ways to optimize the choice of these regularization parameters. In [67], Guo et al. present a Bayesian approach to operator inference, in which the maximum marginal likelihood provides insight into the selection of the regularization parameters specified in the OpInf minimization problem. An alternate remedy known as nested operator inference is being pursued by Aretz et al. [68].

While OpInf originated in the context of ROMs in which the solution is approximated using an affine POD basis, the method has recently been extended to balanced truncation [69] and quadratic manifold bases [19,69]. The latter work [69] presents a symplecticity-preserving method based on quadratic manifold bases. The advantage of using quadratic manifold bases over linear bases is that it is often possible to represent the reduced solution using fewer basis vector, especially for problems exhibiting a slow decay of the Kolmogorov n-width [20].

Contributions of this manuscript. The present work extends the literature on structure-preserving OpInf techniques to include the linear operators governing general canonical and noncanonical Hamiltonian systems. Particularly, we contribute

- Non-intrusive methods based on OpInf for learning either: (1) the linear part of the Hamiltonian gradient in the case of canonical Hamiltonian systems, or (2) the constant part of the Poisson matrix in the case of noncanonical Hamiltonian systems. In contrast to previous work, these methods impose no restriction on the separability properties of the Hamiltonian or the algorithm used to compute the ROM basis.
- Theoretical analysis which guarantees that the learned operators converge to their intrusive counterparts in the limits of increasing basis size and increasing amounts of training data.
- Several numerical examples probing the behavior of these Hamiltonian OpInf ROMs in comparison to ROMs based on black-box OpInf as well as more intrusive PMOR techniques.

The remainder of this paper is organized as follows. Section 2 recalls preliminary information on Hamiltonian systems (including methods for their POD), as well as OpInf and average vector field time integration (c.f. [70]). Section 3 describes the present methods for canonical and noncanonical Hamiltonian OpInf, as well as their connection to previous work. Section 4 provides analysis showing that the proposed OpInf methods converge to their intrusive counterparts with the addition of snapshot data and basis modes. Finally, Section 5 provides numerical evidence for the Hamiltonian OpInf approaches in Section 3 using five example problems: a linear wave equation, a nonseparable but canonical quadratic Hamiltonian system, the Korteweg–De Vries (KdV) equation, the Benjamin–Bona–Mahony (BBM) equation, and a 3D linear elastic cantilever plate. Finally, some conclusions and future directions are discussed in Section 6.

2. Preliminaries

Here some preliminary information on Hamiltonian systems, as well as intrusive and nonintrusive methods of model reduction for such systems, is summarized.

2.1. Hamiltonian systems

The Hamiltonian formalism provides a mechanical framework encompassing a wide variety of conservative dynamical systems which arise from a variational principle. In particular, it reduces the problem of understanding a near-arbitrarily complicated dynamical system to the simpler problem of understanding a scalar-valued function H, called the Hamiltonian, and a skew-symmetric Poisson bracket $\{\cdot, \cdot\}$, which encodes a Lie algebra realization on functions. More formally, given a state vector $\mathbf{x} \in \mathbb{R}^N$, it follows that $\nabla \mathbf{x} = \mathbf{I}$, and so any Hamiltonian system can be written in the form

$$\dot{\mathbf{x}} = \{\mathbf{x}, H(\mathbf{x})\} = \nabla \mathbf{x} \cdot \mathbf{L}(\mathbf{x}) \nabla H(\mathbf{x}) = \mathbf{L}(\mathbf{x}) \nabla H(\mathbf{x}), \tag{1}$$

for some $H : \mathbb{R}^N \to \mathbb{R}$ and some potentially degenerate Poisson matrix $\mathbf{L} : \mathbb{R}^N \to \mathbb{R}^{N \times N}$, $\mathbf{L}^{\mathsf{T}} = -\mathbf{L}$ which is antisymmetric and satisfies the Jacobi identity,

$$\sum_{\ell=1}^{N} \left(L_{il} L_{jk,l} + L_{jl} L_{ki,l} + L_{kl} L_{ij,l} \right) = 0, \qquad 1 \le i, j, k \le N,$$

where L_{ij} are the components of $\mathbf{L}$ and $L_{ij,k}$ denotes the derivative with respect to the kth basis vector $\mathbf{e}_k \in \mathbb{R}^N$. From this, it is easy to see that the Poisson bracket (generated by $\mathbf{L}$) is also skew-symmetric, bilinear, and obeys a Leibniz rule. Moreover, the Hamiltonian H is a conserved quantity, since $\{H, H\} = 0$ by antisymmetry.

In the simplest case, Hamiltonian systems are dual (and equivalent) to their Lagrangian counterparts. To see this, consider a Lagrangian function $L(t, \mathbf{q}, \dot{\mathbf{q}})$ defined in terms of a position variable $\mathbf{q} \in \mathbb{R}^N$ and its associated velocity $\dot{\mathbf{q}}$. Then, under some regularity conditions (see e.g. [71]), there is a canonical Legendre transformation

$$H(t, \mathbf{q}, \mathbf{p}) = \sup_{\dot{\mathbf{q}}} \left(\langle \mathbf{p}, \dot{\mathbf{q}} \rangle - L(t, \mathbf{q}, \dot{\mathbf{q}}) \right),$$

which yields the conjugate momentum vector $\mathbf{p} = L_{\dot{\mathbf{q}}} := \nabla_{\dot{\mathbf{q}}} L \in \mathbb{R}^N$. Substituting $L = \langle \mathbf{p}, \dot{\mathbf{q}} \rangle - H$ in the usual action integral $S = \int L \, dt$ and computing the first variation now leads immediately to Hamilton's equations for the state $\mathbf{x} = \begin{pmatrix} \mathbf{q} & \mathbf{p} \end{pmatrix}^{\mathsf{T}} \in \mathbb{R}^{2N}$,

$$\dot{\mathbf{x}} = \begin{pmatrix} \dot{\mathbf{q}} \\ \dot{\mathbf{p}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} H_{\mathbf{q}} \\ H_{\mathbf{p}} \end{pmatrix} = \mathbf{J} \nabla H(\mathbf{x}),$$

similar to the above. Notice that $\mathbf{J}$ is anti-involutive and (trivially) satisfies the Jacobi identity, which implies that this Hamiltonian system is in canonical form. Conversely, systems of the form (1) for which $\mathbf{L} \ne \mathbf{J}$ are said to be noncanonical. Noncanonical Hamiltonian systems are quite flexible and have an important property: elements in the kernel of $\mathbf{L}$, called Casimirs, are invariant quantities, meaning that many (but not all) constants of motion in a noncanonical Hamiltonian system can be identified directly from its Poisson structure. Since Casimir invariants are often directly responsible for the long-time behavior of the system, it is important that they are appropriately respected by model reduction methods. It will be shown in Section 5 that the particular Hamiltonian OpInf methods developed in Section 3 attend to this issue at least as well as the current state of the art.

Although the Hamiltonian and Lagrangian formalisms can often be freely exchanged, many interesting dynamical systems which are readily modeled using the Hamiltonian formalism do not have an unconstrained Lagrangian formulation. For example, every completely integrable equation, including the KdV equation considered in Section 5, has a bi-Hamiltonian structure and therefore a singular Legendre transformation. Therefore, these systems can only be expressed in Lagrangian terms if the argument to the Lagrangian is constrained to be a derivative of the state variable (see, e.g., [72,73] for the case of KdV). This makes working directly with the Hamiltonian formulation of a dynamical system preferable in many cases, and encourages the search for model reduction techniques which are more general than those developed for canonical Hamiltonian systems. In particular, the nonintrusive methods of Section 3 are well adapted to noncanonical Hamiltonian systems and do not appeal to Lagrangians or Legendre transforms.

2.2. Proper orthogonal decomposition for Hamiltonian systems

Given a large semidiscrete Hamiltonian system (1), it is often necessary to perform model reduction in order to produce a feasible surrogate. Typically, this means constructing an informative reduced basis for the solution space to the system onto which the dynamics can be projected. While there are a variety of linear and nonlinear methods for accomplishing this task (including those in [11,74,75], to mention a few), this paper focuses on the linear technique known as Proper Orthogonal Decomposition (POD) which has seen the most widespread success. POD uses snapshots $\mathbf{x} \in \mathbb{R}^N$ of the high-fidelity model solution to construct a variance-maximizing subspace in which reduced solutions can be represented. To explain this more precisely, let $\mathbf{Y} \in \mathbb{R}^{N \times n_s}$ be a matrix with rank $r \le \min\{N, n_s\}$ containing n_s snapshots of the high-fidelity solution $\mathbf{y} = \mathbf{x} - \mathbf{x}_0 \in \mathbb{R}^N$ shifted by the initial condition $\mathbf{x}_0 := \mathbf{x}(0)$. Such snapshots could be collected at, e.g., discrete points in the interval $[0, T]$, where $T \in \mathbb{R}$ represents the final simulation time. If $\mathbf{Y} = \tilde{\mathbf{U}} \boldsymbol{\Sigma} \mathbf{V}^{\mathsf{T}}$ is the singular value decomposition of this mean-centered data matrix, standard computations show that the matrix $\mathbf{U} \in \mathbb{R}^{N \times n}$ comprised of the first $n < r$ columns of $\tilde{\mathbf{U}}$ minimizes

the $L^2([0, T])$ reconstruction error of $\mathbf{y}$, and that this error is precisely the sum of the remaining squared singular values [76]. More precisely, it follows that

$$\|\mathbf{y} - \mathbf{U}\mathbf{U}^\mathsf{T}\mathbf{y}\|^2 := \int_0^T |\mathbf{y} - \mathbf{U}\mathbf{U}^\mathsf{T}\mathbf{y}|^2 \, dt = \sum_{i=n+1}^{r} \sigma_i^2,$$

where σ_i is the ith singular value of $\mathbf{Y}$. This is the basis for the standard Galerkin POD-ROM (G-ROM) procedure, which is applied to the dynamical system governing $\mathbf{x}$ by making the approximation $\tilde{\mathbf{x}} = \mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}} \in \mathbb{R}^N$ for some unknown coefficients $\hat{\mathbf{x}} \in \mathbb{R}^n$ and using that $\mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{I}$ in $\mathbb{R}^n$. In particular, inserting this approximation into the Hamiltonian system (1) yields the update rule

$$\dot{\hat{\mathbf{x}}} = \mathbf{U}^\mathsf{T}\mathbf{L}(\tilde{\mathbf{x}})\nabla H(\tilde{\mathbf{x}}),$$

which is low-order but obviously not Hamiltonian since $\mathbf{U}^\mathsf{T}\mathbf{L} \neq -(\mathbf{U}^\mathsf{T}\mathbf{L})^\mathsf{T} = \mathbf{L}\mathbf{U}$. An effective remedy for this is the strategy developed in [77], which solves the overdetermined least-squares problem $\mathbf{U}^\mathsf{T}\mathbf{L} = \hat{\mathbf{L}}\mathbf{U}^\mathsf{T}$ for $\hat{\mathbf{L}} = \mathbf{U}^\mathsf{T}\mathbf{L}\mathbf{U}$, yielding a skew-symmetric operator which is guaranteed to produce dynamics which preserve the reduced Hamiltonian $\hat{H} = H \circ \tilde{\mathbf{x}} \colon \mathbb{R}^n \to \mathbb{R}$. To see this, consider solving the Hamiltonian POD-ROM (H-ROM)

$$\dot{\hat{\mathbf{x}}} = \hat{\mathbf{L}}(\tilde{\mathbf{x}})\nabla \hat{H}(\hat{\mathbf{x}}).$$

Then, it follows that the change in the value of the reduced Hamiltonian along a solution is given by

$$\dot{\hat{H}} = \dot{\hat{\mathbf{x}}} \cdot \nabla \hat{H} = \hat{\mathbf{L}}\nabla \hat{H} \cdot \nabla \hat{H} = -\hat{\mathbf{L}}\nabla \hat{H} \cdot \nabla \hat{H} = 0,$$

so that this quantity is exactly preserved up to time discretization error.

While noncanonical Hamiltonian systems are (thus far) limited to variants of the "ordinary" POD basis construction described above, it turns out that there are several useful ways to construct the POD basis $\mathbf{U}$ in the case of canonical Hamiltonian systems. In particular, when $N = 2M$ for some $M \in \mathbb{N}$ and $\mathbf{x} = \begin{pmatrix} \mathbf{q} & \mathbf{p} \end{pmatrix}^\mathsf{T}$ separates nicely into position and momentum variables, it is frequently useful to use a basis built block-wise from sections of the snapshot data $\mathbf{Y}$. This is particularly true in the presence of scale separation, where the variance in one of $\mathbf{q}, \mathbf{p}$ will be dominated by the other if a standard POD basis for the full field $\begin{pmatrix} \mathbf{q} & \mathbf{p} \end{pmatrix}^\mathsf{T}$ is used [31,78]. In this case, separating the data $\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_q & \mathbf{Y}_p \end{pmatrix}^\mathsf{T}$ into $M \times n_s$ blocks and carrying out the POD procedure described before yields separate bases $\mathbf{U}_q, \mathbf{U}_p \in \mathbb{R}^{N \times m}$ for position and momentum, which can be combined into the block basis $\mathbf{U} = \mathrm{Diag}(\mathbf{U}_q, \mathbf{U}_p)$ of size $N \times n$ where $n = 2m$. This has the effect of normalizing the importance of $\mathbf{q}$ and $\mathbf{p}$ in the dimension reduction, often leading to better performance in the associated ROMs. As an added benefit, notice that both $\mathbf{U}_q^\mathsf{T}\mathbf{Y}_q\mathbf{Y}_q^\mathsf{T}\mathbf{U}_q$ and $\mathbf{U}_p^\mathsf{T}\mathbf{Y}_p\mathbf{Y}_p^\mathsf{T}\mathbf{U}_p$ are diagonal under this construction, since each POD basis is drawn from the SVD of the snapshots.

In addition to this, another block basis construction which has been demonstrably useful in the model reduction of canonical Hamiltonian systems is known as the "cotangent lift" algorithm from [47]. This procedure constructs a basis such that $\mathbf{U}^\mathsf{T}\mathbf{J} = \mathbf{J}_n\mathbf{U}^\mathsf{T}$ (for $\mathbf{J}_n \in \mathbb{R}^{n \times n}$ the canonical Poisson matrix of dimension n) by choosing $\mathbf{U}$ from the left singular vectors of the concatenated snapshot matrix $\begin{pmatrix} \mathbf{q} & \mathbf{p} \end{pmatrix} \approx \bar{\mathbf{U}}\boldsymbol{\Sigma}\mathbf{V}^\mathsf{T} \in \mathbb{R}^{M \times 2n_s}$. More precisely, if $\bar{\mathbf{U}} \in \mathbb{R}^{M \times m}$ contains the first M left singular vectors, the basis $\mathbf{U} = \mathrm{Diag}(\bar{\mathbf{U}}, \bar{\mathbf{U}})$ satisfies the required condition. This is quite a useful construction, as it follows that $\mathbf{U}^\mathsf{T}\mathbf{J}\nabla H = \mathbf{J}_n\mathbf{U}^\mathsf{T}\nabla H = \mathbf{J}_n\nabla \hat{H}$ and hence the prototypical G-ROM is converted into an H-ROM. On the other hand, it is clear that $\bar{\mathbf{U}}^\mathsf{T}\mathbf{Y}_q\mathbf{Y}_q^\mathsf{T}\bar{\mathbf{U}}$ is not diagonal (and same for p), since $\mathbf{V}^\mathsf{T} = \begin{pmatrix} \mathbf{V}_1 & \mathbf{V}_2 \end{pmatrix}^\mathsf{T} \in \mathbb{R}^{m \times 2n_s}$ and so $\mathbf{V}_1^\mathsf{T}\mathbf{V}_1 \neq \mathbf{V}_2^\mathsf{T}\mathbf{V}_2 \neq \mathbf{V}_1^\mathsf{T}\mathbf{V}_1 + \mathbf{V}_2^\mathsf{T}\mathbf{V}_2 = \mathbf{I}$.

2.3. Generic operator inference

Consider a dynamical system of the form

$$\dot{\mathbf{x}}(t, \boldsymbol{\mu}) = \mathbf{F}(t, \boldsymbol{\mu}, \mathbf{x}(t, \boldsymbol{\mu})), \tag{2}$$

where $\mathbf{x} \colon \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}^N$ is a time-dependent state variable and $\boldsymbol{\mu} \in \mathbb{R}^p$ is a vector of parameters. As mentioned previously, constructing a POD basis $\mathbf{U} \in \mathbb{R}^{N \times n}$ and making the approximation $\tilde{\mathbf{x}} = \mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}}$ leads to the canonical Galerkin ROM,

$$\dot{\hat{\mathbf{x}}}(t, \boldsymbol{\mu}) = \mathbf{U}^\mathsf{T}\mathbf{F}(t, \mu, \tilde{\mathbf{x}}(t, \boldsymbol{\mu})), \tag{3}$$

which is an n-dimensional dynamical system describing the evolution of the POD basis coefficients. While this procedure is well studied and often effective, it clearly requires intrusion into the FOM simulation code via access to the operators governing the high-fidelity system (2), since it is necessary to assemble $\mathbf{U}^\mathsf{T}\mathbf{F}(\tilde{\mathbf{x}})$. In the case that $\mathbf{x}_0 = \mathbf{0}$ and $\mathbf{F}(\mathbf{x}) = \mathbf{D}\mathbf{x}$ is linear, this means direct access to $\mathbf{D} \in \mathbb{R}^{N \times N}$ is needed in order to assemble the reduced operator $\hat{\mathbf{D}} = \mathbf{U}^\mathsf{T}\mathbf{D}\mathbf{U} \in \mathbb{R}^{n \times n}$. However, it is frequently impossible (or prohibitively expensive) to access this information, due to, e.g., complicated or proprietary legacy codes. This motivates the black-box operator inference technique (OpInf) of [6] which is used for the non-intrusive modeling of dynamical systems such as (2). More precisely, OpInf uses snapshot data to learn the tensor coefficients of a polynomial approximation $\mathbf{D}_0, \ldots, \mathbf{D}_n$ to the action of $\mathbf{F}$, so that if $\tilde{\mathbf{x}}$ satisfies

$$\dot{\mathbf{z}} = \mathbf{D}_0 + \mathbf{D}_1\mathbf{z} + \mathbf{D}_2\left(\mathbf{z} \otimes \mathbf{z}\right) + \cdots + \mathbf{D}_n\left(\mathbf{z} \otimes \cdots \otimes \mathbf{z}\right),$$

then $\mathbf{z} \approx \mathbf{x}$ remains close to a solution to the original system. This ansatz is clearly exact in the case that the model in question is polynomial (or differentially polynomial), but has been shown to be useful even outside of this case, see, e.g., [22,63] and lifting transformations [59,59,62–64]. Moreover, it readily extends to learning the coefficients of a POD-based ROM, since reduced basis projection preserves this polynomial structure.

To see this in detail, consider learning a linear approximation $\dot{\mathbf{z}} = \mathbf{D}\mathbf{z}$ (for simplicity), and suppose an $N \times n_s$ matrix $\mathbf{X}$ of snapshot data is provided containing (partial) trajectories of the original system (2). Then, an approximation $\mathbf{X}_t \approx \dot{\mathbf{X}}$ to the temporal derivative of each snapshot can be formed through, e.g., finite differences, and the matrix least-squares problem,

$$\underset{\mathbf{D} \in \mathbb{R}^{N \times N}}{\operatorname{argmin}} R\left(\mathbf{D}\right) = \underset{\mathbf{D} \in \mathbb{R}^{N \times N}}{\operatorname{argmin}} \left|\mathbf{X}_t - \mathbf{D}\mathbf{X}\right|^2,$$

can be solved to yield the desired operator $\mathbf{D}$. More precisely, exterior differentiation yields

$$dR\left(\mathbf{D}\right) = -2\left\langle \mathbf{X}_t - \mathbf{D}\mathbf{X}, d\mathbf{D}\,\mathbf{X}\right\rangle = -2\left\langle \left(\mathbf{X}_t - \mathbf{D}\mathbf{X}\right)\mathbf{X}^\mathsf{T}, d\mathbf{D}\right\rangle, = \left\langle \nabla R\left(\mathbf{D}\right), d\mathbf{D}\right\rangle,$$

so that solving $\nabla R\left(\mathbf{D}\right) = -2\left(\mathbf{X}_t - \mathbf{D}\mathbf{X}\right)\mathbf{X}^\mathsf{T} = \mathbf{0}$ reduces to solving the linear system

$$\mathbf{D}\mathbf{X}\mathbf{X}^\mathsf{T} = \mathbf{X}_t\mathbf{X}^\mathsf{T}.$$

Alternatively, there is the equivalent vectorized system

$$\left(\mathbf{X}\mathbf{X}^\mathsf{T} \otimes \mathbf{I}\right)\operatorname{vec}\mathbf{D} = \operatorname{vec}\left(\mathbf{X}_t\mathbf{X}^\mathsf{T}\right),$$

where $\otimes$ denotes Kronecker's matricized tensor product and equivalence follows via the "vec trick" (see Appendix A.1 for a review of these ideas).

On the other hand, in practical application settings, it is usually undesirable (or even infeasible) to infer the full $N \times N$ operator $\mathbf{D}$ in this way, as this requires solving a linear system which scales with N^2. Instead, it is more useful to combine OpInf with dimension reduction techniques such as POD, since, if $\tilde{\mathbf{x}} = \mathbf{U}\hat{\mathbf{x}}$ where $\mathbf{U} \in \mathbb{R}^{N \times n}$ is a POD basis and $\hat{\mathbf{x}} \in \mathbb{R}^n$, then the snapshot data $\mathbf{X}$ and its approximate time derivative $\mathbf{X}_t$ can be projected onto this basis before inferring a reduced operator. In particular, there are the $n \times n_s$ matrices $\hat{\mathbf{X}} = \mathbf{U}^\mathsf{T}\mathbf{X}$ and $\hat{\mathbf{X}}_t = \mathbf{U}^\mathsf{T}\mathbf{X}_t$, which can be used to infer a lower-dimensional operator $\hat{\mathbf{D}} \in \mathbb{R}^{n \times n}$ governing the non-intrusive reduced dynamical system $\dot{\hat{\mathbf{x}}} = \hat{\mathbf{D}}\hat{\mathbf{x}}$. In this case, $\hat{\mathbf{D}}$ is inferred through the reduced OpInf problem of size n,

$$\underset{\hat{\mathbf{D}} \in \mathbb{R}^{n \times n}}{\operatorname{argmin}} \left|\hat{\mathbf{X}}_t - \hat{\mathbf{D}}\hat{\mathbf{X}}\right|^2,$$

which is solved as described above.

Besides reducing computational costs relative to inference of the full operator $\mathbf{D}$, inferring the reduced operator $\hat{\mathbf{D}}$ has the following added benefit due to the hierarchical order of the columns of $\mathbf{U}$. While this result appears to be well known, the lack of a standard reference has motivated the inclusion of a proof in Appendix A.2.

Proposition 2.1. *Suppose* $\mathbf{U} \in \mathbb{R}^{N \times n}$ *is the matrix of left singular vectors of some data matrix* $\mathbf{X}$*, and* $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ *is the corresponding diagonal matrix of (nonzero) singular values* $\{\sigma_j\}_{j=1}^n$*. Then, the unique solution to the OpInf problem of size n is given by*

$$\hat{\mathbf{D}} = \underset{\mathbf{D} \in \mathbb{R}^{n \times n}}{\operatorname{argmin}} \left|\hat{\mathbf{X}}_t - \mathbf{D}\hat{\mathbf{X}}\right|^2 = \hat{\mathbf{X}}_t\hat{\mathbf{X}}^\mathsf{T}\boldsymbol{\Sigma}^{-2}.$$

Moreover, for any $n' < n$, the submatrix $\hat{\mathbf{D}}' \in \mathbb{R}^{n' \times n'}$ formed by extracting the first n rows and columns of $\hat{\mathbf{D}}$ is the solution to the corresponding OpInf problem of size n.

Remark 2.1. Note that the conclusion of Proposition 2.1 continues to hold if the minimization objective is Tikhonov regularized by a multiple of $\hat{\mathbf{D}}$, as can be checked by considering the minimization objective $\left|\hat{\mathbf{X}}_t - \hat{\mathbf{D}}\hat{\mathbf{X}}\right|^2 + \eta\left|\hat{\mathbf{D}}\right|^2$ for some $\eta > 0$ and repeating the arguments above. Moreover, the conclusion also holds block-wise if $\mathbf{U}$ is a block basis as discussed in Section 2.2 and $\hat{\mathbf{D}}$ is block diagonal, since the relevant problem decouples over the blocks of $\hat{\mathbf{D}}$.

2.4. Previous Hamiltonian operator inference

The idea of using OpInf in conjunction with Hamiltonian systems has been previously explored in [40], where it was used to learn the linear part of the Hamiltonian gradient ∇H for a sub-class of canonical Hamiltonian systems which are known as separable.[2] The separability assumption implies that the system Hamiltonian decomposes as $H(\mathbf{q}, \mathbf{p}) = T(\mathbf{q}) + V(\mathbf{p})$ for some real-valued functions T, V depending only on $\mathbf{q}$, $\mathbf{p}$, respectively. A consequence of this is that the linear part of the gradient ∇H becomes block-diagonal in the variables $\mathbf{q}$, $\mathbf{p}$, a fact which is preserved at the POD-ROM level as long as a cotangent lift POD basis $\mathbf{U} \in \mathbb{R}^{N \times n}$ satisfying $\mathbf{U}^\mathsf{T}\mathbf{J}\mathbf{U} = \mathbf{J}_n$ is employed, where $\mathbf{J}_n$ is the canonical symplectic matrix of dimension $n = 2m$. With this additional restriction on the reduced basis, it follows that the intrusive H-ROM for the approximation $\tilde{\mathbf{x}} = \mathbf{U}\hat{\mathbf{x}} \approx \mathbf{x}$ decouples over $\mathbf{q}$, $\mathbf{p}$, becoming $\dot{\hat{\mathbf{x}}} = \mathbf{J}_n \nabla \hat{H}(\tilde{\mathbf{x}}) = \mathbf{J}_n \left(\hat{\mathbf{A}}\hat{\mathbf{x}} + \nabla \hat{f}(\tilde{\mathbf{x}})\right)$ where $\hat{\mathbf{A}} = \text{diag}\left(\hat{\mathbf{A}}_\mathbf{q}, \hat{\mathbf{A}}_\mathbf{p}\right)$ is block-diagonal and we have written $H(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} + f(\mathbf{x})$. This allows the authors of [40] to formulate an inference procedure for the linear operator $\hat{\mathbf{A}}$ which decouples block-wise into two $m^2 \times m^2$ subproblems for $\hat{\mathbf{A}}_\mathbf{q}, \hat{\mathbf{A}}_\mathbf{p}$, provided snapshots of ∇f can be obtained and this quantity can be simulated online. More precisely, given snapshots $\mathbf{X} \in \mathbb{R}^{N \times n_s}$ of the full order solution along with snapshots $\nabla f(\mathbf{X}) \in \mathbb{R}^{N \times n_s}$ of the nonlinear part of ∇H, the problems to solve are

$$\underset{\hat{\mathbf{A}} \in \mathbb{R}^{m \times m}}{\arg\min} \left|\hat{\mathbf{X}}_{\mathbf{p},t} + \hat{\mathbf{A}}\hat{\mathbf{X}}_\mathbf{q} + \hat{\nabla}_\mathbf{q} f(\mathbf{X})\right|^2, \quad \text{s.t.} \quad \hat{\mathbf{A}}^\mathsf{T} = \hat{\mathbf{A}},$$

$$\underset{\hat{\mathbf{A}} \in \mathbb{R}^{m \times m}}{\arg\min} \left|\hat{\mathbf{X}}_{\mathbf{q},t} - \hat{\mathbf{A}}\hat{\mathbf{X}}_\mathbf{p} - \hat{\nabla}_\mathbf{p} f(\mathbf{X})\right|^2 \quad \text{s.t.} \quad \hat{\mathbf{A}}^\mathsf{T} = \hat{\mathbf{A}},$$

where subscripts on $\mathbf{q}$, $\mathbf{p}$ denote either the first or second m rows in the snapshot matrix, $\nabla_\mathbf{q}$, $\nabla_\mathbf{p}$ denote partial derivatives, subscript t denotes a finite difference approximation to the time derivative, and "hat" indicates the application of a basis projection $\mathbf{U}^\mathsf{T}$. This yields $\mathbf{A}_\mathbf{q}$ and $\mathbf{A}_\mathbf{p}$, respectively, which once learned can be used to simulate the differential equation for $\dot{\hat{\mathbf{x}}}$ as usual to yield the approximation $\tilde{\mathbf{x}}$. While this procedure has been previously useful for simulating several systems of interest, it will be shown in Section 3 that the restrictions inherent in this algorithm can be removed, leading to a Hamiltonian OpInf ROM for canonical systems which does not require separability of the quadratic terms in the Hamiltonian or a specific choice of reduced basis. This extends the applicability of canonical Hamiltonian operator inference methods to, e.g., problems in molecular dynamics involving temperature-dependent potentials [79] or the estimation of Schwarzschild geodesics [80] in general relativity.

2.5. Linear ROMs and average vector field integration

To simplify the presentation of later results, it is worth mentioning some facts about linear ROMs and the particular timestepping scheme used in this work. First, note that the average vector field (AVF) method [70,81] is employed for time integration of all numerical examples, meaning that the Hamiltonian dynamical system (1) is discretized as

$$\frac{\mathbf{x}^{k+1} - \mathbf{x}^k}{\Delta t} = \mathbf{L}\left(\mathbf{x}^{k+\frac{1}{2}}\right) \int_0^1 \nabla H\left(t\mathbf{x}^{k+1} + (1-t)\mathbf{x}^k\right) dt,$$

---

[2] Since the nonlinear part of ∇H is assumed to be available intrusively, the method in [40] actually requires the weaker condition of separability in the quadratic part of the Hamiltonian.

where $\mathbf{x}^{k+\frac{1}{2}} = \frac{1}{2}\left(\mathbf{x}^k + \mathbf{x}^{k+1}\right)$, which amounts to linearizing the trajectory of the state between time steps k and $k+1$ and fixing evaluation of $\mathbf{L}$ at the midpoint. This integration scheme has appealing properties including exact quadrature for polynomial nonlinearities, as well as second-order convergence in time. Moreover, it is easy to see that AVF integration is globally energy-conserving: if $\boldsymbol{\ell}(t) = t\mathbf{x}^{k+1} + (1-t)\mathbf{x}^k$, it follows from the symmetry relation $\mathbf{L}^\intercal = -\mathbf{L}$ that

$$\frac{H\left(\mathbf{x}^{k+1}\right) - H\left(\mathbf{x}^k\right)}{\Delta t} = \frac{1}{\Delta t} \int_0^1 \frac{d}{dt} H\left(\boldsymbol{\ell}(t)\right) dt = \frac{\mathbf{x}^{k+1} - \mathbf{x}^k}{\Delta t} \cdot \int_0^1 \nabla H\left(\boldsymbol{\ell}(t)\right) dt$$

$$= \left[\mathbf{L}\left(\mathbf{x}^{k+\frac{1}{2}}\right) \int_0^1 \nabla H\left(\boldsymbol{\ell}(t)\right) dt\right] \cdot \int_0^1 \nabla H\left(\boldsymbol{\ell}(t)\right) dt = 0,$$

so that there can be no loss of energy during AVF timestepping.

Now, when $\dot{\mathbf{x}} = \mathbf{D}\mathbf{x}$ is linear, it is clear that AVF integration reduces to the implicit midpoint method

$$\frac{\mathbf{x}^{k+1} - \mathbf{x}^k}{\Delta t} = \mathbf{D}\mathbf{x}^{k+\frac{1}{2}},$$

which can be easily solved at each time step k by writing $\mathbf{x}^{k+1} = \mathbf{x}^k + \Delta\mathbf{x}^k$, where $\Delta\mathbf{x}^k = \mathbf{x}^{k+1} - \mathbf{x}^k$ satisfies the linear system

$$\left(\mathbf{I} - \frac{\Delta t}{2}\mathbf{D}\right)\Delta\mathbf{x}^k = \Delta t\,\mathbf{D}\mathbf{x}^k.$$

Note additionally that if $\tilde{\mathbf{x}} = \mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}} \approx \mathbf{x}$ is a mean-centered Galerkin projection and $\hat{\mathbf{D}} = \mathbf{U}^\intercal\mathbf{D}\mathbf{U}$ is the intrusive projection of $\mathbf{D}$, this implies the low-order update formula $\hat{\mathbf{x}}^{k+1} = \hat{\mathbf{x}}^k + \Delta\hat{\mathbf{x}}^k$, where

$$\left(\hat{\mathbf{I}} - \frac{\Delta t}{2}\hat{\mathbf{D}}\right)\Delta\hat{\mathbf{x}}^k = \Delta t\left(\mathbf{U}^\intercal\mathbf{D}\mathbf{x}_0 + \hat{\mathbf{D}}\hat{\mathbf{x}}^k\right),$$

which is an $n \times n$ linear solve leading to the full-order approximate $\tilde{\mathbf{x}}^{k+1} = \mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}}^{k+1}$. Of course, in the event that $\mathbf{D}$ is not available and so $\hat{\mathbf{D}}$ must be inferred, this mean-centering can be ignored. Finally, to specify these expressions to linear Hamiltonian systems $\dot{\mathbf{x}} = \mathbf{L}\mathbf{A}\mathbf{x}$, it suffices to replace $\hat{\mathbf{D}} = \widehat{\mathbf{L}\mathbf{A}} = \mathbf{U}^\intercal\mathbf{L}\mathbf{A}\mathbf{U}$ in the case of the G-ROM and $\hat{\mathbf{D}} = \hat{\mathbf{L}}\hat{\mathbf{A}} = \mathbf{U}^\intercal\mathbf{L}\mathbf{U}\mathbf{U}^\intercal\mathbf{A}\mathbf{U}$ in the case of the H-ROM.

3. Hamiltonian operator inference

It is now possible to discuss the present methods for canonical and noncanonical Hamiltonian OpInf, which will be referred to as C-H-OpInf and NC-H-OpInf, respectively. First, note the following computational result central to these techniques.

Theorem 3.1. *Let* $\mathbf{A} \in \mathbb{R}^{N \times N}$, $\mathbf{B}, \mathbf{C} \in \mathbb{R}^{N \times n_s}$, *and define* $\mathbf{A}\bar{\oplus}\mathbf{B} = \mathbf{A} \otimes \mathbf{B} + \mathbf{B} \otimes \mathbf{A}$. *Then, every solution to the symmetry-constrained least-squares regularization problem*

$$\underset{\mathbf{D} \in \mathbb{R}^{N \times N}}{\mathrm{argmin}} |\mathbf{C} - \mathbf{A}\mathbf{D}\mathbf{B}|^2, \quad \text{s.t.} \quad \mathbf{D}^\intercal = \pm\mathbf{D},$$

corresponds to a solution to the vectorized problem

$$\left(\mathbf{A}^\intercal\mathbf{A}\bar{\oplus}\mathbf{B}\mathbf{B}^\intercal\right)\mathrm{vec}\,\mathbf{D} = \mathrm{vec}\left(\mathbf{A}^\intercal\mathbf{C}\mathbf{B}^\intercal \pm \mathbf{B}\mathbf{C}^\intercal\mathbf{A}\right).$$

In particular, the first system is uniquely solvable if and only if the second is also, which holds whenever $\mathbf{A}, \mathbf{B}$ *have maximal rank.*

Proof. First, note that the uniqueness condition follows immediately from the fact that the objective is convex, the symmetry constraint is linear, and $\mathrm{rank}\left(\mathbf{B} \otimes \mathbf{A}\right) = \mathrm{rank}\left(\mathbf{B}\right)\mathrm{rank}(\mathbf{A})$. The remainder will follow from a direct calculation using the method of Lagrange multipliers. More precisely, define the Lagrangian $L\left(\mathbf{D}, \boldsymbol{\Lambda}\right) = \frac{1}{2}\left|\mathbf{C} - \mathbf{A}\mathbf{D}\mathbf{B}\right|^2 + \langle\boldsymbol{\Lambda}, \mathbf{D} \mp \mathbf{D}^\intercal\rangle$ where $\boldsymbol{\Lambda} \in \mathbb{R}^{N \times N}$ is a matrix of Lagrange multipliers. Then, exterior differentiation yields

$$dL\left(\mathbf{D}, \boldsymbol{\Lambda}\right) = -\langle\mathbf{C} - \mathbf{A}\mathbf{D}\mathbf{B}, \mathbf{A}\,d\mathbf{D}\,\mathbf{B}\rangle + \langle\boldsymbol{\Lambda}, d\mathbf{D} \mp d\mathbf{D}^\intercal\rangle + \langle d\boldsymbol{\Lambda}, \mathbf{D} \mp \mathbf{D}^\intercal\rangle$$

$$= \langle d\mathbf{D}, -\mathbf{A}^\mathsf{T} (\mathbf{C} - \mathbf{ADB}) \mathbf{B}^\mathsf{T} + \Lambda \mp \Lambda^\mathsf{T} \rangle + \langle d\Lambda, \mathbf{D} \mp \mathbf{D}^\mathsf{T} \rangle.$$

Setting this to zero yields the first-order optimality conditions

$$\mathbf{A}^\mathsf{T} (\mathbf{C} - \mathbf{ADB}) \mathbf{B}^\mathsf{T} = \Lambda \mp \Lambda^\mathsf{T},$$
$$\mathbf{D} \mp \mathbf{D}^\mathsf{T} = \mathbf{0}.$$

Examining the first condition the right-hand side implies symmetry in the left-hand side, allowing for easy elimination of Λ through the expression

$$\mathbf{A}^\mathsf{T} (\mathbf{C} - \mathbf{ADB}) \mathbf{B}^\mathsf{T} \pm \mathbf{B} (\mathbf{C} - \mathbf{ADB})^\mathsf{T} \mathbf{A} = \mathbf{0}.$$

Expanding the above and using the second condition $\mathbf{D} \mp \mathbf{D}^\mathsf{T} = \mathbf{0}$ then yields

$$\mathbf{A}^\mathsf{T}\mathbf{ADBB}^\mathsf{T} \pm \mathbf{BB}^\mathsf{T}\mathbf{D}^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{A} = \mathbf{A}^\mathsf{T}\mathbf{ADBB}^\mathsf{T} + \mathbf{BB}^\mathsf{T}\mathbf{DA}^\mathsf{T}\mathbf{A} = \mathbf{A}^\mathsf{T}\mathbf{CB}^\mathsf{T} \pm \mathbf{BC}^\mathsf{T}\mathbf{A},$$

which vectorizes through the "vec trick" (c.f. Appendix A.1) to yield the claimed result. □

Theorem 3.1 provides the solution to a generic symmetric or skew-symmetric operator inference problem, which will be seen to include the C-H-OpInf and NC-H-OpInf procedures employed presently. As mentioned before, a notable benefit of the generic OpInf procedure is that its solutions satisfy Proposition 2.1, meaning that a solution computed using a reduced basis of size n remains optimal via truncation for all $n' < n$. The next result shows that, under some (fairly strong) assumptions on $\mathbf{A}, \mathbf{B}$, this "one-shot" ability continues to hold for the system in Theorem 3.1. Since the proof is straightforward but technical, it is deferred to Appendix A.2.

Proposition 3.1. *Let $\mathbf{U} \in \mathbb{R}^{N \times n}$ be a POD basis. Suppose $\hat{\mathbf{D}} \in \mathbb{R}^{n \times n}$ uniquely solves the optimization problem in Theorem 3.1 for given $\hat{\mathbf{A}} = \mathbf{U}^\mathsf{T}\mathbf{AU} \in \mathbb{R}^{n \times n}$ and $\hat{\mathbf{B}}, \hat{\mathbf{C}} \in \mathbb{R}^{n \times n_s}$ defined by $\hat{\mathbf{B}} = \mathbf{U}^\mathsf{T}\mathbf{B}, \hat{\mathbf{C}} = \mathbf{U}^\mathsf{T}\mathbf{C}$. Let $n' < n$, and for any matrix $\hat{\mathbf{M}}$ which is multiplied with the POD basis $\mathbf{U}$, let $\hat{\mathbf{M}}'$ denote the submatrix obtained by removing the $n - n'$ highest-frequency basis vectors of $\mathbf{U}$ from every relevant multiplication. If $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}\hat{\mathbf{B}}^\mathsf{T}$ are both diagonal, then the unique solution to*

$$\underset{\mathbf{D} \in \mathbb{R}^{n' \times n'}}{\operatorname{argmin}} \left| \hat{\mathbf{C}}' - \hat{\mathbf{A}}'\mathbf{D}\hat{\mathbf{B}}' \right|^2, \quad \text{s.t.} \quad \mathbf{D}^\mathsf{T} = \pm\mathbf{D},$$

is given by the truncation $\hat{\mathbf{D}}'$.

Proof. See Appendix A.2. □

While Proposition 3.1 is useful to know, it is worth mentioning that its conclusion generally does not hold for any of the structure-preserving OpInf methods known to date, including those discussed here. Indeed, while the diagonality of $\hat{\mathbf{A}}$ can often be arranged, it is more difficult to construct a suitable $\hat{\mathbf{B}}\hat{\mathbf{B}}^\mathsf{T}$ which is diagonal. On the other hand, there are many cases when the truncated solution to Theorem 3.1 is close enough to optimal to produce a well performing ROM, making it useful to employ truncation without the guarantee of Proposition 3.1 provided this property is empirically verified.

3.1. Canonical Hamiltonian systems

The first goal is to present an OpInf method applicable to canonical Hamiltonian systems, and connect it to previous work in [40]. Suppose snapshots of the form $\mathbf{x} = \begin{pmatrix} \mathbf{q} & \mathbf{p} \end{pmatrix}^\mathsf{T}$ can be obtained, say, as the result of post-processing data from a Lagrangian system via a Legendre transformation (c.f. Sections 2.1 and 5.5). Then, given that $\mathbf{q}, \mathbf{p}$ are the canonical position and momentum variables, it must be true that $\mathbf{L} = \mathbf{J}$ in (1) and the Hamiltonian system to be modeled is in canonical form. In this case, a Hamiltonian OpInf procedure can be considered which requires only knowledge of the nonlinear part of ∇H. To see this, recall that the discrete Hamiltonian can be expressed as

$$H(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{Ax} + f(\mathbf{x}),$$

for a symmetric, potentially unknown $\mathbf{A} \in \mathbb{R}^{N \times N}$, and a known nonlinear function $\mathbf{f}: \mathbb{R}^N \to \mathbb{R}$. It follows that the gradient is given by $\nabla H(\mathbf{x}) = \mathbf{A}\mathbf{x} + \nabla f(\mathbf{x})$, and any POD basis $\mathbf{U} \in \mathbb{R}^{N \times n}$ yields a reduced Poisson operator $\hat{\mathbf{J}} = \mathbf{U}^\mathsf{T}\mathbf{J}\mathbf{U}$ corresponding to the H-ROM discussed in Section 2.2. Notice that $\mathbf{J}$ has a canonical form, so that this operator can be computed without intrusion into any simulation code. Making the obvious Galerkin projection $\tilde{\mathbf{x}} = \mathbf{U}\hat{\mathbf{x}}$ and writing $\hat{H} = H \circ \tilde{\mathbf{x}}$, $\hat{\mathbf{f}} = \mathbf{f} \circ \tilde{\mathbf{x}}$ then yields the reduced Hamiltonian

$$\hat{H}\left(\hat{\mathbf{x}}\right) = \frac{1}{2}\hat{\mathbf{x}}^\mathsf{T}\hat{\mathbf{A}}\hat{\mathbf{x}} + \hat{f}\left(\hat{\mathbf{x}}\right),$$

which depends on the symmetric, potentially unknown reduced operator $\hat{\mathbf{A}} \in \mathbb{R}^{n \times n}$. Provided $\hat{\mathbf{A}}$ can be computed or inferred, access to $\nabla \mathbf{f}$ then implies solvability of the H-ROM

$$\dot{\hat{\mathbf{x}}} = \hat{\mathbf{J}}\nabla\hat{H}(\hat{\mathbf{x}}) = \hat{\mathbf{J}}\left(\hat{\mathbf{A}}\hat{\mathbf{x}} + \nabla\hat{f}(\hat{\mathbf{x}})\right), \tag{4}$$

which will be a low-order Hamiltonian system approximating the original dynamics.

Since $\mathbf{A}$ is unavailable in the present setting, (4) is most readily solved by setting up a tractable inference problem for $\hat{\mathbf{A}}$. This means forming the appropriate reduced quantities from snapshot data and solving the constrained least-squares problem

$$\underset{\hat{\mathbf{A}} \in \mathbb{R}^{n \times n}}{\operatorname{argmin}} \left|\hat{\mathbf{X}}_t - \hat{\mathbf{J}}\hat{\mathbf{A}}\hat{\mathbf{X}} + \hat{\nabla} f(\mathbf{X})\right|^2, \quad \text{s.t.} \quad \hat{\mathbf{A}}^\mathsf{T} = \pm\hat{\mathbf{A}}, \tag{5}$$

which has minimizer $\hat{\mathbf{A}}$ satisfying the desired symmetry. In (5), $\hat{\nabla} f(\mathbf{X}) = \mathbf{U}^\mathsf{T}\nabla f(\mathbf{X})$ denotes the projection of the snapshot data for the derivative of the nonlinear term. Applying Theorem 3.1 with $\mathbf{C} = \hat{\mathbf{X}}_t - \hat{\mathbf{J}}\hat{\nabla} f(\mathbf{X})$ yields the equivalent linear system

$$\left(\hat{\mathbf{J}}^\mathsf{T}\hat{\mathbf{J}}\bar{\oplus}\hat{\mathbf{X}}\hat{\mathbf{X}}^\mathsf{T}\right)\operatorname{vec}\hat{\mathbf{A}} = \operatorname{vec}\left(\hat{\mathbf{J}}^\mathsf{T}\hat{\mathbf{X}}_t\hat{\mathbf{X}}^\mathsf{T} + \hat{\mathbf{X}}\hat{\mathbf{X}}_t^\mathsf{T}\hat{\mathbf{J}} - \hat{\mathbf{J}}^\mathsf{T}\hat{\mathbf{J}}\hat{\nabla}\mathbf{f}(\mathbf{X})\hat{\mathbf{X}}^\mathsf{T} - \hat{\mathbf{X}}\hat{\nabla} f(\mathbf{X})^\mathsf{T}\hat{\mathbf{J}}^\mathsf{T}\hat{\mathbf{J}}\right), \tag{6}$$

which is guaranteed (see Section 4) to yield an operator $\hat{\mathbf{A}}$ which converges to $\mathbf{U}^\mathsf{T}\mathbf{A}\mathbf{U}$ in an appropriate limit. Interestingly, it is even more useful in practice to make the approximation $\hat{\mathbf{J}}^\mathsf{T}\hat{\mathbf{J}} \approx \mathbf{I}$ in (6), which is exact for the cotangent lift algorithm discussed in Section 2.2, yielding the alternative linear system

$$\left(\mathbf{I}\bar{\oplus}\hat{\mathbf{X}}\hat{\mathbf{X}}^\mathsf{T}\right)\operatorname{vec}\hat{\mathbf{A}} = \operatorname{vec}\left(\hat{\mathbf{J}}^\mathsf{T}\hat{\mathbf{X}}_t\hat{\mathbf{X}}^\mathsf{T} + \hat{\mathbf{X}}\hat{\mathbf{X}}_t^\mathsf{T}\hat{\mathbf{J}} - \hat{\nabla} f(\mathbf{X})\hat{\mathbf{X}}^\mathsf{T} - \hat{\mathbf{X}}\hat{\nabla} f(\mathbf{X})^\mathsf{T}\right), \tag{7}$$

which satisfies Proposition 3.1 whenever the POD basis used is drawn from the SVD of $\mathbf{X}$. Inferring $\hat{\mathbf{A}}$ by way of solving (7) will be called the C-H-OpInf procedure, and is summarized in Algorithm 1.

Remark 3.2. Notice that both inference procedures (6) and (7) preserve an approximation to the reduced Hamiltonian $\hat{H}\left(\hat{\mathbf{x}}\right) = \frac{1}{2}\hat{\mathbf{x}}^\mathsf{T}\hat{\mathbf{A}}\hat{\mathbf{x}} + \hat{\mathbf{f}}\left(\hat{\mathbf{x}}\right)$, which can be considered a perturbation of the true $\hat{H}$. The analysis in Section 4, particularly Theorem 4.5, guarantees that this perturbation remains bounded throughout the range of the training data for a high enough snapshot density and large enough basis size, although, in practice, this property seems to hold for much longer time integrations as well (see Section 5).

Algorithm 1 Canonical Hamiltonian Operator Inference (C-H-OpInf)

Input: Snapshots $\mathbf{X} \in \mathbb{R}^{N \times n_s}$ of model solution; snapshots $\nabla\mathbf{f}(\mathbf{X}) \in \mathbb{R}^{N \times n_s}$ of nonlinear term in the gradient ∇H of the Hamiltonian; integer $n > 0$ and real number $\eta > 0$.
Output: Symmetric, reduced operator $\hat{\mathbf{A}} \in \mathbb{R}^{n \times n}$ approximating the linear term in the gradient $\nabla\hat{H}$ of the reduced Hamiltonian.
  1: Employ the user's preferred algorithm to build a reduced basis $\mathbf{U} \in \mathbb{R}^{N \times n}$ from snapshot data.
  2: Form reduced Poisson operator $\hat{\mathbf{J}} = \mathbf{U}^\mathsf{T}\mathbf{J}\mathbf{U} \in \mathbb{R}^{n \times n}$, as well as projected quantities $\hat{\mathbf{X}} = \mathbf{U}^\mathsf{T}\mathbf{X} \in \mathbb{R}^{n \times n_s}$ and $\hat{\nabla} f(\mathbf{X}) = \mathbf{U}^\mathsf{T}\nabla f(\mathbf{X}) \in \mathbb{R}^{n \times n_s}$.
  3: Solve the $n^2 \times n^2$ linear system (7) for $\hat{\mathbf{A}} \in \mathbb{R}^{n \times n}$.

Before moving to the case of noncanonical systems, it is worth discussing how the C-H-OpInf procedure discussed here relates to the previous H-OpInf work in [40] summarized in Section 2.4. Particularly, if $\mathbf{U}$ is

chosen via the cotangent lift algorithm so that $\mathbf{U}^\mathsf{T}\mathbf{J}\mathbf{U} = \mathbf{J}_n$ is the canonical symplectic matrix of dimension n, and $\hat{\mathbf{A}} = \mathrm{diag}\left(\hat{\mathbf{A}}_{\mathbf{qq}}, \hat{\mathbf{A}}_{\mathbf{pp}}\right)$ is assumed to be block diagonal, then the algorithm presented here reduces to [40, Algorithm 1]. This is because the C-H-OpInf problem (7) decouples into a pair of problems for each diagonal block $\hat{\mathbf{A}}_{\mathbf{q}}, \hat{\mathbf{A}}_{\mathbf{p}}$ in $\hat{\mathbf{A}}$, recovering exactly the minimization problems solved by that algorithm. The formulation from [40] has the advantage of requiring the solution to two problems of size $m^2 \times m^2$ (still solvable with Theorem 3.1) as opposed to one problem of size $2m^2 \times 2m^2$, but does not allow any flexibility in the choice of basis $\mathbf{U}$ and cannot accurately represent any systems with a nonseparable Hamiltonian. Therefore, it should only be used when the problem in question is canonical and the continuous operator ∇H is diagonal in phase space. Conversely, the inference described in Algorithm 1 can accommodate any reduced basis, requires the solution of only one linear system, and is applicable to any Hamiltonian system in canonical form.

3.2. Noncanonical Hamiltonian systems

A primary advantage of the OpInf technique inspired by Theorem 3.1 is that it extends to Hamiltonian systems in noncanonical form. To see this, suppose snapshots of a potentially unknown Hamiltonian system are collected in an $(N \times n_s)$-matrix $\mathbf{X}$, and that a candidate Hamiltonian function H has been identified. This may occur if, for example, a conserved quantity has been identified but the corresponding Hamiltonian structure remains unknown. Then, an analytic expression for ∇H can be obtained, and hence it is possible to compute the matrix $\nabla H(\mathbf{X}) \in \mathbb{R}^{N \times n_s}$ of gradients at the snapshot data $\mathbf{X}$, as well as a finite difference approximation $\mathbf{X}_t \approx \dot{\mathbf{X}}$. As before, this enables the construction of a POD basis $\mathbf{U} \in \mathbb{R}^{N \times n}$ via the SVD of the mean-centered data matrix $\mathbf{Y} = \mathbf{X} - \mathbf{X}_0$, where $\mathbf{X}_0$ denotes the matrix each column of which is the initial state $\mathbf{x}_0$. Writing the Galerkin approximation $\tilde{\mathbf{x}} \approx \mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}}$ again yields the prototypical H-ROM (see Section 2.2) $\dot{\mathbf{x}} = \hat{\mathbf{L}}\nabla \hat{H}$ where $\hat{\mathbf{L}} = \mathbf{U}^\mathsf{T}\mathbf{L}\mathbf{U}$. When $\mathbf{L}$ is inaccessible, this suggests a similar inference procedure based on Theorem 3.1 which preserves the antisymmetry necessary for Hamiltonian preservation. Particularly, it is possible to form the $n \times n_s$ reduced quantities

$$\hat{\mathbf{X}} = \mathbf{U}^\mathsf{T}\mathbf{X}, \qquad \hat{\mathbf{X}}_t = \mathbf{U}^\mathsf{T}\mathbf{X}_t, \qquad \hat{\nabla} H(\mathbf{X}) = \mathbf{U}^\mathsf{T}\nabla H(\mathbf{X}),$$

and solve the optimization problem

$$\underset{\hat{\mathbf{L}} \in \mathbb{R}^{n \times n}}{\mathrm{argmin}} \left|\hat{\mathbf{X}}_t - \hat{\mathbf{L}}\hat{\nabla} H(\mathbf{X})\right|^2, \quad \text{s.t.} \quad \hat{\mathbf{L}}^\mathsf{T} = -\hat{\mathbf{L}},$$

which is a straightforward least-squares inference for the antisymmetric $\hat{\mathbf{L}}$. As shown in Theorem 3.1, this is equivalent to solving the $n^2 \times n^2$ linear system

$$\left(\mathbf{I} \oplus \hat{\nabla} H(\mathbf{X})\hat{\nabla} H(\mathbf{X})^\mathsf{T}\right) \mathrm{vec}\,\hat{\mathbf{L}} = \mathrm{vec}\left(\hat{\mathbf{X}}_t \hat{\nabla} H(\mathbf{X})^\mathsf{T} - \hat{\nabla} H(\mathbf{X})\hat{\mathbf{X}}_t^\mathsf{T}\right). \tag{8}$$

Inferring $\hat{\mathbf{L}}$ based on solving (8) will be called the NC-H-OpInf method, and is summarized in Algorithm 2. While this inference similarly does not satisfy the hypotheses of Proposition 3.1, it is interesting to note that "one shot" computation of $\hat{\mathbf{L}}$ using Algorithm 2 occasionally works quite well in practice when the basis $\mathbf{U}$ is chosen from the SVD of $\mathbf{Y}$ (see Section 5).

Algorithm 2 Noncanonical Hamiltonian Operator Inference (NC-H-OpInf)

Input: Snapshots $\mathbf{X} \in \mathbb{R}^{N \times n_s}$ of model solution; snapshots $\nabla H(\mathbf{X}) \in \mathbb{R}^{N \times n_s}$ of the gradient ∇H of the Hamiltonian; integer $n > 0$ and real number $\eta > 0$.
Output: Antisymmetric, reduced operator $\hat{\mathbf{L}} \in \mathbb{R}^{n \times n}$ approximating the Poisson operator governing the H-ROM $\dot{\hat{\mathbf{x}}} = \hat{\mathbf{L}}\nabla \hat{H}$.
  1: Employ the user's preferred algorithm to build a (mean-centered) POD basis $\mathbf{U} \in \mathbb{R}^{N \times n}$ from snapshot data.
  2: Form projected quantities $\hat{\mathbf{X}} = \mathbf{U}^\mathsf{T}\mathbf{X} \in \mathbb{R}^{n \times n_s}$ and $\hat{\nabla} H(\mathbf{X}) = \mathbf{U}^\mathsf{T}\nabla H(\mathbf{X}) \in \mathbb{R}^{n \times n_s}$.
  3: Solve the $n^2 \times n^2$ linear system (8) for $\hat{\mathbf{L}} \in \mathbb{R}^{n \times n}$.

Remark 3.3. Note that NC-H-OpInf can be used (along with a symplectic time integrator) to obtain dynamics which preserve any quantity H, regardless of whether or not it corresponds to a true Hamiltonian structure. In this way, it can be considered a gray-box method requiring only snapshots and a desired conserved quantity.

4. Analysis

Now that the C-H-OpInf and NC-H-OpInf procedures have been described, it is important to validate that the inferred operators approximate their intrusive counterparts in an appropriate sense. To accomplish this, the following mild assumptions are needed.

Assumption 4.1. The span of the POD basis $\mathbf{U} \in \mathbb{R}^{N \times n}$ tends to $\mathbb{R}^N$ as $n \to N$, i.e., for any $\mathbf{x} \in \mathbb{R}^N$,

$$\lim_{n \to N} \left| \mathbf{P}^\perp \mathbf{x} \right| = 0,$$

where $\mathbf{P}^\perp := \mathbf{I} - \mathbf{U}\mathbf{U}^\mathsf{T}$.

Assumption 4.2. The approximate time derivatives $\mathbf{x}_t$ converge to the true derivatives $\dot{\mathbf{x}}$ as the time step $\Delta t \to 0$, i.e.,

$$\lim_{\Delta t \to 0} \max_i |\mathbf{x}_t(t_i) - \dot{\mathbf{x}}(t_i)| = 0.$$

Assumption 4.3. The snapshot matrices $\mathbf{X}, \nabla H(\mathbf{X}) \in \mathbb{R}^{N \times n_s}$ have maximal rank.

This allows for the following result regarding the convergence of NC-H-OpInf.

Theorem 4.4. *Under Assumptions 4.1, 4.2, and 4.3, the inferred operator $\hat{\mathbf{L}}$ from the NC-H-OpInf procedure in Algorithm 2 converges to the intrusive operator $\bar{\mathbf{L}} = \mathbf{U}^\mathsf{T}\mathbf{L}\mathbf{U}$ as $\Delta t \to 0$ and $n \to N$.*

Proof. First, notice that

$$\left| \hat{\mathbf{X}}_t - \hat{\mathbf{L}}\hat{\nabla}H(\mathbf{X}) \right| = \left| \left(\hat{\mathbf{X}}_t - \dot{\hat{\mathbf{X}}} \right) + \left(\dot{\hat{\mathbf{X}}} - \bar{\mathbf{L}}\hat{\nabla}H(\mathbf{X}) \right) + \left(\bar{\mathbf{L}} - \hat{\mathbf{L}} \right)\hat{\nabla}H(\mathbf{X}) \right|$$

$$= \left| \mathbf{U}^\mathsf{T}\left(\mathbf{X}_t - \dot{\mathbf{X}} \right) + \mathbf{U}^\mathsf{T}\left(\dot{\mathbf{X}} - \mathbf{L}\nabla H(\mathbf{X}) \right) + \mathbf{U}^\mathsf{T}\mathbf{L}\mathbf{P}^\perp\nabla H(\mathbf{X}) + \left(\bar{\mathbf{L}} - \hat{\mathbf{L}} \right)\hat{\nabla}H(\mathbf{X}) \right|$$

$$= \left| \mathbf{U}^\mathsf{T}\left(\mathbf{X}_t - \dot{\mathbf{X}} \right) + \mathbf{U}^\mathsf{T}\mathbf{L}\mathbf{P}^\perp\nabla H(\mathbf{X}) + \left(\bar{\mathbf{L}} - \hat{\mathbf{L}} \right)\hat{\nabla}H(\mathbf{X}) \right|$$

$$\leq |\mathbf{U}|\left(\left| \mathbf{X}_t - \dot{\mathbf{X}} \right| + |\mathbf{L}|\left| \mathbf{P}^\perp\nabla H(\mathbf{X}) \right| \right) + \left| \bar{\mathbf{L}} - \hat{\mathbf{L}} \right|\left| \hat{\nabla}H(\mathbf{X}) \right|.$$

Therefore, for each $n \leq N$,

$$\min_{\hat{\mathbf{L}}} \left| \hat{\mathbf{X}}_t - \hat{\mathbf{L}}\hat{\nabla}H(\mathbf{X}) \right| \leq \min_{\hat{\mathbf{L}}} \left[|\mathbf{U}|\left(\left| \mathbf{X}_t - \dot{\mathbf{X}} \right| + |\mathbf{L}|\left| \mathbf{P}^\perp\nabla H(\mathbf{X}) \right| \right) + \left| \bar{\mathbf{L}} - \hat{\mathbf{L}} \right||\nabla H(\mathbf{X})| \right]$$

$$= |\mathbf{U}|\left(\left| \mathbf{X}_t - \dot{\mathbf{X}} \right| + |\mathbf{L}|\left| \mathbf{P}^\perp\nabla H(\mathbf{X}) \right| \right).$$

By Assumptions 4.1 and 4.2, for any $\varepsilon > 0$ there exists an $n' < N$ and $\Delta t' > 0$ such that

$$\left| \mathbf{U}^\mathsf{T}\left(\mathbf{X}_t - \dot{\mathbf{X}} \right) + \mathbf{U}^\mathsf{T}\mathbf{L}\mathbf{P}^\perp\nabla H(\mathbf{X}) \right| \leq |\mathbf{U}|\left(\left| \mathbf{X}_t - \dot{\mathbf{X}} \right| + |\mathbf{L}|\left| \mathbf{P}^\perp\nabla H(\mathbf{X}) \right| \right) < \frac{\varepsilon}{2}.$$

Therefore, for $n \geq n'$ and $\Delta t \leq \Delta t'$ it follows from an elementary calculation that

$$\min_{\hat{\mathbf{L}}} \left| \left(\bar{\mathbf{L}} - \hat{\mathbf{L}} \right)\hat{\nabla}H(\mathbf{X}) \right| < \varepsilon,$$

from which it can be concluded that $\hat{\mathbf{L}} \to \bar{\mathbf{L}}$, since $\hat{\nabla}H(\mathbf{X})$ has maximal rank. $\square$

A similar result holds for C-H-OpInf provided a cotangent lift basis $\mathbf{U}$ is used.

Theorem 4.5. *Under Assumptions 4.1, 4.2, 4.3, and using a cotangent lift POD basis $\mathbf{U}$, the inferred operator $\hat{\mathbf{A}}$ from the C-H-OpInf procedure in Algorithm 1 converges to the intrusive operator $\bar{\mathbf{A}} = \mathbf{U}^\mathsf{T}\mathbf{A}\mathbf{U}$ as $\Delta t \to 0$ and $n \to N$.*

Remark 4.6. Note that the assumption of a cotangent lift basis in Theorem 4.5 can be dropped provided (6) is solved instead of (7) in the C-H-OpInf Algorithm 1.

Proof. First, notice that

$$\dot{\mathbf{X}} - \mathbf{J}\mathbf{U}\mathbf{U}^\mathsf{T}\left(\mathbf{A}\mathbf{U}^\mathsf{T}\mathbf{X} + \nabla f(\mathbf{X})\right)$$
$$= \left(\dot{\mathbf{X}} - \mathbf{J}\left(\mathbf{A}\mathbf{X} + \nabla f(\mathbf{X})\right)\right) + \mathbf{J}\left(\mathbf{P}^\perp\mathbf{A}\mathbf{X} + \mathbf{P}^\perp\nabla f(\mathbf{X}) + \mathbf{U}\mathbf{U}^\mathsf{T}\mathbf{A}\mathbf{P}^\perp\mathbf{X}\right)$$
$$= \mathbf{J}\left(\mathbf{P}^\perp\mathbf{A}\mathbf{X} + \mathbf{P}^\perp\nabla f(\mathbf{X}) + \mathbf{U}\mathbf{U}^\mathsf{T}\mathbf{A}\mathbf{P}^\perp\mathbf{X}\right),$$

since $\mathbf{J}\nabla H(\mathbf{X}) = \mathbf{J}\left(\mathbf{A}\mathbf{X} + \nabla f(\mathbf{X})\right)$. Therefore, it follows as before that for every $n \leq N$,

$$\left|\hat{\mathbf{X}}_t - \hat{\mathbf{J}}\left(\hat{\mathbf{A}}\hat{\mathbf{X}} + \hat{\nabla} f(\mathbf{X})\right)\right| = \left|\left(\hat{\mathbf{X}}_t - \dot{\mathbf{X}}\right) + \left(\dot{\mathbf{X}} - \hat{\mathbf{J}}\left(\bar{\mathbf{A}}\hat{\mathbf{X}} + \hat{\nabla} f(\mathbf{X})\right)\right) + \hat{\mathbf{J}}\left(\bar{\mathbf{A}} - \hat{\mathbf{A}}\right)\hat{\mathbf{X}}\right|$$
$$= \left|\mathbf{U}^\mathsf{T}\left(\mathbf{X}_t - \dot{\mathbf{X}}\right) + \mathbf{U}^\mathsf{T}\mathbf{J}\left(\mathbf{P}^\perp\mathbf{A}\mathbf{X} + \mathbf{P}^\perp\nabla f(\mathbf{X}) + \mathbf{U}\mathbf{U}^\mathsf{T}\mathbf{A}\mathbf{P}^\perp\mathbf{X}\right) + \hat{\mathbf{J}}\left(\bar{\mathbf{A}} - \hat{\mathbf{A}}\right)\hat{\mathbf{X}}\right|,$$

Now, for any $\varepsilon > 0$ we can choose $n' < N$ and $\Delta t' > 0$ so that

$$\left|\mathbf{U}^\mathsf{T}\left(\mathbf{X}_t - \dot{\mathbf{X}}\right) + \mathbf{U}^\mathsf{T}\mathbf{J}\left(\mathbf{P}^\perp\mathbf{A}\mathbf{X} + \mathbf{P}^\perp\nabla f(\mathbf{X}) + \mathbf{U}\mathbf{U}^\mathsf{T}\mathbf{A}\mathbf{P}^\perp\mathbf{X}\right)\right|$$
$$\leq |\mathbf{U}|\left(\left|\mathbf{X}_t - \dot{\mathbf{X}}\right| + \left|\mathbf{P}^\perp\right||\mathbf{J}|\left(|\mathbf{A}||\mathbf{X}| + |\nabla f(\mathbf{X})| + |\mathbf{U}|^2|\mathbf{A}||\mathbf{X}|\right)\right) < \frac{\varepsilon}{2},$$

and therefore we have

$$\min_{\hat{\mathbf{A}}}\left|\hat{\mathbf{J}}\left(\bar{\mathbf{A}} - \hat{\mathbf{A}}\right)\hat{\mathbf{X}}\right| < \varepsilon,$$

provided $n \geq n'$ and $\Delta t < \Delta t'$. Hence, $\hat{\mathbf{A}} \to \bar{\mathbf{A}}$ as desired, since $\hat{\mathbf{J}}, \hat{\mathbf{X}}$ have maximal rank. $\quad\square$

Remark 4.7. While useful, the results of this section only hold in the "infinite data limit", and so cannot guarantee good performance of the OpInf methods (and projection-based ROMs in general) in all cases of practical interest, particularly in the predictive regime. It is an ongoing effort to develop rigorous estimates which are more valuable in the presence of partial or limited data.

5. Numerical examples

Here, numerical results are reported on several benchmark problems from hydrodynamics and linear elasticity, including a linear wave equation, a manufactured test case which has a non-separable canonical Hamiltonian form, the Korteweg–de Vries equation, the Benjamin–Bona–Mahony equation, and a 3D linear elastic clamped plate problem undergoing high-frequency oscillations. The primary error metrics used for comparison will be relative ℓ_2 error in the state approximation,

$$R\ell_2\left(\mathbf{X}, \tilde{\mathbf{X}}\right) = \frac{\left|\mathbf{X} - \tilde{\mathbf{X}}\right|_2}{|\mathbf{X}|_2},$$

as well as signed error in the Hamiltonian (or other conserved quantity) approximation $H(\mathbf{x}(t)) - H_0$ where $H_0 = H(\mathbf{x}(0))$. When speaking about the properties of POD bases, it will also be useful to evaluate the snapshot energy, computed for a given rank r snapshot matrix $\mathbf{X}$ with singular values $\{\sigma_i\}_{i=1}^r$ and POD basis size $n \leq r$ as

$$E_s\left(\mathbf{X}, n\right) = \frac{\sum_{k=1}^n \sigma_k}{\sum_{k=1}^r \sigma_k}.$$

Note that, when appropriate, both uncentered ($\tilde{\mathbf{x}} = \mathbf{U}\hat{\mathbf{x}}$) and mean-centered ($\tilde{\mathbf{x}} = \mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}}$) Galerkin projections will be considered. This will be denoted by the letters "MC" in the figures below. Of course, mean-centering requires a POD of the centered snapshot matrix discussed in Section 2.2, and is infeasible for a general OpInf method. On the other hand, NC-H-OpInf is amenable to this technique, since the inferred operator $\hat{\mathbf{L}}$ does not interface directly with the approximate solution $\tilde{\mathbf{x}}$.

When evaluating the performance of the H-OpInf methods in Section 3, comparisons are drawn with the standard intrusive Galerkin ROM (G-ROM) and Hamiltonian ROM (H-ROM) discussed previously, as well as the standard Galerkin OpInf (G-OpInf) when appropriate. Reproductive as well as predictive problems are considered, encompassing both prediction in time as well as prediction across parameter space. Note that all ROMs considered are equally efficient online; since the chosen examples have polynomial nonlinearities, their resulting ROMs do not depend on the full-order state space N, instead scaling only with the reduced basis size n.

Remark 5.1. On canonical Hamiltonian examples, the NC-H-OpInf algorithm will infer only $\hat{\mathbf{J}} \approx \mathbf{U}^\mathsf{T}\mathbf{J}\mathbf{U}$, which is already known. Since it is instructive to see that the NC-H-OpInf ROM behaves appropriately on these examples, comparisons including it are presented for these cases, although it should be noted that this is not the intended purpose of NC-H-OpInf.

5.1. Linear wave equation

First, consider the one-dimensional linear wave equation with constant speed c,

$$\begin{aligned}
\varphi_{tt} &= c^2 \varphi_{ss}, & 0 \le s \le l, \\
\varphi(0) &= h(y(s)), & \varphi_t(0) = 0,
\end{aligned} \tag{9}$$

where the boundary conditions are periodic and the (parameterized) initial condition is a cubic spline defined by

$$h(y) = \begin{cases} 1 - \frac{3}{2}y^2 + \frac{3}{4}y^3 & 0 \le y \le 1, \\ \frac{1}{4}(2-y)^3 & 1 < y \le 2, \\ 0 & y > 2, \end{cases} \qquad y(s,\alpha) = \alpha \left| s - \frac{1}{2} \right|.$$

Letting $\mathbf{x} = \begin{pmatrix} q & p \end{pmatrix}^\mathsf{T} \in \mathbb{R}^2$ where $q = \varphi$ and $p = \varphi_t$, this problem is readily recast in the canonical Hamiltonian form

$$\dot{\mathbf{x}} = \mathbf{J}\nabla H(\mathbf{x}) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} -c^2 \partial_{ss} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix},$$

where the Hamiltonian functional is given by

$$H(\mathbf{x}) = \frac{1}{2} \int_0^l \left(p^2 + c^2 q_s^2 \right) ds,$$

and it follows quickly from differentiation that $H_q = -c^2 q_{ss}$, $H_p = p$. As discussed in Section 2.5, semi-discretizing in $\mathbf{x}$ and applying AVF integration to this system yields the implicit midpoint rule

$$\frac{\mathbf{x}^{k+1} - \mathbf{x}^k}{\Delta t} = \mathbf{J}\mathbf{A} \left(\frac{\mathbf{x}^{k+1} + \mathbf{x}^k}{2} \right) = \begin{pmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{pmatrix} \begin{pmatrix} -c^2 \mathbf{D}_2 & \\ 0 & \mathbf{I} \end{pmatrix} \left(\frac{\mathbf{x}^{k+1} + \mathbf{x}^k}{2} \right),$$

where $\mathbf{x} = \begin{pmatrix} \mathbf{q} & \mathbf{p} \end{pmatrix}^\mathsf{T}$ has been overloaded, $\mathbf{D}_2$ denotes the circulant matrix which results from using a three-point stencil finite difference method to discretize the 1-D Laplace operator, and the discrete Hamiltonian (also overloaded as H) is given by

$$H(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{N/2} \left(p_i^2 + c^2 \frac{(q_{i+1} - q_i)^2 + (q_i - q_{i-1})^2}{4\Delta x^2} \right).$$

Note that the AVF method will preserve this discrete Hamiltonian exactly by construction. Some snapshots of this solution for different values of α are displayed in Fig. 1.

To evaluate the performance of the ROMs discussed thus far, two experiments will be conducted: one testing prediction in time, and one testing prediction in parameter space. For each, the wave speed is fixed to $c = 0.1$, the length to $l = 1$, and the spatial domain is divided into $M = 500$ equally sized intervals (yielding a state vector $\mathbf{x}$ of dimension $N = 2M = 1000$).

5.1.1. Reproductive versus predictive dynamics

The first goal is to compare the C-H-Opinf and NC-H-Opinf ROM methods discussed in Section 3 to their intrusive counterparts when predicting trajectories outside the temporal range of their training data. For this, a total of 501 snapshots of the FOM solution with initial condition parameter $\alpha = 5$ are uniformly collected on the time interval $[0, T]$ where $T = 10$. These data are used to train three POD bases: one constructed in the "ordinary way" by forming the SVD of a data matrix of size $N \times n_t$ containing snapshots of $\mathbf{x}$, another constructed using the cotangent lift algorithm described in Section 2.2, and the final constructed block-wise using the SVD of snapshot data for position $\mathbf{q}$ and momentum $\mathbf{p}$ separately (also described in Section 2.2). The snapshot energies and projection errors associated to these bases are shown in Fig. 2. It is evident that all bases are capable of capturing roughly 99%
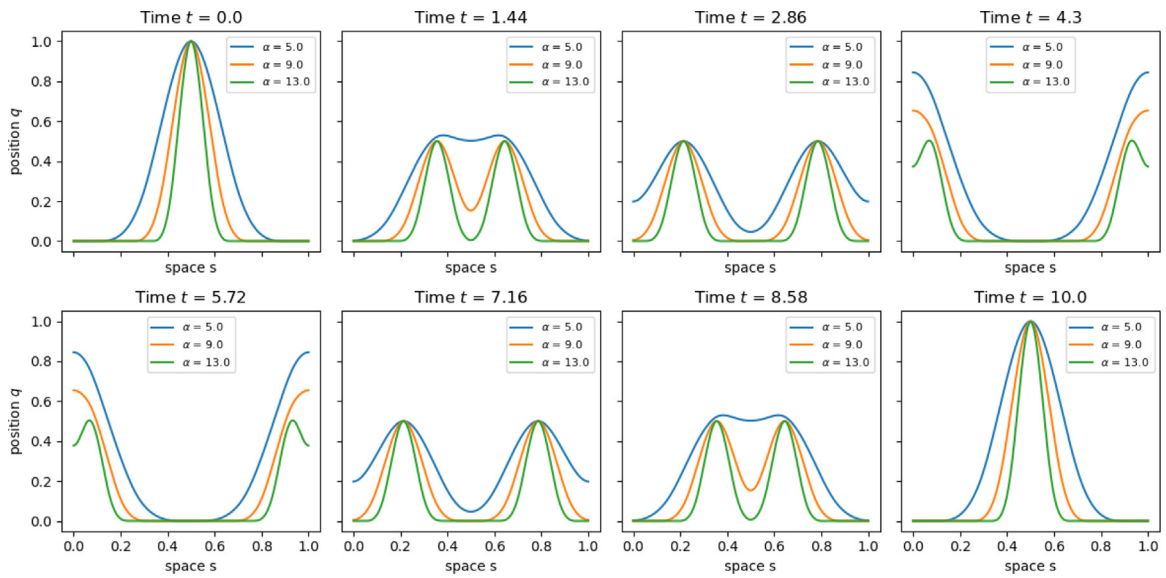
Fig. 1. Solution snapshots from the linear wave example for different values of the parameter α.
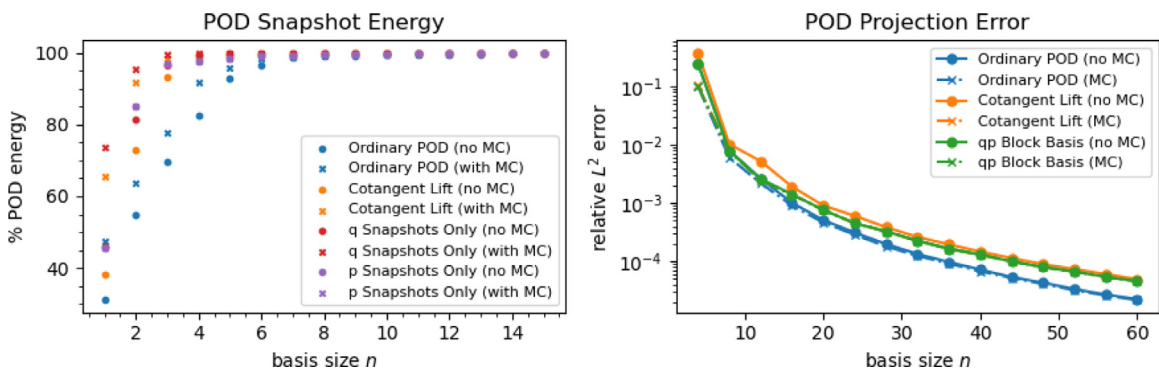

Fig. 2. POD snapshot energies (left) and projection errors (right) corresponding to the bases used in the nonparametric ($\alpha = 5$) linear wave example. "MC" indicates mean-centering of the snapshots was performed.

of the snapshot energy with only $n = 15$ modes, despite exhibiting a slowly decaying projection error characteristic of hyperbolic problems.

Fig. 3 plots the relative ROM errors as a function of basis size in the case where the ROMs are integrated only in the range of the training data, i.e. $t \in [0, 10]$. Notice that both the intrusive G-ROM and the G-OpInf ROM are less accurate than their Hamiltonian counterparts, and that the G-OpInf ROM is somewhat unstable with the addition of basis modes. It is further interesting to observe the differences in performance between the ROM algorithms as the underlying basis is changed. Particularly, both the cotangent lift and (q, p)-block basis lead to lower relative errors than ordinary POD, although ordinary POD has the significant (empirical) advantage of stability under OpInf truncation. Indeed, in the case of the ordinary POD basis, all operators used in the OpInf ROMs were computed in "one shot" via truncation from the operators learned at the largest basis size. While this is not guaranteed to be optimal according to Proposition 3.1, it is interesting to note that this resulted in almost no degradation of performance. This contrasts highly with the case of the cotangent lift and (q, p) block bases, for which OpInf truncation led to unusable results (not pictured).

To show the effect of each ROM on energy preservation, Fig. 4 uses the block (q, p) basis case with $n = 16$ modes to show the change in the Hamiltonian H over time. From this, it is seen that the intrusive H-ROM and NC-H-OpInf ROM conserve energy exactly, while the C-H-OpInf ROM conserves energy to order 10^{-8}. Of course,
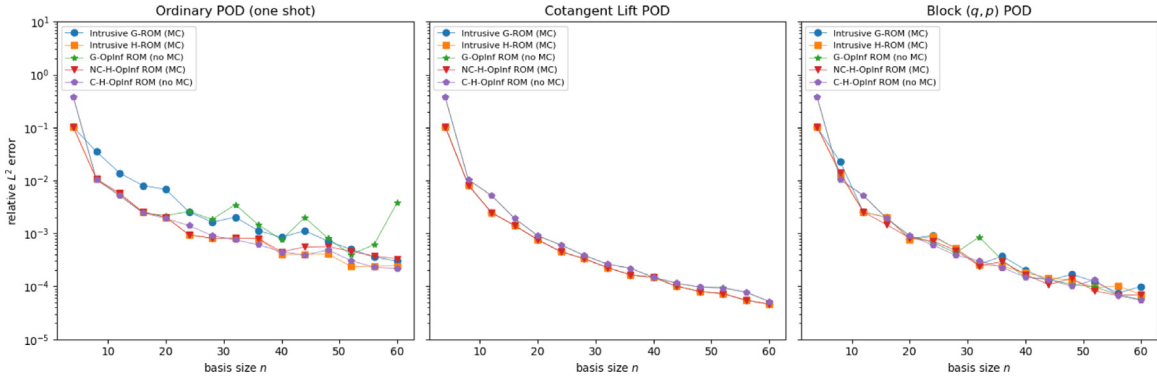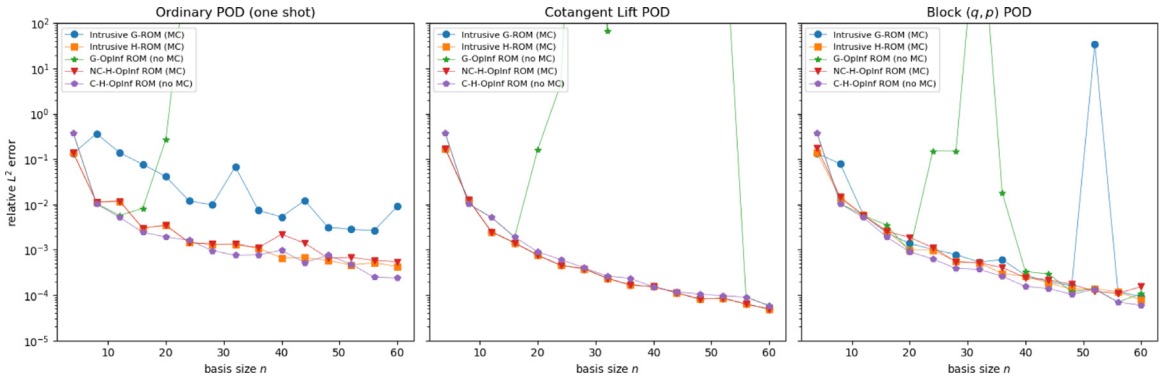
Fig. 3. Relative state errors as a function of basis modes for the ROMs in the linear wave example (reproductive case $T = 10$). "MC" indicates the use of a mean-centered reconstruction.
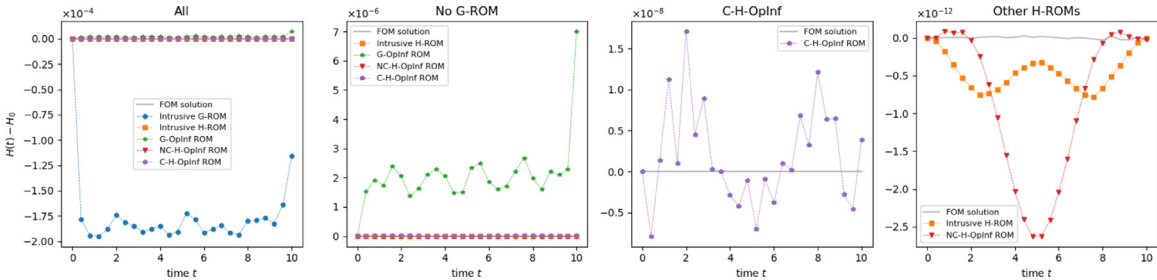
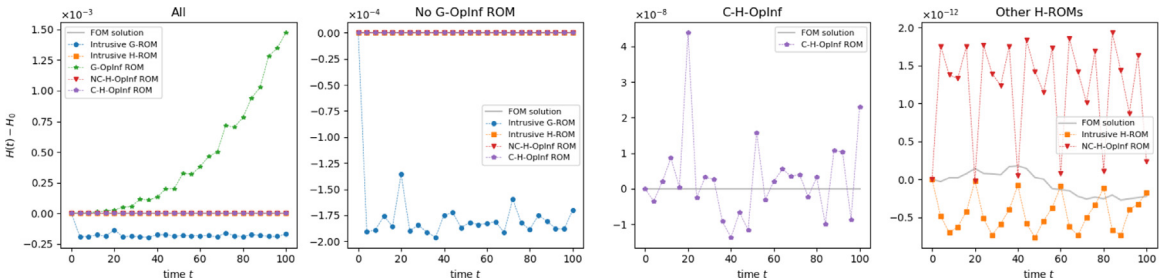**Fig. 4.** ROM energy errors for the linear wave example in the reproductive case ($T = 10$) when using a block $(\mathbf{q}, \mathbf{p})$ POD basis with mean-centering (where applicable) and with $n = 16$ modes.
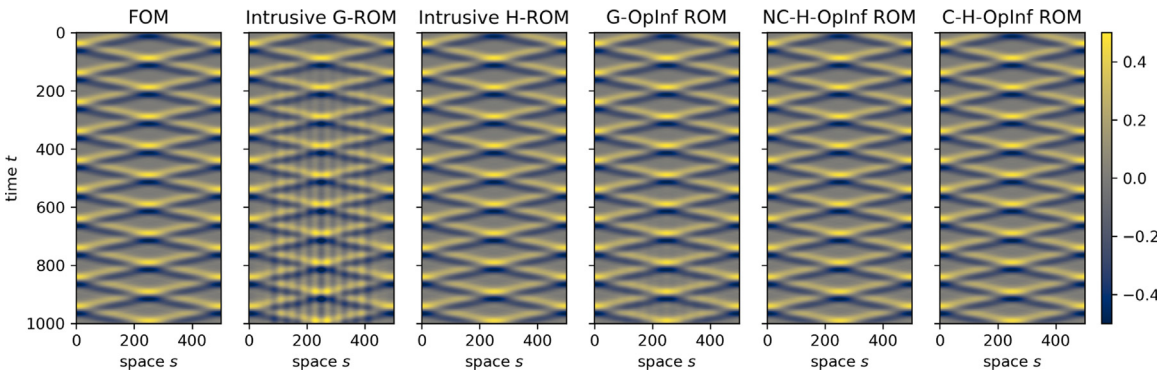
Fig. 5. Plots of the FOM and ROM solutions to the linear wave equation in the predictive case ($T = 100$) when using a standard POD basis with $n = 16$ modes. Note that mean-centered reconstructions were used for all but the G-OpInf and C-H-OpInf ROMs.

this is a consequence of the fact that the matrix $\hat{\mathbf{A}}$ learned by C-H-OpInf represents only an approximation to the gradient of the true reduced Hamiltonian $\hat{H}$. On the other hand, note that C-H-OpInf still conserves H much better than G-OpInf or the intrusive G-ROM, and is guaranteed to exactly preserve the approximate reduced energy $\tilde{H} = \frac{1}{2}\hat{\mathbf{x}}^{\mathsf{T}}\hat{\mathbf{A}}\hat{\mathbf{x}}$ (not pictured), which follows since the matrix $\hat{\mathbf{J}} = \mathbf{U}^{\mathsf{T}}\mathbf{J}\mathbf{U}$ is skew-symmetric. It is further remarkable that the conservation properties of the H-ROMs displayed in these plots do not depend on the basis construction mechanism or the number of basis modes, n.

Moving beyond the reproductive case, it is useful to see what happens when the ROMs are tested on an interval of integration which is much larger. Fig. 6 plots the relative ROM errors as a function of basis modes when the ROMs are tested over an interval of $[0, T]$ with $T = 100$, which is ten times the interval of training. Here the instabilities

Fig. 6. Relative state errors as a function of basis modes for the ROMs in the linear wave equation example (predictive case $T = 100$). "MC" indicates the use of a mean-centered reconstruction.

Fig. 7. ROM energy errors for the linear wave equation example in the predictive case ($T = 100$) when using a block ($\mathbf{q}, \mathbf{p}$) POD basis with mean-centering (where applicable) and with $n = 16$ modes.

in the G-OpInf ROM are made readily apparent, as certain numbers of modes lead extreme blow-ups regardless of the underlying basis construction. It is interesting to note that the intrusive G-ROM also exhibits similar blow-up in the cases (not pictured here) when the POD basis is constructed with ordinary POD and no mean-centering is applied. Conversely, both the intrusive and OpInf H-ROMs exhibit a steady and predictable decrease in error with the addition of basis modes. Note that a comparative visualization of the FOM and ROM solutions is shown in Fig. 5, which plots each solution when an ordinary POD basis is used with $n = 16$ modes.

Fig. 7 displays the variation in the value of the Hamiltonian over this larger integration range when the ROMs are computed using a (q, p) block basis of $n = 16$ modes. As before, the intrusive G-ROM and OpInf G-ROM are not sufficiently conservative, which has consequences for their accuracy and stability. Conversely, the NC-H-OpInf ROM conserves H on the same order as the intrusive H-ROM, and the C-H-OpInf ROM conserves H to order 10^{-8}, exhibiting similar performance to integration over the training interval.

5.1.2. Parametric case

In addition to prediction in time, it is also useful to consider applying ROMs for the prediction of solutions across the parameter space spanned by $\alpha \in \mathbb{R}$, which controls the initial state of the wave (c.f. Fig. 10). To that end, the next experiment examines how well the present ROM methods are able to predict solutions with variable initial conditions. To accomplish this, eleven uniformly distributed parameters $\alpha \in [5, 15]$ are chosen for training, and 501 snapshots of the FOM solution in the range $[0, 10]$ are collected using each parameter instance. These data are then concatenated to form the snapshot matrix which is used to train the POD decompositions. The snapshot energies and projection errors associated to this procedure are shown in Fig. 8, where it is evident that the inclusion of multiple solution trajectories slows down both the increase in the snapshot energy and the decay of the projection error.

For testing, six uniformly distributed parameters $\alpha \in [5.5, 14.5]$ are chosen (note that these are disjoint from the training parameters), and snapshot data of each solution in the temporal range $t \in [0, 100]$ is collected for
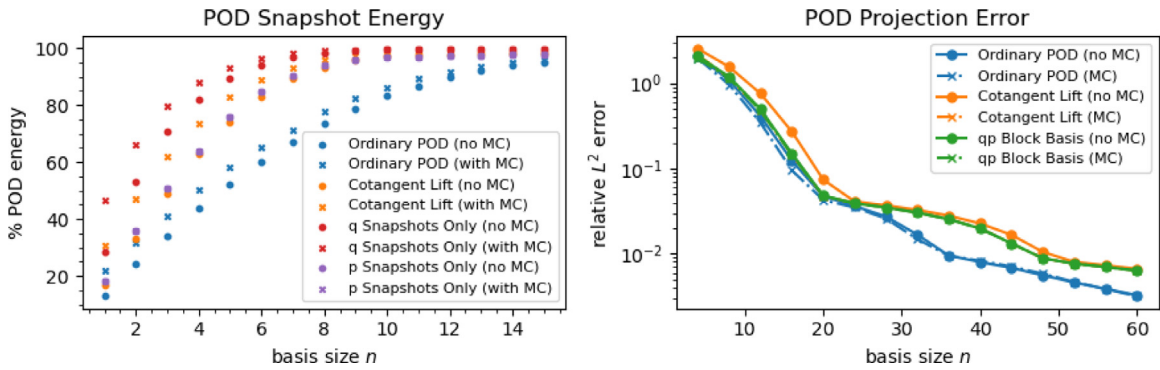
Fig. 8. POD snapshot energies (left) and projection errors (right) corresponding to the bases used in the parametric linear wave example. "MC" indicates that mean-centering of the snapshots was performed.
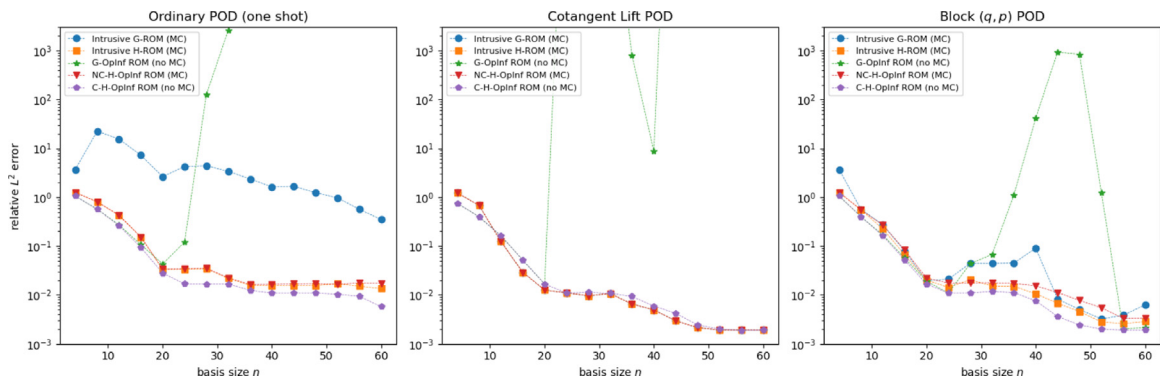

Fig. 9. Relative state errors as a function of basis modes for the ROMs in the linear wave equation example (parametric predictive case $T = 100$). "MC" indicates the use of a mean-centered reconstruction.

comparison with the ROM integration. The ROMs are then tested over this interval beginning from each unseen initial condition, and the average relative error over all test snapshots is reported.

Fig. 9 illustrates the results of this experiment. As in the purely predictive case, we see that the G-OpInf ROM is highly sensitive to basis size, while the intrusive H-ROM and H-OpInf ROMs exhibit a predictable increase in accuracy with the addition of basis modes. Moreover, the intrusive G-ROM is significantly less accurate in the case of an ordinary POD basis, and indeed blows up similarly to the G-OpInf ROM in the case (not pictured) that mean-centering is not applied. A consequence of this is illustrated in Fig. 10, which shows the FOM and ROM solutions in the case that $\alpha = 9.1$ and the ROMs are computed using an ordinary POD basis with mean-centering and with $n = 28$ modes. Notice that the G-OpInf ROM becomes increasingly unstable while the others remain bounded and close to the FOM solution throughout the range of integration.

Finally, it is illustrative to observe the energy plots in Fig. 11, computed using a block (q, p) basis with mean-centering and with $n = 16$ modes. Here it is obvious that the improved conservation properties of the intrusive and OpInf H-ROMs persist in this setting as well, leading to improved accuracy and stability over time when compared to the G-ROMs which do not have this property.

5.2. A non-separable canonical example

Since the linear wave equation can be similarly handled with the techniques in [40], it is worth considering a simple canonical example where the C-H-OpInf method is necessary. Consider the following Hamiltonian and its
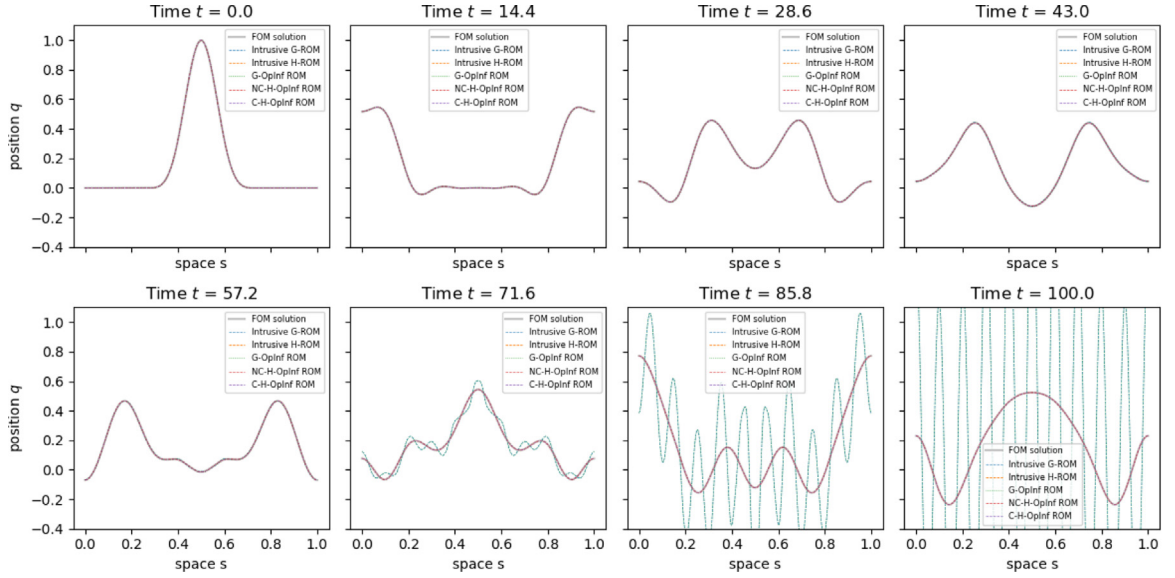
Fig. 10. FOM and ROM solutions for the linear wave equation example in the parametric predictive case ($\alpha = 9.1$, $T = 100$) when using an ordinary POD basis without mean-centering and with $n = 28$ modes.
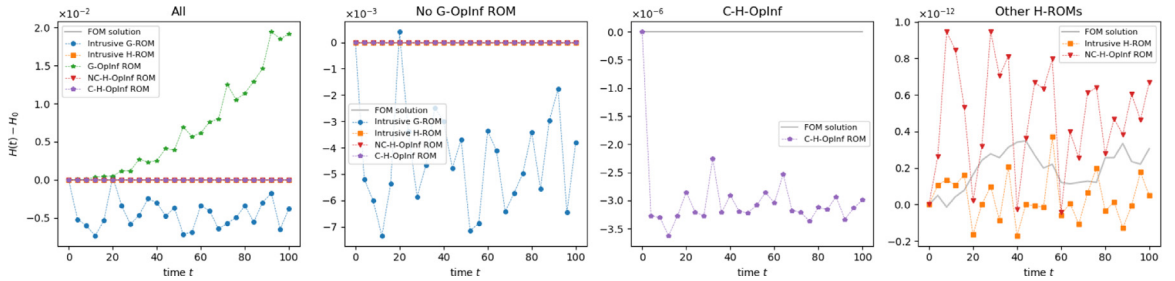

Fig. 11. ROM energy errors for the linear wave equation example in the parametric predictive case ($\alpha = 9.1$, $T = 100$) when using a block (q, p) POD basis with mean-centering (where applicable) and with $n = 16$ modes.

(overloaded) discrete counterpart

$$H(q, p) = \int_0^l \left(1 + q(s)p(s)\right) ds, \qquad H(\mathbf{x}) = \sum_{i=1}^{N/2} \left(1 + q_i p_i\right),$$

where $\mathbf{x} = \begin{pmatrix} \mathbf{q} & \mathbf{p} \end{pmatrix}^\mathsf{T}$ is the semidiscrete state variable of dimension $N = 2M$. Applying the implicit midpoint method to the canonical Hamiltonian system $\dot{\mathbf{x}} = \mathbf{J}\nabla H(\mathbf{x})$ generates the canonical dynamics

$$\frac{\mathbf{x}^{k+1} - \mathbf{x}^k}{\Delta t} = \mathbf{JA}\left(\frac{\mathbf{x}^{k+1} + \mathbf{x}^k}{2}\right) = \begin{pmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{pmatrix}\begin{pmatrix} 0 & \mathbf{I} \\ \mathbf{I} & 0 \end{pmatrix}\left(\frac{\mathbf{x}^{k+1} + \mathbf{x}^k}{2}\right).$$

Clearly, $\nabla H(\mathbf{x}) = \mathbf{Ax}$ does not satisfy the separability hypothesis of [40, Algorithm 1], and therefore that method should not be effective at learning this system. Conversely, the C-H-OpInf Algorithm 1 applies regardless of the separability of H, so it is expected that this system can still be learned through this approach. To see that this is the case, a parameterized initial condition is considered with dimension $M = 500$,

$$\mathbf{x}_0(\alpha) = \begin{pmatrix} \mathbf{q}_0 & \mathbf{p}_0 \end{pmatrix}^\mathsf{T} = \begin{pmatrix} e^{-\alpha(\mathbf{q}+1)}\sin(\alpha\mathbf{q}) & \mathbf{p} \end{pmatrix}^\mathsf{T},$$

and, as before, eleven uniformly distributed parameters $\alpha \in [5, 15]$ are chosen for training, and 501 snapshots of the FOM solution in the range $[0, 2]$ are collected using each parameter instance. The resulting POD projection errors corresponding to these data are shown in Fig. 13, along with some solution snapshots in Fig. 12.
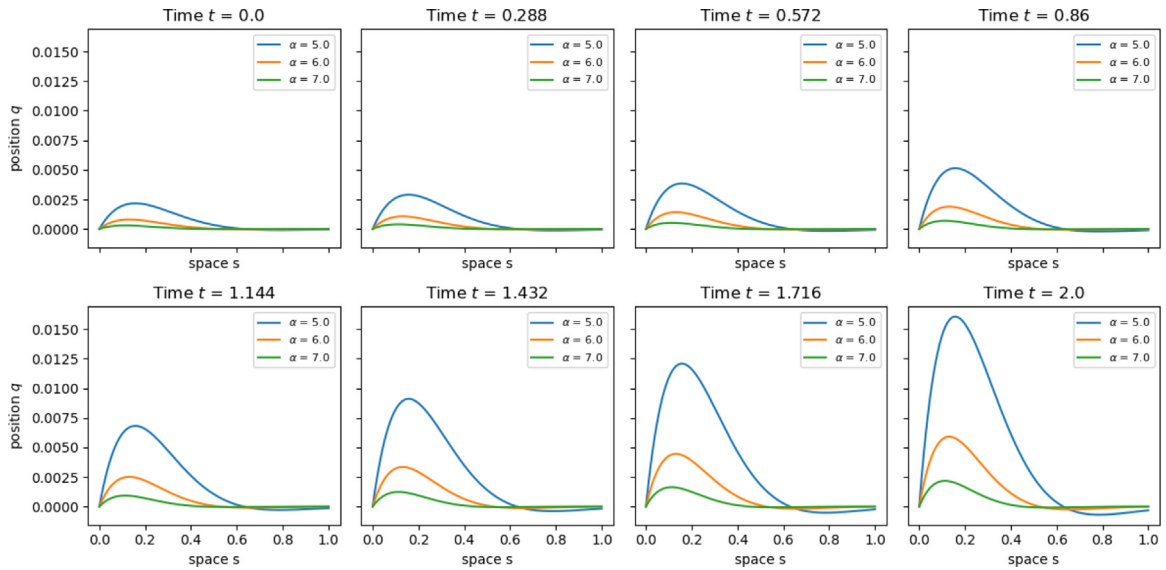
Fig. 12. Solution snapshots from the non-separable canonical example for different values of the parameter α.
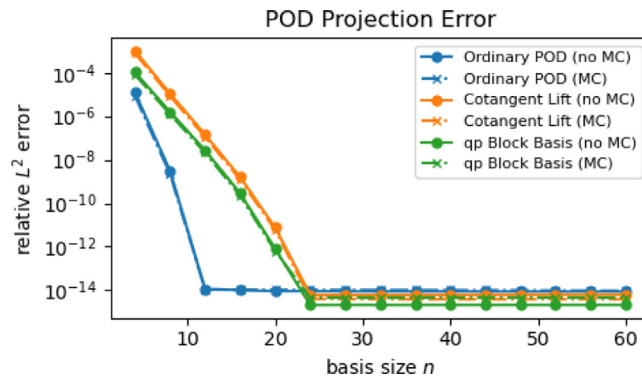

Fig. 13. POD projection errors corresponding to the bases used in the non-separable canonical example. "MC" indicates that mean-centering of the snapshots was performed.

Again, the predictive case is considered. For testing, six uniformly distributed parameters $\alpha \in [5.5, 14.5]$, disjoint from the training data, are chosen, and snapshot data of each solution in the range $[0, 10]$ is collected for comparison with the ROM integration. The ROMs are then tested over this interval beginning from each unseen initial condition, and the average relative error over all test snapshots is reported. In addition to the ROMs seen in the linear wave equation example, note that the H-OpInf ROM of [40] discussed in Section 2.4 is also reported.

The results of this experiment are displayed in Fig. 14. As expected, H-OpInf cannot produce a useful ROM, while C-H-OpInf is effective whenever the POD basis is built block-wise or with the cotangent lift algorithm. Interestingly, no Hamiltonian ROM algorithm is useful in the case where the POD basis is built from an SVD of the full snapshot matrix, while the intrusive Galerkin ROM appears to work quite well. This could be due to the fact that this Hamiltonian system decouples over $\mathbf{q}$ and $\mathbf{p}$: a quick calculation shows that $\mathbf{q} = e^t \mathbf{q}_0$ and $\mathbf{p} = e^{-t} \mathbf{p}_0$, so the scale separation in $\mathbf{q}, \mathbf{p}$ grows exponentially as t increases. Conversely, C-H-OpInf with a cotangent lift basis learns an accurate and stable ROM, while the H-OpInf algorithm is unable to do so due to its assumption of a block diagonal $\hat{\mathbf{A}}$. In addition to the state errors, conservation of the system Hamiltonian is displayed in Fig. 15, where it is clear that C-H-OpInf is conservative to a much higher order than either the Galerkin ROMs or the H-OpInf ROM.
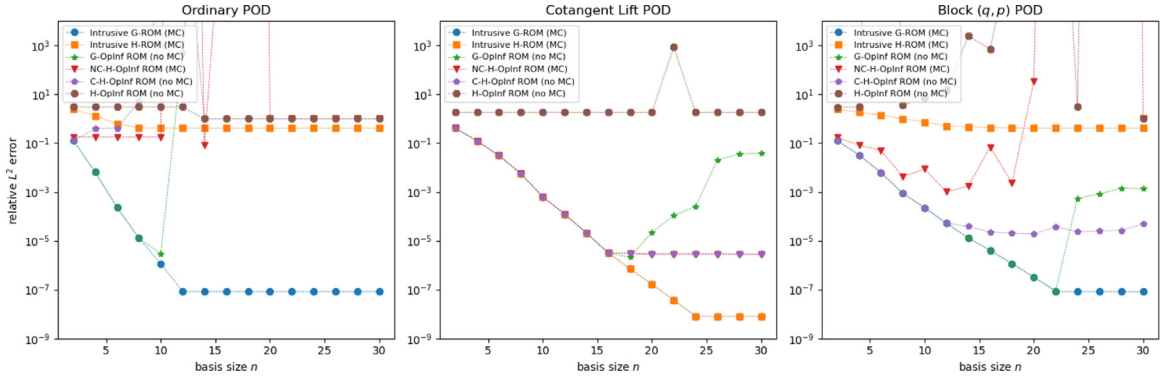
Fig. 14. Relative state errors as a function of basis modes for the ROMs in the non-separable canonical example (parametric predictive case $T = 10$). "MC" indicates the use of a mean-centered reconstruction.
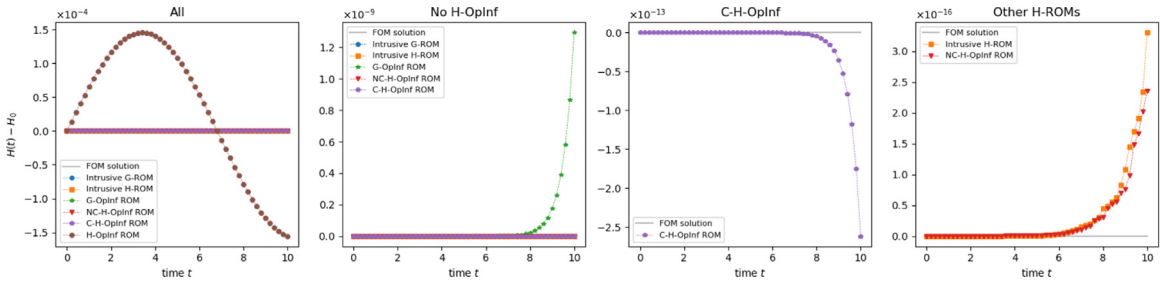

Fig. 15. ROM energy errors for the non-separable canonical example in the parametric predictive case ($\alpha = 7.3, T = 10$) when using a cotangent lift POD basis with mean-centering (where applicable) and with $n = 10$ modes. Note that the intrusive G-ROM and intrusive H-ROM are identical in this case.

5.3. Korteweg–De Vries equation

Moving beyond canonical Hamiltonian systems, consider the Korteweg–De Vries (KdV) equation [81]

$$\dot{x} = \alpha x x_s + \rho x_s + \gamma x_{sss}, \qquad x \in [-l, l] \times [0, T],$$

which depends on the parameters $\alpha, \rho, \gamma \in \mathbb{R}$. This equation has infinitely many integrals of motion [72], the first few of which are mass, momentum, and energy:

$$M(x) = \int_{-l}^{l} x \, ds, \qquad P(x) = \int_{-l}^{l} x^2 \, ds, \qquad E(x) = \int_{-l}^{l} \left(\frac{\alpha}{6} x^3 + \frac{\rho}{2} x^2 - \frac{\gamma}{2} x_s^2 \right) ds.$$

Moreover, KdV has a noncanonical bi-Hamiltonian structure, meaning that it can be recast as a Hamiltonian system in two distinct ways. While only the first form will be considered here, the second form is also interesting and (to date) no POD-ROMs for it have been seen in the literature. Therefore, some additional discussion regarding this second form is included in Appendix A.3.

5.3.1. First Hamiltonian formulation

Consider the Hamiltonian functional $H(x) = E(x)$, and note that its gradient satisfies

$$\nabla H(x) = \frac{\alpha}{2} x^2 + \rho x + \gamma x_{ss}.$$

Then, recalling that $L := \partial_s$ is an antisymmetric operator with respect to the usual metric on $L^2(\mathbb{R})$, it follows that $\dot{x} = L \nabla H(x)$ is a Hamiltonian system equivalent to the KdV equation. Since L has nontrivial kernel, this system is not canonical, meaning that there is no obvious way to separate the state x into position and momentum variables.

Assuming periodic boundary conditions and a discretization $\mathbf{x} \in \mathbb{R}^N$, the differential operators ∂_s and ∂_{ss} can be discretized with central finite differences as the circulant matrices

$$\mathbf{L} = \frac{1}{2\Delta x} \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & -1 \\ -1 & 0 & 1 & 0 & \dots & 0 \\ & & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & -1 & 0 & 1 \\ 1 & \dots & 0 & 0 & -1 & 0 \end{pmatrix}, \qquad \mathbf{B} = \frac{1}{(\Delta x)^2} \begin{pmatrix} -2 & 1 & 0 & 0 & \dots & 1 \\ 1 & -2 & 1 & 0 & \dots & 0 \\ & & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & 1 & -2 & 1 \\ 1 & \dots & 0 & 0 & 1 & -2 \end{pmatrix},$$

yielding the semidiscrete Hamiltonian system

$$\dot{\mathbf{x}} = \mathbf{L} \nabla H(\mathbf{x}) = \mathbf{L} \left(\frac{\alpha}{2} \mathbf{x}^2 + \rho \mathbf{x} + \nu \mathbf{B} \mathbf{x} \right).$$

Notice that the only nonlinearity in this system is polynomial in $\mathbf{x}$, meaning that the quadrature necessary for AVF time discretization (see Section 2.5) can be computed exactly. This leads to the fully discrete system

$$\frac{\mathbf{x}^{k+1} - \mathbf{x}^k}{\Delta t} = \mathbf{L} \left[\frac{\alpha}{6} \left((\mathbf{x}^k)^2 + \mathbf{x}^k \mathbf{x}^{k+1} + (\mathbf{x}^{k+1})^2 \right) + (\rho \mathbf{I} + \nu \mathbf{B}) \mathbf{x}^{k+\frac{1}{2}} \right],$$

where $\mathbf{x}^{k+\frac{1}{2}} = (1/2) \left(\mathbf{x}^k + \mathbf{x}^{k+1} \right)$ and vector products are interpreted element-wise. This represents the KdV FOM and is solved by Newton iteration. More precisely, at each time step k we have the (Δt-normalized) residual and Jacobian functions

$$\mathbf{R}^k (\mathbf{v}) = \mathbf{v} - \mathbf{x}^k - \Delta t \, \mathbf{L} \left[\frac{\alpha}{6} \left((\mathbf{x}^k)^2 + \mathbf{x}^k \mathbf{v} + \mathbf{v}^2 \right) + \left(\frac{\rho \mathbf{I} + \nu \mathbf{B}}{2} \right) (\mathbf{x}^k + \mathbf{v}) \right],$$

$$\mathbf{J}^k (\mathbf{v}) = \mathbf{I} - \frac{\Delta t}{2} \mathbf{L} \left[\frac{\alpha}{3} \left(\operatorname{Diag} (\mathbf{x}^k) + 2 \operatorname{Diag} (\mathbf{v}) \right) + \rho \mathbf{I} + \nu \mathbf{B} \right],$$

which are easily constructed and used to iterate $\mathbf{v}^{i+1} = \mathbf{v}^i - \mathbf{J}^k \left(\mathbf{v}^i \right)^{-1} \mathbf{R}^k \left(\mathbf{v}^i \right)$ until convergence. It can be checked that this scheme exactly preserves the discrete Hamiltonian,

$$H(\mathbf{x}) = \frac{1}{2} \sum_{j=1}^{N} \left(\frac{\alpha}{3} x_j^3 + \rho x_j^2 - \nu \left(\frac{x_{j+1} - x_j}{\Delta x} \right)^2 \right) \Delta x.$$

From this, it is possible to compute the intrusive G-ROM and intrusive H-ROM as described in Section 2.5. Particularly, straightforward Galerkin projection onto a reduced basis contained in the columns of $\mathbf{U}$ yields the reduced-order G-ROM system

$$\begin{aligned} \dot{\hat{\mathbf{x}}} &= \mathbf{U}^\mathsf{T} \mathbf{L} \nabla H \left(\mathbf{x}_0 + \mathbf{U} \hat{\mathbf{x}} \right) = \mathbf{U}^\mathsf{T} \mathbf{L} \left[\frac{\alpha}{2} \left(\mathbf{x}_0 + \mathbf{U} \hat{\mathbf{x}} \right)^2 + (\rho \mathbf{I} + \nu \mathbf{B}) \left(\mathbf{x}_0 + \mathbf{U} \hat{\mathbf{x}} \right) \right] \\ &= \left(\frac{\alpha}{2} \mathbf{U}^\mathsf{T} \mathbf{L} \mathbf{x}_0^2 + \mathbf{U}^\mathsf{T} \mathbf{L} (\rho \mathbf{I} + \nu \mathbf{B}) \mathbf{x}_0 \right) + \mathbf{U}^\mathsf{T} \mathbf{L} \left(\alpha \operatorname{Diag} (\mathbf{x}_0) + (\rho \mathbf{I} + \nu \mathbf{B}) \right) \mathbf{U} \hat{\mathbf{x}} + \frac{\alpha}{2} \mathbf{U}^\mathsf{T} \mathbf{L} \left(\mathbf{U} \hat{\mathbf{x}} \right)^2 \\ &:= \hat{\mathbf{c}} + \hat{\mathbf{C}} \hat{\mathbf{x}} + \hat{\mathbf{T}} \left(\hat{\mathbf{x}}, \hat{\mathbf{x}} \right), \end{aligned}$$

where $\hat{\mathbf{T}}$ is a precomputable order-three tensor with components $\hat{T}_{bc}^a = (\alpha/2) \, U_i^a L_j^i U_b^j U_c^j$. Similarly, a reduced-order H-ROM system is given by

$$\begin{aligned} \dot{\hat{\mathbf{x}}} &= \hat{\mathbf{L}} \nabla \hat{H} \left(\hat{\mathbf{x}} \right) = \hat{\mathbf{L}} \mathbf{U}^\mathsf{T} \left[\frac{\alpha}{2} \left(\mathbf{x}_0 + \mathbf{U} \hat{\mathbf{x}} \right)^2 + (\rho \mathbf{I} + \nu \mathbf{B}) \left(\mathbf{x}_0 + \mathbf{U} \hat{\mathbf{x}} \right) \right] \\ &= \hat{\mathbf{L}} \left[\frac{\alpha}{2} \left(\mathbf{U}^\mathsf{T} \mathbf{x}_0^2 + 2 \, \mathbf{U}^\mathsf{T} \operatorname{Diag} (\mathbf{x}_0) \mathbf{U} \hat{\mathbf{x}} + \mathbf{U}^\mathsf{T} \left(\mathbf{U} \hat{\mathbf{x}} \right)^2 \right) + (\rho \mathbf{I} + \nu \mathbf{B}) \left(\mathbf{x}_0 + \mathbf{U} \hat{\mathbf{x}} \right) \right], \\ &= \hat{\mathbf{L}} \left[\left(\frac{\alpha}{2} \mathbf{U}^\mathsf{T} \mathbf{x}_0^2 + \mathbf{U}^\mathsf{T} (\rho \mathbf{I} + \nu \mathbf{B}) \mathbf{x}_0 \right) + \mathbf{U}^\mathsf{T} \left(\alpha \operatorname{Diag} (\mathbf{x}_0) + (\rho \mathbf{I} + \nu \mathbf{B}) \right) \mathbf{U} \hat{\mathbf{x}} + \frac{\alpha}{2} \mathbf{U}^\mathsf{T} \left(\mathbf{U} \hat{\mathbf{x}} \right)^2 \right] \\ &:= \hat{\mathbf{L}} \left(\hat{\mathbf{c}} + \hat{\mathbf{C}} \hat{\mathbf{x}} + \hat{\mathbf{T}} \left(\hat{\mathbf{x}}, \hat{\mathbf{x}} \right) \right), \end{aligned}$$

where $\hat{\mathbf{T}} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is a precomputable order-three tensor with components $\hat{T}_{bc}^a = (\alpha/2) U_i^a U_b^i U_c^i$. In either case, applying AVF for temporal discretization and using the fact that $\hat{\mathbf{T}}$ is symmetric in its lower indices yields
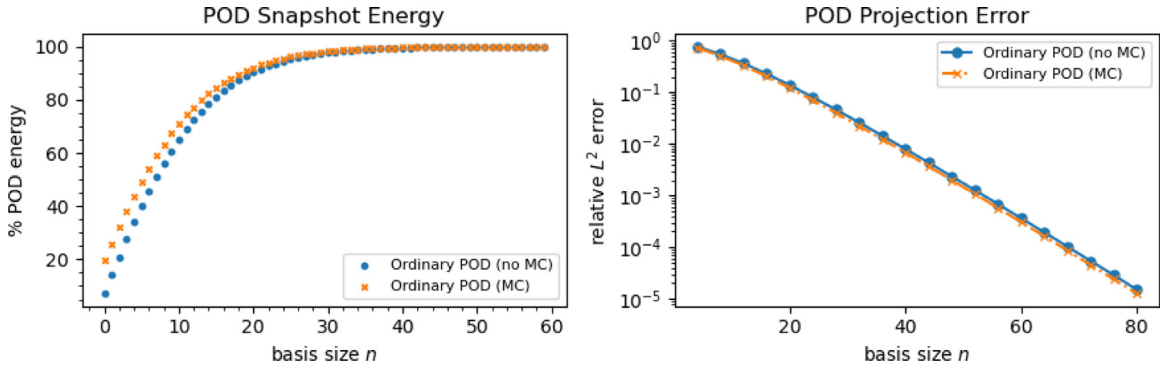
Fig. 16. POD snapshot energies (left) and projection errors (right) corresponding to the bases used in the KdV equation example. "MC" indicates mean-centering of the snapshots was performed.

the fully discrete ROM (note that $\hat{\mathbf{L}} = \mathbf{I}$ in the G-ROM),

$$\frac{\hat{\mathbf{x}}^{k+1} - \hat{\mathbf{x}}^k}{\Delta t} = \hat{\mathbf{L}} \left[\hat{\mathbf{c}} + \hat{\mathbf{C}} \hat{\mathbf{x}}^{k+\frac{1}{2}} + \frac{1}{3} \left(2 \hat{\mathbf{T}} \left(\hat{\mathbf{x}}^k, \hat{\mathbf{x}}^{k+\frac{1}{2}} \right) + \hat{\mathbf{T}} \left(\hat{\mathbf{x}}^{k+1}, \hat{\mathbf{x}}^{k+1} \right) \right) \right],$$

which is again solvable with Newton iterations. In this case, the (Δt-normalized) residual and Jacobian at time step k are given by

$$\hat{\mathbf{R}}^k \left(\hat{\mathbf{v}} \right) = \hat{\mathbf{v}} - \hat{\mathbf{x}}^k - \Delta t \, \hat{\mathbf{L}} \left[\hat{\mathbf{c}} + \frac{1}{2} \mathbf{C} \left(\hat{\mathbf{x}}^k + \hat{\mathbf{v}} \right) + \frac{1}{3} \left(\hat{\mathbf{T}} \left(\hat{\mathbf{x}}^k, \hat{\mathbf{x}}^k + \hat{\mathbf{v}} \right) + \hat{\mathbf{T}} \left(\hat{\mathbf{v}}, \hat{\mathbf{v}} \right) \right) \right],$$

$$\hat{\mathbf{J}}^k \left(\hat{\mathbf{v}} \right) = \mathbf{I} - \Delta t \, \hat{\mathbf{L}} \left[\frac{1}{2} \mathbf{C} + \frac{1}{3} \left(\hat{\mathbf{T}} \left(\hat{\mathbf{x}}^k \right) + 2 \hat{\mathbf{T}} \left(\hat{\mathbf{v}} \right) \right) \right],$$

where $\hat{\mathbf{T}} \left(\hat{\mathbf{x}}^k \right)$ indicates that the symmetric tensor $\hat{\mathbf{T}}$ is applied to the vector $\hat{\mathbf{x}}$ in either of its lower indices, yielding an $n \times n$ matrix.

The goal is now to compare these intrusive ROMs to the NC-H-OpInf ROM from Section 3 as well as a G-OpInf ROM which does not incorporate any structure information. To facilitate a fair comparison, the G-OpInf procedure employed presently will not be black-box, but will instead aim to infer $\hat{\mathbf{L}}$ in the intrusive H-ROM $\dot{\hat{\mathbf{x}}} = \hat{\mathbf{L}} \hat{\nabla} H(\hat{\mathbf{x}})$ similarly to NC-H-OpInf, but using the generic technique of Section 2.3. This way, both the G-OpInf ROM and the NC-H-OpInf ROM are assumed to use analytic knowledge of the nonlinear part of $\nabla \hat{H}$, and both OpInf ROMs can be integrated similarly to the intrusive H-ROM, but with the intrusive governing operator replaced by the inferred one. For experimental parameters, we choose $l = 20$, $(\alpha, \beta, \gamma) = (-6, 0, -1)$, $N = 500$, and an initial condition

$$x_0(s) = \text{sech}^2 \left(\frac{s}{\sqrt{2}} \right),$$

which generates a soliton solution for $s \in \mathbb{R}$. To train the OpInf ROMs, 1001 snapshots of the solution $\mathbf{x}$ and the gradient $\nabla H(\mathbf{x})$ are collected uniformly on the interval $[0, T]$ with $T = 20$.

Recall that there is no analogue of a block basis or cotangent lift method in the case of noncanonical Hamiltonian systems, so the POD bases $\mathbf{U}$ employed here are trained using the full snapshot matrix. The associated snapshot energies and projection errors are displayed in Fig. 16, where it is evident that the snapshot energy accumulates quite slowly with the addition of basis modes. On the other hand, the use of ordinary POD bases again allows for the one-shot computation of all OpInf ROMs via truncation from the OpInf solution at the highest number of modes, creating large savings in computational cost. While this is unlikely to be provably optimal in view of Proposition 3.1, the empirical difference in performance is small enough to justify the substantial decrease in computational time necessary for computing the ROMs.

As before, the performance of these ROMs in both predictive and reproductive cases is considered. The relative state errors of each ROM as a function of basis modes are shown in Fig. 17, with the reproductive case ($T = 20$) on the left and the predictive case ($T = 100$) on the right. Reported are the errors with and without mean-centering by $\mathbf{x}_0$, as it is interesting to observe the effect of this choice. Notice that mean-centering in the POD basis appears
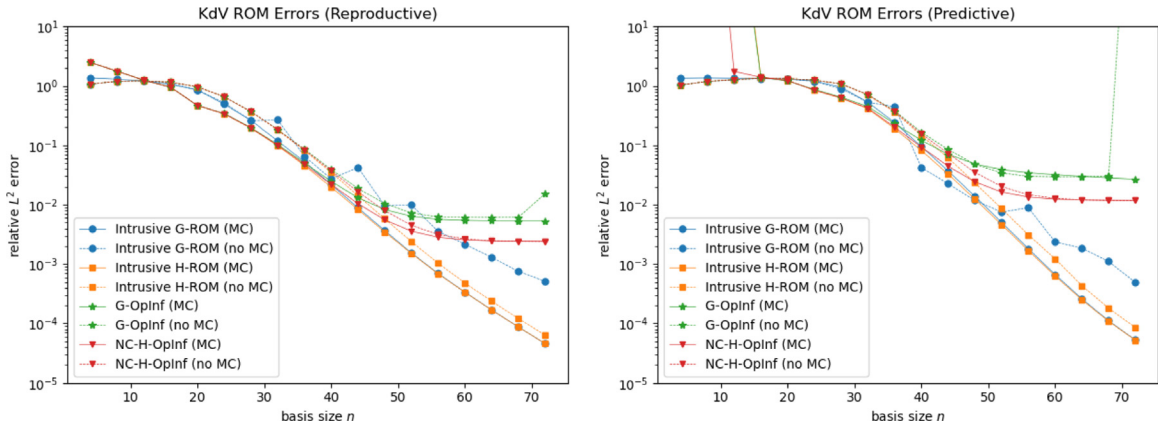
Fig. 17. Relative state errors as a function of basis modes for the ROMs in the KdV equation example. Left: reproductive case ($T = 20$). Right: predictive case ($T = 100$). "MC" indicates the use of a mean-centered reconstruction.
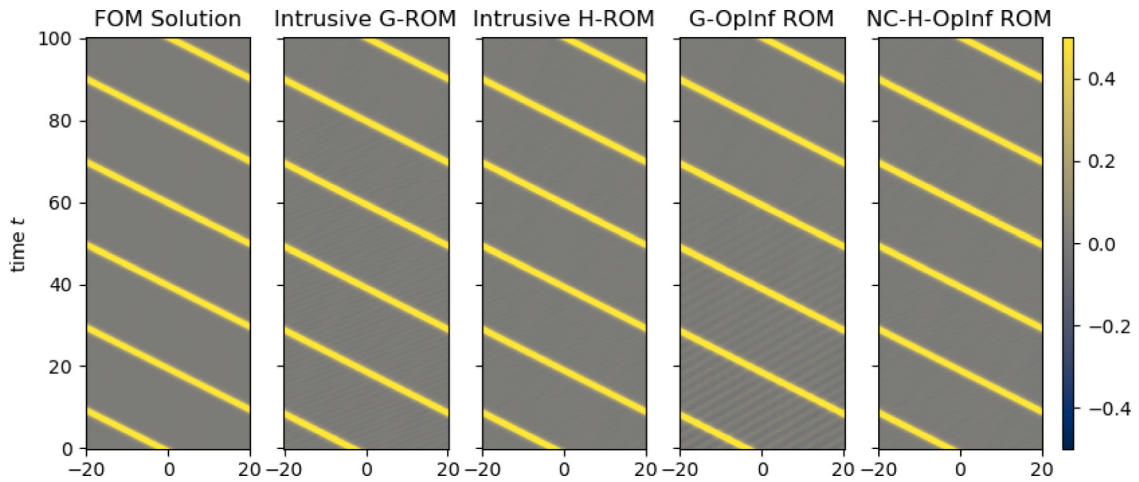

Fig. 18. FOM and ROM solutions to the KdV equation example with mean-centering (where applicable) and $n = 32$ modes (predictive case $T = 100$).

to make the intrusive ROMs more accurate and the OpInf ROMs more stable, perhaps because it ensures that the value of the Hamiltonian is exact at $\hat{\mathbf{x}} = \mathbf{0}$. However, in either case the NC-H-OpInf ROM remains more accurate and stable than the G-OpInf ROM, demonstrating the benefits of preserving antisymmetry in the learned operator. Figs. 18 and 19 provide a comparative illustration of the FOM and ROM solutions in the case that $n = 32$ modes and mean-centering is applied. While both OpInf ROMs are capable of predicting the general trajectory of the soliton, the NC-H-OpInf ROM exhibits much less artifacting over the rest of the domain—a consequence of capturing the correct latent space dynamics. Note that, in either case, the performance of the OpInf ROM improves substantially as the number of modes increases, eventually leveling off around $n = 60$ as a consequence of the failure of the learned dynamics to remain Markovian (see [82]).

Besides decreased state errors, Fig. 20 shows the improved conservation of energy, mass, and momentum displayed by the H-ROMs over the G-ROMs when a mean-centered POD basis with $n = 48$ modes is used. Again, the conservation behavior of the intrusive H-ROM and the NC-H-OpInf ROM is orders of magnitude more accurate than the intrusive G-ROM or the G-OpInf ROM, reflecting the notion of the Hamiltonian as a conserved quantity. It is also clear that the mass and momentum are preserved by the H-ROMs at least as well as the by the G-ROMs, demonstrating that other conserved quantities are not sacrificed for Hamiltonian preservation. Finally, it is useful to note that, as before in the canonical case, this behavior persists regardless of the number of basis modes used in the ROM.
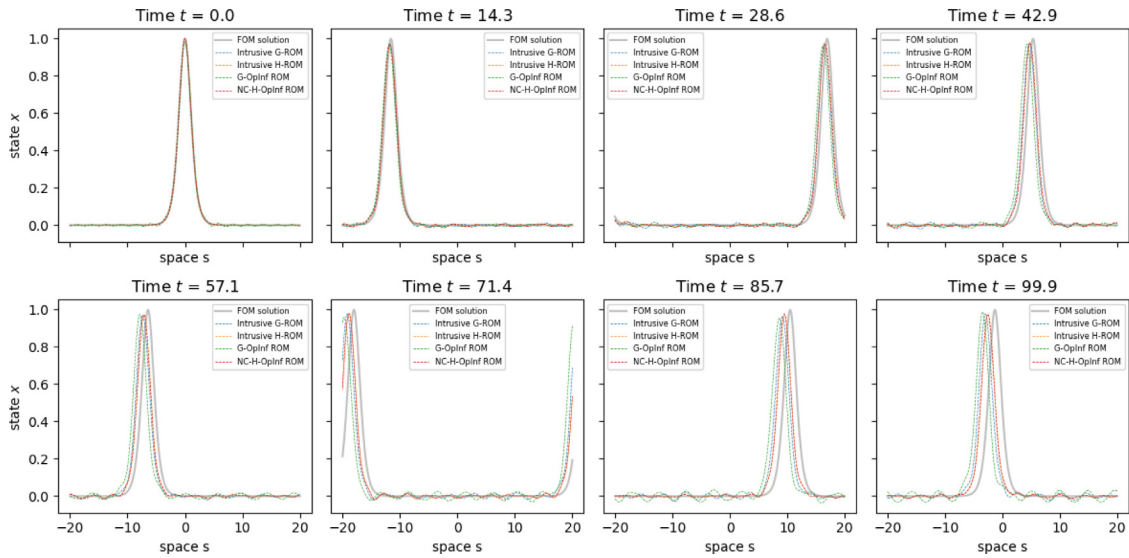
Fig. 19. Snapshots in time corresponding to the FOM and ROM solutions to the KdV equation example with mean-centering (where applicable) and $n = 32$ modes (predictive case $T = 100$).
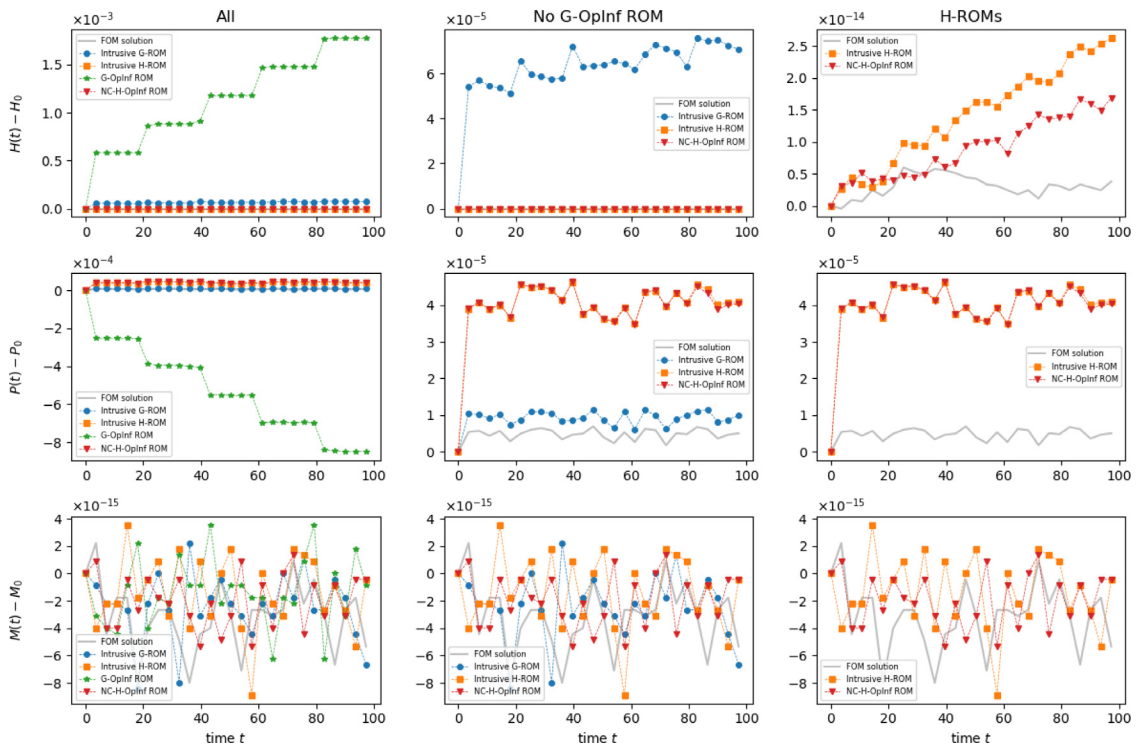

Fig. 20. Errors in conserved quantities for the (mean-centered) ROMs in the KdV equation example in the predictive case ($T = 100$) when using a POD basis with $n = 48$ modes.

5.4. Benjamin-Bona-Mahoney equation

As another example, consider the Benjamin–Bona–Mahony (BBM) equation, also referred to as the regularized long-wave equation,

$$\dot{x} = \alpha x_s + \beta x x_s - \gamma \dot{x}_{ss}.$$

The BBM equation represents an alternative to the KdV equation introduced in [83] and later in [84], intended as a model for the unidirectional propagation of long-range water waves with small amplitude. This equation has a noncanonical Hamiltonian form defined by the data:

$$L = -\left(1 - \partial_s^2\right)^{-1} \partial_s, \qquad H(x) = \frac{1}{2} \int_0^\ell \alpha x^2 + \frac{\beta}{3} x^3 \, ds,$$

where $L^2(\mathbb{R})$ is skew-symmetric and H is the Hamiltonian. The BBM equation is distinct from KdV in that it is not completely integrable, possessing only three globally conserved quantities. In addition to H, these are the momentum and kinetic energy:

$$P(x) = \int_0^\ell (x - \gamma x_{ss}) \, ds, \qquad KE(x) = \frac{1}{2} \int_0^\ell \left(x^2 + \gamma x_s^2\right) ds.$$

Because the governing operator L is unwieldy to spatially discretize, the BBM equation has (to the authors' knowledge) never been simulated in Hamiltonian form. On the other hand, it is straightforward to discretize this system with pseudospectral techniques. In particular, denote the Fourier and inverse Fourier transforms of a function $f : \mathbb{R} \to \mathbb{R}$ by

$$\hat{f}(\xi) := (\mathcal{F}f)(\xi) = \int_{-\infty}^\infty f(\xi) e^{-2\pi i \xi x}, \qquad f(x) = \left(\mathcal{F}^{-1} \hat{f}\right)(x) = \int_{-\infty}^\infty \hat{f}(\xi) e^{2\pi i \xi x}.$$

Then, basic properties of the Fourier transform (see e.g. [85]) show that the BBM equation has the equivalent (non-Hamiltonian) expression

$$\dot{x} = \mathcal{F}^{-1} \left(\frac{-2\pi i \, \mathcal{F}\left(\nabla H(x)\right)(\xi)}{1 + 4\gamma \pi^2 \xi^2} \right)(x),$$

where $\nabla H(x) = \alpha x + (\beta/2)x^2$. The FOM used presently is generated from this expression by semidiscretizing x as $\mathbf{x} \in \mathbb{R}^N$ with $N = 1024$ and utilizing the fast Fourier transform and "solve_ivp" functions found in the SciPy library [86]. More precisely, given the discrete Hamiltonian

$$H(\mathbf{x}) = \frac{1}{2} \sum_{j=1}^N \left(\alpha x_j^2 + \frac{\beta}{3} x_j^3 \right) \Delta x,$$

the FOM is computed by solving the system

$$\dot{\mathbf{x}} = \mathcal{F}^{-1} \left(\frac{-2\pi i \, \mathcal{F}\left(\nabla H(\mathbf{x})\right)}{1 + 4\gamma \pi^2 \boldsymbol{\xi}^2} \right),$$

with an explicit Runge–Kutta method of order 8, and $\mathcal{F}, \mathcal{F}^{-1}$ are the discrete Fourier and inverse Fourier transforms defined in terms of the vector $\mathbf{k} = \mathbf{m} = \begin{pmatrix} 0 & 1 & ... & N-1 \end{pmatrix}^{\mathsf{T}}$ of nonnegative integers at most $N-1$,

$$\mathcal{F}(\mathbf{x}) = \sum_{m=0}^{N-1} x_m \exp\left(-\frac{2\pi i}{N} \mathbf{k}m \right), \qquad \mathcal{F}^{-1}(\boldsymbol{\xi}) = \frac{1}{N} \sum_{k=0}^{N-1} \xi_k \exp\left(\frac{2\pi i}{N} k\mathbf{m} \right).$$

Since it is challenging to build an intrusive Hamiltonian ROM for the BBM system, it is useful to see if the governing operator can be effectively learned by the OpInf methods seen in Section 3. This would allow for a nonintrusive spatial ROM which preserves the underlying Hamiltonian structure, which could be valuable in cases where conservation is paramount. As before, accomplishing this means inferring $\hat{\mathbf{L}}$ in $\dot{\hat{\mathbf{x}}} = \hat{\mathbf{L}} \nabla \hat{H}(\hat{\mathbf{x}})$, which is readily done by solving the linear system in Eq. (8). Similar to the case of KdV, this result will be compared to the
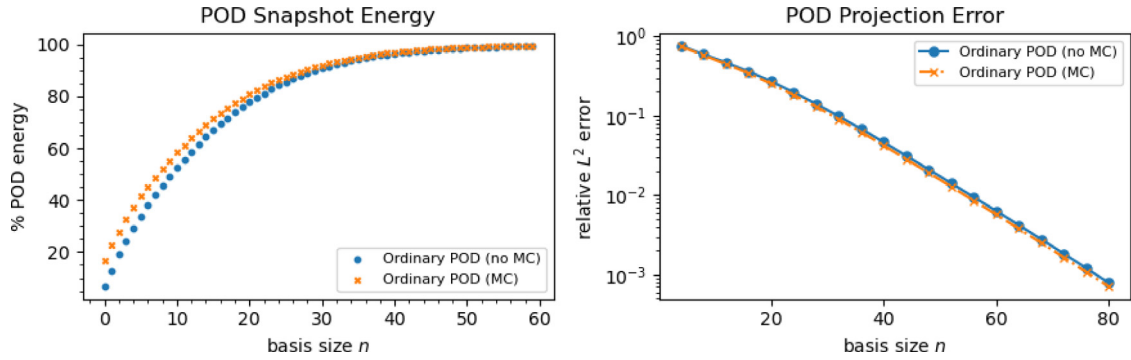
Fig. 21. Left: POD snapshot energies (left) and projection errors (right) corresponding to the bases used in the BBM equation example. "MC" indicates mean-centering of the snapshots was performed.

nonintrusive ROM generated by inferring $\hat{\mathbf{L}}$ using the generic OpInf technique described in Section 2.3. Provided a suitable $\hat{\mathbf{L}}$ is available, the desired Hamiltonian ROM becomes

$$
\dot{\hat{\mathbf{x}}} = \hat{\mathbf{L}}\nabla\hat{H}\left(\mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}}\right) = \hat{\mathbf{L}}\mathbf{U}^{\mathsf{T}}\left[\alpha\left(\mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}}\right) + \frac{\beta}{2}\left(\mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}}\right)^2\right]
$$

$$
= \hat{\mathbf{L}}\left[\mathbf{U}^{\mathsf{T}}\left(\alpha\mathbf{x}_0 + \frac{\beta}{2}\mathbf{x}_0^2\right) + \mathbf{U}^{\mathsf{T}}\left(\alpha\mathbf{I} + \beta\operatorname{Diag}\mathbf{x}_0\right)\mathbf{U} + \hat{\mathbf{T}}\left(\hat{\mathbf{x}},\hat{\mathbf{x}}\right)\right]
$$

$$
= \hat{\mathbf{L}}\left(\hat{\mathbf{c}} + \hat{\mathbf{C}}\hat{\mathbf{x}} + \hat{\mathbf{T}}\left(\hat{\mathbf{x}},\hat{\mathbf{x}}\right)\right),
$$

where $\hat{\mathbf{T}}$ is precomputable with components $T_{bc}^a = (\beta/2)U_i^a U_b^i U_c^i$ similar to the case of KdV. With these definitions of $\hat{\mathbf{c}}, \hat{\mathbf{C}}, \hat{\mathbf{T}}$, AVF integration yields the BBM ROM system

$$
\frac{\hat{\mathbf{x}}^{k+1} - \hat{\mathbf{x}}^k}{\Delta t} = \hat{\mathbf{L}}\left[\hat{\mathbf{c}} + \hat{\mathbf{C}}\hat{\mathbf{x}}^{k+\frac{1}{2}} + \frac{1}{3}\left(2\,\hat{\mathbf{T}}\left(\hat{\mathbf{x}}^k, \hat{\mathbf{x}}^{k+\frac{1}{2}}\right) + \hat{\mathbf{T}}\left(\hat{\mathbf{x}}^{k+1}, \hat{\mathbf{x}}^{k+1}\right)\right)\right],
$$

which is solvable with Newton iterations identically to the KdV system.

For the present experiment, the parameters in the governing equation are set to $(\alpha, \beta, \gamma) = \left(1, 1, 10^{-4}\right)$ and 2001 snapshots of $\mathbf{x}, \nabla H(\mathbf{x})$ are collected in the interval $[0, T]$ for $T = 0.5$, starting from the initial condition

$$
\mathbf{x}_0 = 7\operatorname{sech}^2\left(\sqrt{\frac{1}{5\gamma}}\,(\mathbf{s} - 0.25)\right) + 3\operatorname{sech}^2\left(\sqrt{\frac{1}{6\gamma}}\,(\mathbf{s} - 0.35)\right).
$$

This generates a nonperiodic 2-solitary wave solution, which experiences an inelastic collision over the length of the training integration. The relative POD energies and reconstruction errors of the computed POD bases are shown in Fig. 21, where it is seen that the reconstruction error decays quite slowly as a function of basis modes. It is further interesting to observe that the first eigenvector of the mean-centered basis contains much more information than the others, although this does not appear to yield a faster decrease in reconstruction error. These data are used to train the NC-H-OpInf ROM and a corresponding G-OpInf ROM.

For testing, AVF time integration is carried out to $T = 0.5$ and $T = 1$, respectively, representing reproductive and predictive scenarios. The relative errors of these ROMs as a function of basis size are displayed in Fig. 22, where it can be seen that the errors for the NC-H-OpInf ROMs are about half of those for the G-OpInf ROMs. However, it is also clear that this example poses a much greater challenge for either OpInf ROM, likely due to the nonperiodic and inelastic nature of the solitary wave collisions present in the BBM solution, as well as the complicated form of the governing operator $\mathbf{L}$. It is interesting to note the effect of mean-centering here: in either case, there is a significant gain in performance for middling numbers of modes (20–60) which diminishes as more modes are added.

Visual comparisons of the FOM and ROM solutions in the predictive case are shown in Figs. 23 and 24, where two inelastic collisions are pictured and the second collision occurs outside the range of the training data (note
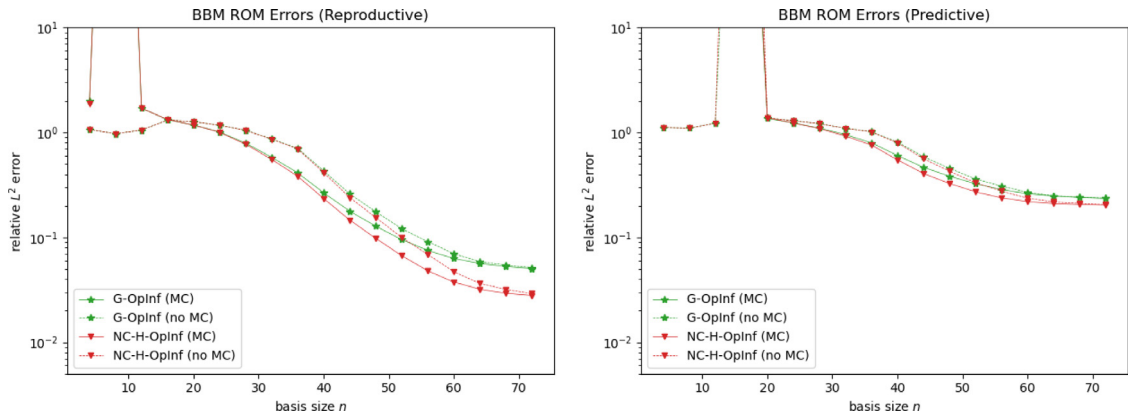
Fig. 22. Relative state errors as a function of basis modes for the ROMs in the BBM equation example. Left: reproductive case ($T = 0.5$). Right: predictive case ($T = 1$). "MC" indicates the use of a mean-centered reconstruction.
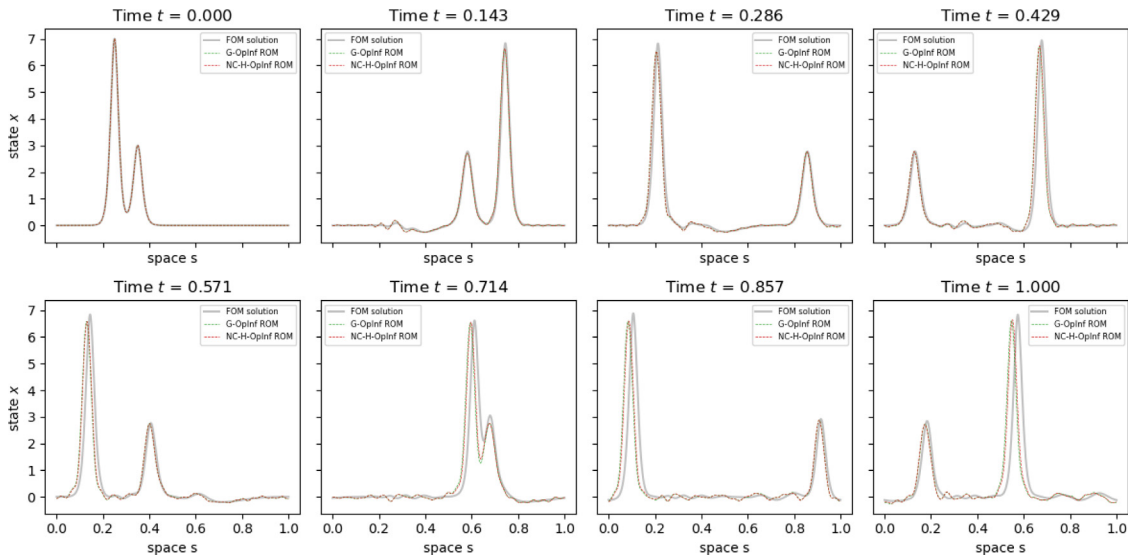

Fig. 23. Space–time plots showing the evolution of the BBM ROMs in the predictive case ($T = 1$) with mean-centering and using $n = 44$ modes.

the difference in the tails). Even at $n = 44$ modes, the collisions are relatively well captured, validating the hypothesis that the NC-H-OpInf procedure can produce a useful and nonintrusive spatial ROM even when the FOM is pseudospectral and the involved operator $\hat{\mathbf{L}}$ cannot be readily discretized by standard techniques.

Moving beyond state errors, the difference in conserved quantities between the NC-H-OpInf and G-OpInf ROMs is displayed in Fig. 25, using mean-centered POD bases and $n = 44$ modes. From this, it is evident that the Hamiltonian is conserved exactly by the NC-H-OpInf ROM but not by the G-ROM (note that the FOM is conservative to $\mathcal{O}\left(10^{-12}\right)$), likely enabling the NC-H-OpInf ROM to capture small-scale features like the tails of the solitons in Fig. 23 much more realistically. Moreover, it appears that both OpInf ROMs are capable of conserving momentum exactly and kinetic energy to a relatively low-order. It is interesting to note that mean-centering makes a difference here: without this choice (not shown here), the momentum conservation of both ROMs is on the same order as the kinetic energy conservation.
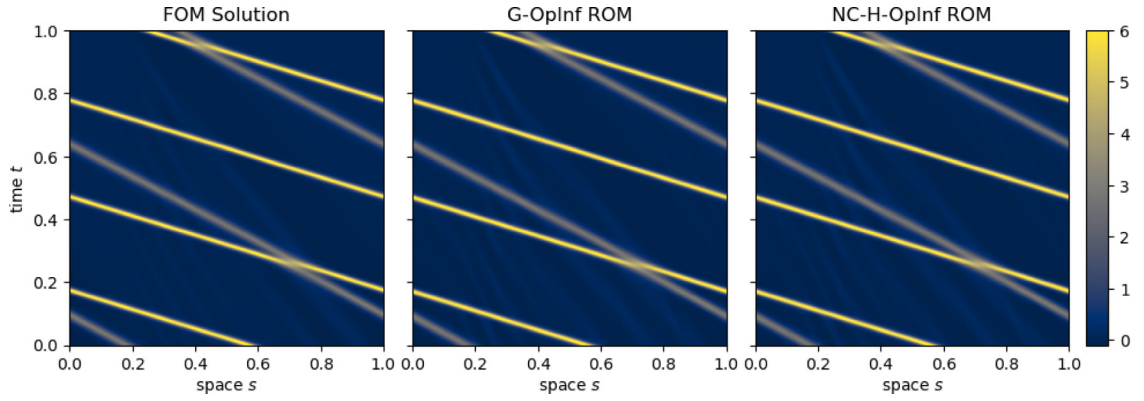
Fig. 24. Snapshots showing the evolution of the BBM ROMs in the predictive case $(T = 1)$ with mean-centering and using $n = 44$ modes.
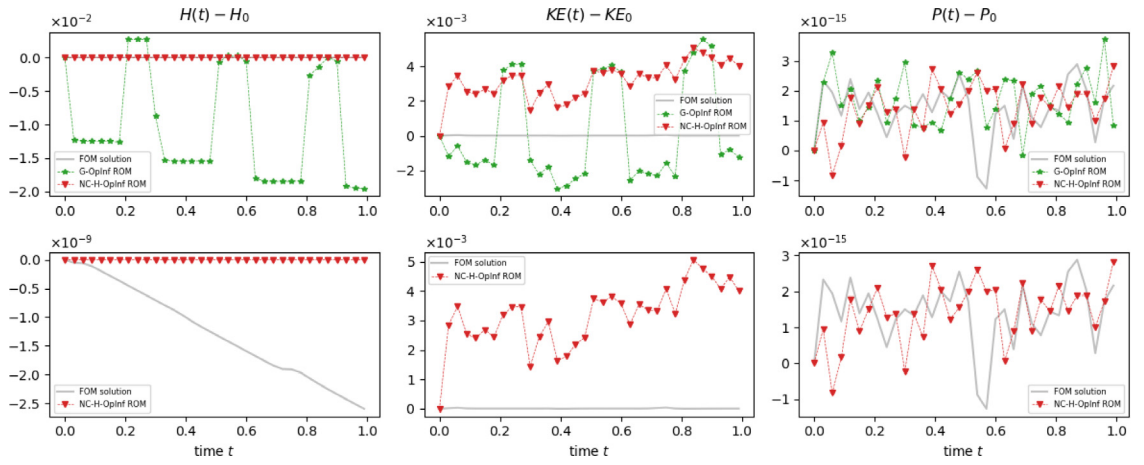

Fig. 25. Plots showing the evolution of the conserved quantities for the mean-centered ROMs in the BBM equation example in the predictive case $(T = 1)$ using $n = 44$ modes.

5.5. Three-dimensional linear elasticity

The final example considered in this work involves a moderate size three-dimensional (3D) linear elasticity problem, given by the following equations of motion:

$$\rho \ddot{\mathbf{q}} = \nabla \cdot \boldsymbol{\sigma}, \qquad \text{on } \Omega \in \mathbb{R}^3. \tag{10}$$

In (10), $\mathbf{q} \in \mathbb{R}^3$ is the displacement vector, $\rho > 0$ is the material density, and $\boldsymbol{\sigma}$ is the Cauchy stress tensor. We assume that the material is elastic and follows Hooke's law, so that the components of $\boldsymbol{\sigma}$ satisfy

$$\sigma_{ij} := \lambda \text{Tr}(\boldsymbol{\epsilon}) \delta_{ij} + 2\mu \epsilon_{ij}, \qquad 1 \le i \le j \le 3, \tag{11}$$

where $\lambda, \mu > 0$ are the Lamé coefficients and

$$\boldsymbol{\epsilon} := \frac{1}{2} \left[\nabla \mathbf{q} + (\nabla \mathbf{q})^\mathsf{T} \right] \tag{12}$$

is the infinitesimal strain tensor. It can be shown [87] that the Hamiltonian for (10) can be expressed in terms of (noncanonical) position and velocity variables:

$$H(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2} \int_\Omega \left(\rho |\dot{\mathbf{q}}|^2 + \lambda [\text{Tr}(\boldsymbol{\epsilon})]^2 + 2\mu |\boldsymbol{\epsilon}|^2 \right) dV. \tag{13}$$

(a) $t = 0$ s   (b) $t = 1.0 \times 10^{-3}$ s   (c) $t = 2.0 \times 10^{-3}$ s

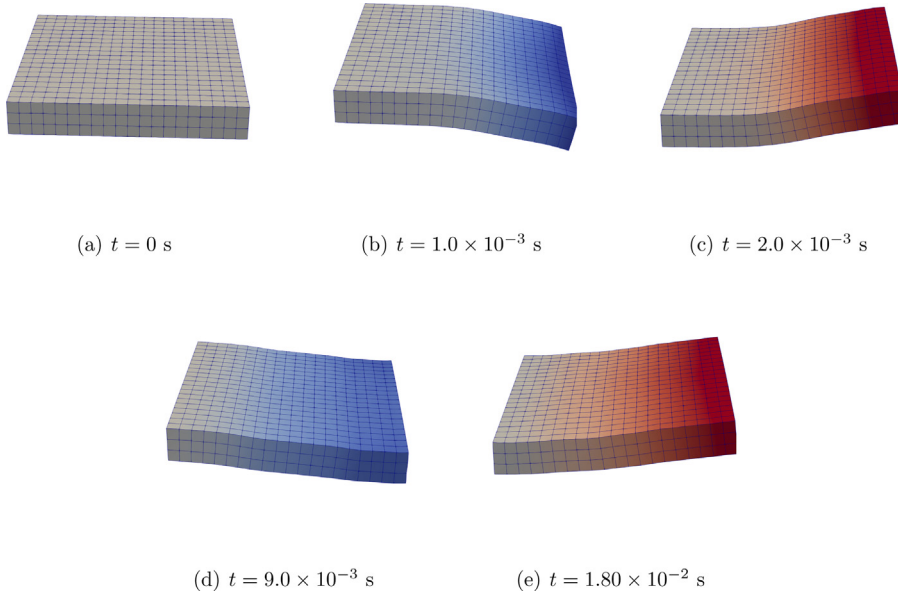(d) $t = 9.0 \times 10^{-3}$ s   (e) $t = 1.80 \times 10^{-2}$ s

Fig. 26. Plot of FOM s_3–displacement, scaled by a factor of ten, at several times for the 3D linear elastic cantilever beam problem. The colorbar range is -2.3×10^{-3} m (blue) to 2.3×10^{-3} m (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We remark that, in 1D, (13) reduces to the linear wave equation (9) considered earlier in Section 5.1. The main purpose of this example is to demonstrate the utility of non-intrusive ROMs on an application in which the FOM is implemented within a large HPC code without embedded ROM capabilities, making intrusive model reduction infeasible.

Herein, Eq. (10) is assumed to be discretized in space using the finite element method (FEM), per common practice in the field of solid mechanics. Doing so gives a semi-discrete system of the form

$$\mathbf{M\ddot{q}} + \mathbf{Kq} = \mathbf{0}, \tag{14}$$

where (overloading notation) $\mathbf{q} \in \mathbb{R}^M$ is the discretized displacement field, and $\mathbf{M} \in \mathbb{R}^{M \times M}$ and $\mathbf{K} \in \mathbb{R}^{M \times M}$ are the mass and stiffness matrices, respectively. Letting $\mathbf{p} := \mathbf{M\dot{q}}$ denote the (overloaded) momentum, $N = 2M$, and defining $\mathbf{x} := (\mathbf{q} \quad \mathbf{p})^\mathsf{T} \in \mathbb{R}^N$, (14) can be written as the following canonical Hamiltonian system:

$$\dot{\mathbf{x}} = \mathbf{J}\nabla H(\mathbf{x}) = \begin{pmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{K} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^{-1} \end{pmatrix} \mathbf{x}, \tag{15}$$

where H is a quadratic discrete Hamiltonian of the form

$$H(\mathbf{x}) = \frac{1}{2} \left(\mathbf{q}^\mathsf{T}\mathbf{Kq} + \mathbf{p}^\mathsf{T}\mathbf{M}^{-1}\mathbf{p} \right). \tag{16}$$

The test case considered presently is a classical solid mechanics benchmark involving a vibrating rectangular cantilever plate of size $0.2 \times 0.2 \times 0.03$ m, so that $\Omega = (0, 0.2) \times (0, 0.2) \times (0, 0.03) \in \mathbb{R}^3$. Let $\mathbf{s}^\mathsf{T} := (s_1, s_2, s_3)^\mathsf{T} \in \mathbb{R}^3$ denote the coordinate (position) vector. Here, the left side of the plate is clamped, meaning that a homogeneous Dirichlet boundary condition $\mathbf{q} = \mathbf{0}$ is imposed on $\Gamma_l := \{s_2, s_3 \in \Omega : s_1 = 0\}$. Homogeneous Neumann boundary conditions are prescribed on the remaining boundaries of Ω, indicating that these boundaries are free surfaces. The problem is initialized by prescribing an initial velocity of 100 m/s in the s_3-direction on the right boundary of the domain, $\Gamma_r := \{s_2, s_3 \in \bar{\Omega} : s_1 = 0.2\}$:

$$\dot{\mathbf{q}}(\mathbf{s}, 0) = \begin{pmatrix} 0 \\ 0 \\ 100 \end{pmatrix}, \quad \text{for } \mathbf{s} \in \Gamma_r. \tag{17}$$
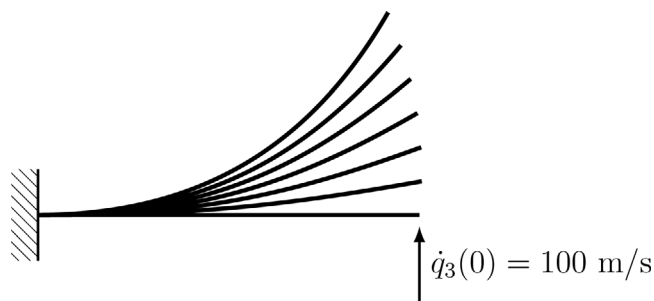
Fig. 27. One-dimensional cartoon illustrating 3D linear elastic cantilever plate problem setup.
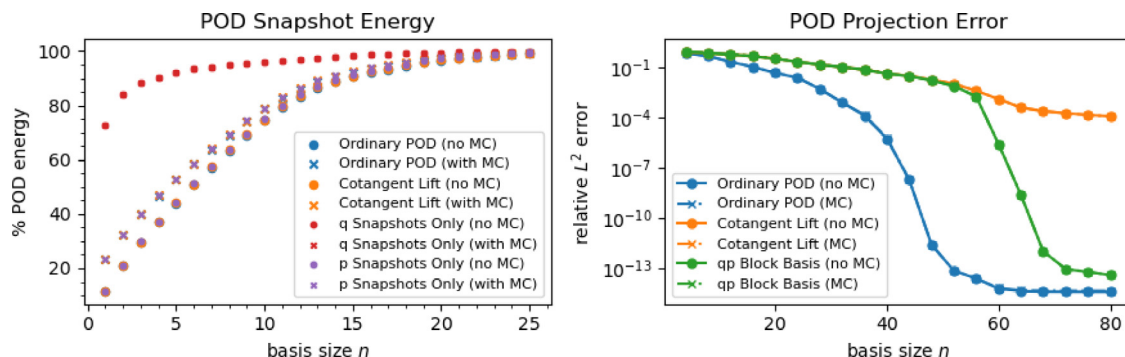

Fig. 28. POD snapshot energies (left) and projection errors (right) corresponding to the bases used in the 3D cantilever plate example. "MC" indicates that mean-centering of the snapshots was performed.

A one-dimensional cartoon illustrating the problem setup is shown in Fig. 27. The initial velocity perturbation (17) will cause the plate to vibrate and undergo a flapping motion, as shown in Fig. 26. As the plate vibrates, waves will form and propagate in all three coordinate directions within the plate. Assuming the plate is made of steel, the material parameters[3] are as follows: $E = 200$ GPa (Young's modulus), $\nu = 0.25$ (Poisson's ratio), and $\rho = 7800$ kg/m^3 (density).

To build the full order model from which our non-intrusive OpInf ROMs are constructed, we utilize the open-source[4] Albany-LCM multi-physics code base [88–90][5] and discretize the domain Ω with a uniform mesh of $20 \times 20 \times 3$ hexahedral elements. To generate snapshots, the FOM system (14) is advanced forward from time $t = 0$ to time $t = 2 \times 10^{-2}$ s using a symplectic implicit Newmark time-stepping scheme with parameters $\beta = 0.25$ and $\gamma = 0.5$, and time-step $\Delta t = 1.0 \times 10^{-4}$ s. Plots of the s_3 component of the displacement are shown at several different times in Fig. 26. The resulting 201 snapshots, each of length 5292, are used to build POD bases of varying sizes, from 4 to 100 POD modes. Fig. 28 shows the snapshot energies and reconstruction errors of these bases as a function of basis modes. Once the POD bases are constructed, several intrusive and non-intrusive ROMs are created and evaluated as discussed earlier in this manuscript. All ROMs are evaluated in the time-predictive regime, by integrating the governing system forward in time until $t = 0.1$ s (5× longer than the training time).

The results of this experiment are displayed in Fig. 29 and Table 1. Clearly, the ROMs are highly sensitive to the basis construction technique as well as the number of modes used. While the intrusive G-ROM and G-OpInf ROMs constructed with a block (q, p) basis yield the lowest minimum errors, they are highly volatile, exhibiting unpredictable behavior as basis modes are added. Conversely, the NC-H-OpInf ROM constructed with a block (q, p) basis and the C-H-OpInf ROM constructed with a cotangent lift basis exhibit some attempts at convergence,

---
[3] It is straightforward to calculate the Lamé coefficients appearing in (11) from the Young's modulus E and the Poisson ratio ν using the formulas $\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)}$ and $\mu = \frac{E}{2(1+\nu)}$.
[4] Albany-LCM is available on github at the following URL: https://github.com/sandialabs/LCM.
[5] For details on how to reproduce the results in this subsection, the reader is referred to Reproducibility and software availability section.
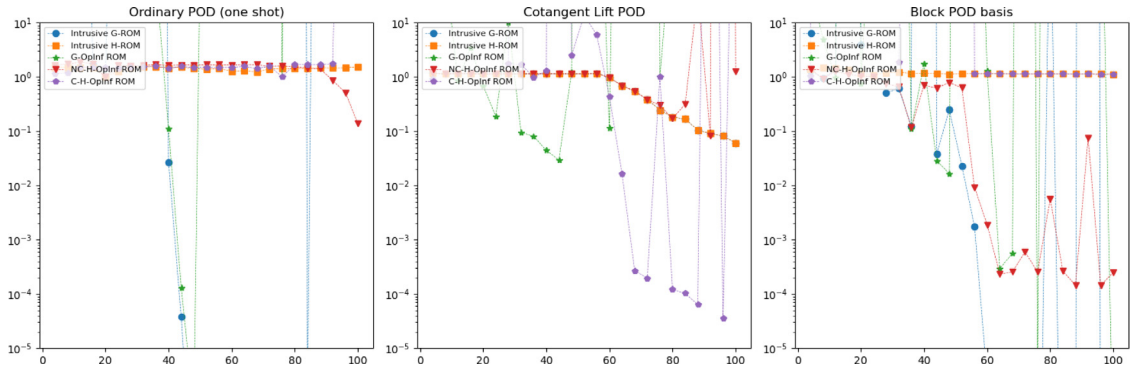
Fig. 29. Relative state errors as a function of basis modes for the ROMs in the 3D cantilever plate example (predictive case $T = 0.1$). "MC" indicates the use of a mean-centered reconstruction.

Table 1
Table corresponding to the plots in Fig. 29, corresponding to the 3D cantilever plate example and showing the ROM errors as a function of basis size. Dashes indicate lack of convergence.

POD basis	ROM type	Basis size n												
		4	12	20	28	36	44	52	60	68	76	84	92	100
Ordinary POD	Intrusive G-ROM	5.67×10^2	2.30×10^{10}	1.00	2.66×10^{10}	1.43×10^{24}	3.79×10^{-5}	–	–	1.72×10^{23}	5.56×10^{23}	1.25×10^{-6}	8.71×10^2	6.56×10^4
	Intrusive H-ROM	1.50	1.50	1.00	1.51	1.56	1.53	1.39	1.28	1.23	1.43	1.45	1.47	1.53
	G-OpInf ROM	8.58×10^{19}	1.51×10^{22}	–	2.75×10^5	6.55×10^1	1.29×10^{-4}	2.46×10^{10}	4.13×10^{34}	–	1.00	3.78×10^{146}	1.05×10^{44}	6.77×10^{55}
	NC-H-OpInf ROM	1.66	1.87	1.00	1.60	1.68	1.66	1.69	1.66	1.68	1.60	1.47	8.60×10^{-1}	1.38×10^{-1}
	C-H-OpInf ROM	1.16	1.48	1.51	1.49	1.59	1.57	1.50	1.51	1.43	1.00	1.70	1.74	5.09×10^3
Cotangent lift	Intrusive G-ROM	1.18	1.16	1.14	1.14	1.15	1.15	1.15	9.75×10^{-1}	5.51×10^{-1}	2.46×10^{-1}	1.68×10^{-1}	9.14×10^{-2}	6.10×10^{-2}
	Intrusive H-ROM	1.18	1.16	1.14	1.14	1.15	1.15	1.15	9.75×10^{-1}	5.51×10^{-1}	2.46×10^{-1}	1.68×10^{-1}	9.14×10^{-2}	6.10×10^{-2}
	G-OpInf ROM	4.83×10^1	8.50×10^{-1}	6.88×10^{-1}	1.00×10^1	8.07×10^{-2}	2.97×10^{-2}	1.90×10^{21}	1.15×10^{-1}	6.08×10^9	1.00	7.34×10^{35}	1.90×10^{60}	5.65×10^{121}
	NC-H-OpInf ROM	1.18	1.16	1.14	1.15	1.15	1.15	1.15	9.72×10^{-1}	5.50×10^{-1}	3.01×10^{-1}	1.00	8.21×10^{-2}	1.25
	C-H-OpInf ROM	9.17×10^{-1}	4.01×10^{90}	1.22×10^5	1.77	9.91×10^{-1}	3.15×10^{67}	1.39×10^1	4.34×10^{-1}	2.70×10^{-4}	1.00	1.03×10^{-4}	1.48×10^{14}	5.62×10^6
Block (q, p)	Intrusive G-ROM	1.32	1.39	3.95	5.12×10^{-1}	1.22×10^{-1}	3.79×10^{-2}	2.30×10^{-2}	2.92×10^{-6}	3.03×10^{-11}	2.94×10^{-11}	2.91×10^{-11}	1.29×10^{127}	1.75×10^{-11}
	Intrusive H-ROM	1.38	1.46	1.16	1.19	1.17	1.17	1.16	1.15	1.14	1.14	1.14	1.13	1.12
	G-OpInf ROM	4.74×10^1	9.65×10^{-1}	7.61×10^{-1}	5.46×10^{80}	1.12×10^{-1}	2.89×10^{-2}	1.93×10^{28}	1.31	5.59×10^{-4}	4.05×10^{-6}	4.10×10^{60}	2.45×10^{43}	5.95×10^{-8}
	NC-H-OpInf ROM	1.43	1.29	1.06	7.49×10^{-1}	1.17×10^{-1}	6.23×10^{-1}	6.32×10^{-1}	1.90×10^{-3}	2.55×10^{-4}	2.55×10^{-4}	2.63×10^{-4}	7.50×10^{-2}	2.50×10^{-4}
	C-H-OpInf ROM	1.05	1.68×10^{69}	4.29×10^{36}	–	1.55×10^{10}	1.44×10^5	5.04×10^{76}	1.15	1.14	1.14	1.14	1.14	1.13

although still with significant oscillations. It is interesting to note that the intrusive H-ROM represents a different extreme with all choices of bases: it is perfectly stable with the addition of modes, but exhibits $\mathcal{O}(1)$ errors unless a cotangent lift basis is used. It is further remarkable that the ROM errors in the reproductive case $T = 0.02$ (not pictured) are slightly lower (within one order of magnitude), but their stability properties remain unchanged.

Remark 5.2. While not pictured here, we have observed that intrusive ROMs based directly on the second-order Euler–Lagrange equations (14) do not suffer from the same degree of instability with respect to the addition of basis modes as those based on the first-order Hamiltonian formulation (15). This could be due to the fact that Galerkin projection of Lagrangian systems onto a reduced basis automatically respects energy conservation, which is not true in the Hamiltonian case, where an additional corrective projection is needed.

For another visualization, Fig. 30 shows plots of the displacement magnitude at the final time $t = 0.1$ s for the FOM (a) and various OpInf ROMs (b)–(d). Here, we showcase "best-case scenarios" for each ROM: (b) a G-OpInf ROM with 100 POD modes calculated via the (q, p)-block basis approach, (c) a C-H-OpInf ROM with 96 POD modes calculated via the cotangent lift basis, and (d) an NC-H-OpInf ROM with 96 POD modes calculated via the (q, p)-block basis approach. The reader can observe that each ROM is capable of producing solutions which are visually indistinguishable from the FOM solution (see subplots (b)–(d)), although their error distribution patterns are quite different (see subplots (e)–(g)). We emphasize that, while the G-OpInf ROM is the most accurate, it is also by far the most sensitive to the size of the reduced basis (see e.g. Fig. 29): there is no visible trend in terms of the basis size, in contrast with the NC-H-OpInf and C-H-OpInf ROMs which are still volatile but roughly decreasing. Additionally, it is likely that the results seen here could be improved somewhat by regularizing the OpInf problem in some way; since the choice of regularization technique is a non-obvious matter which is currently under active investigation (e.g., [91]), this is left for future work.
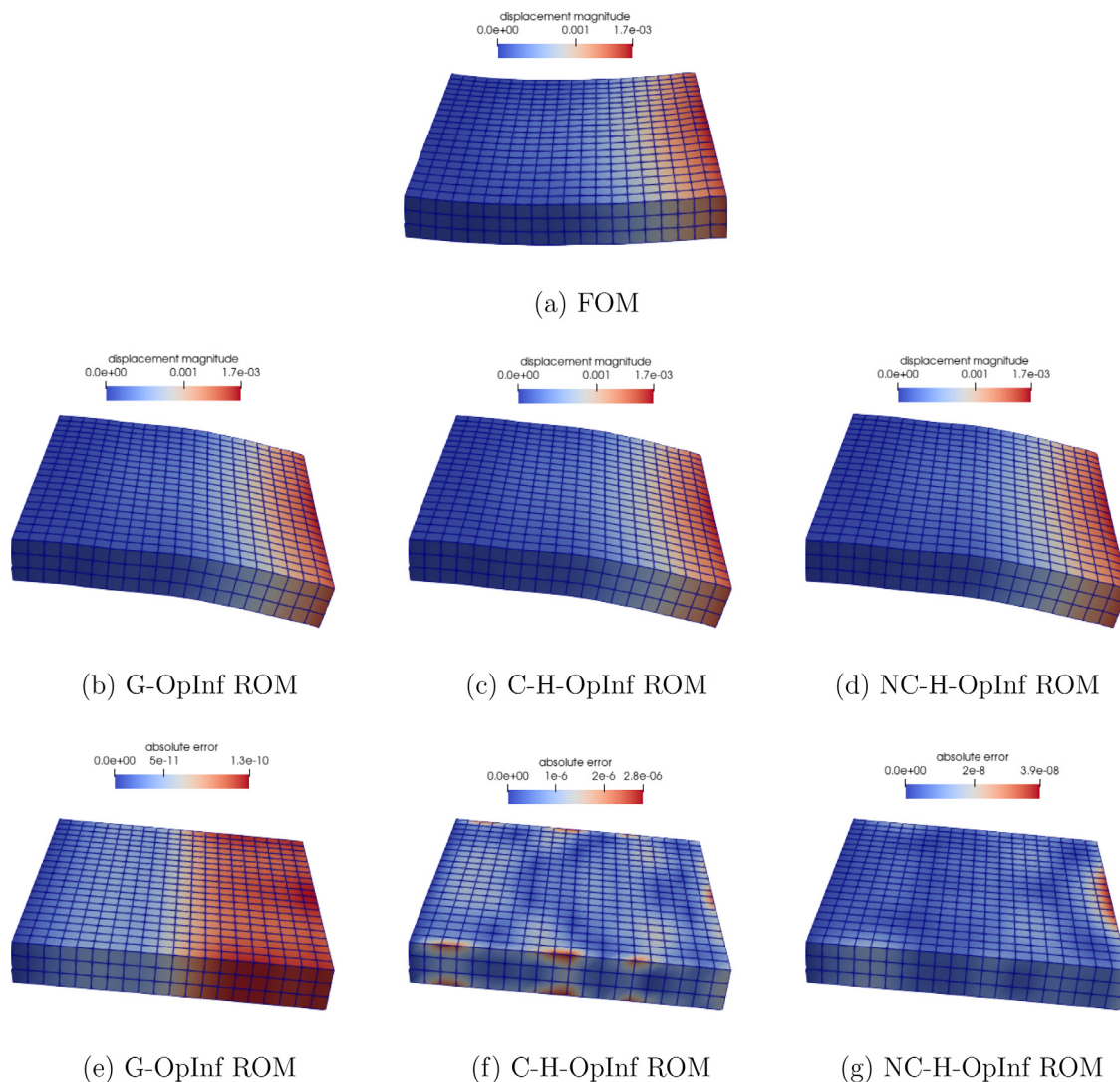
(a) FOM

(b) G-OpInf ROM

(c) C-H-OpInf ROM

(d) NC-H-OpInf ROM

(e) G-OpInf ROM

(f) C-H-OpInf ROM

(g) NC-H-OpInf ROM

Fig. 30. "Best case" plots of the displacement magnitude at the final time $t = 0.1$ s for the FOM (a) and various OpInf ROMs (b)–(d) for the 3D linear elastic cantilever plate problem. Subplots (e)–(g) show the spatial distribution of the absolute errors in the displacement magnitude for the various ROMs evaluated, again at the final time $t = 0.1$ s.

To test conservation, Fig. 31 plots the errors in the value of the Hamiltonian. Again, it can be seen that the C-H-OpInf and NC-H-OpInf preserve this quantity much better than the G-OpInf ROM, even in cases where the G-OpInf ROM is more accurate. Unsurprisingly, the conservation properties of the intrusive H-ROM are still superior in all cases, although it is remarkable that this does not always translate to better accuracy in the ROM solution. This could be due to the fact that the H-ROMs require an additional projection step onto the column space of $\mathbf{U}$, limiting their accuracy in order to gain exact property preservation.

6. Conclusions and future work

Two gray-box operator inference (OpInf) methods for the nonintrusive model reduction of Hamiltonian dynamical systems have been introduced, and their utility has been demonstrated on several canonical and noncanonical benchmarks. Being provably convergent to their intrusive counterparts in the limit of infinite data, these OpInf ROMs are shown to recover desirable properties of carefully built intrusive Hamiltonian ROMs such as improved
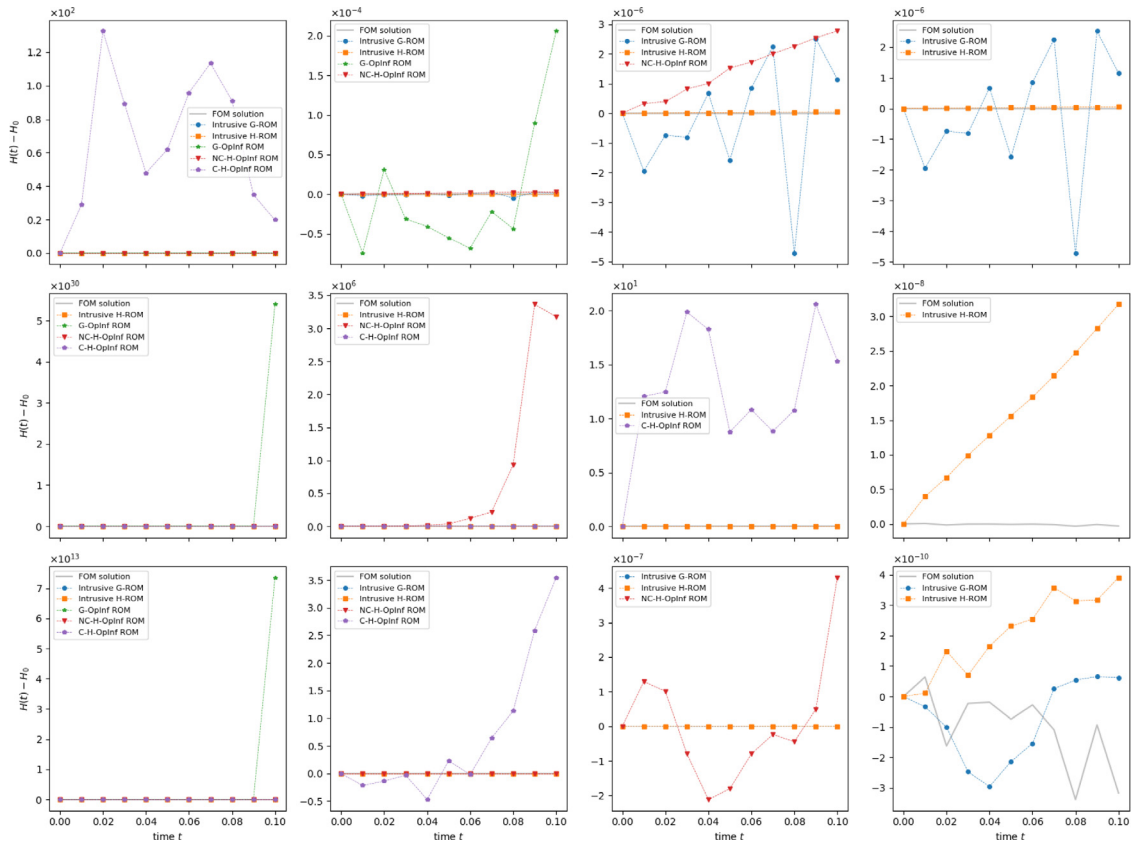
Fig. 31. Plots of the error in the Hamiltonian for three different simulations corresponding to the 3D cantilever beam example. (Top) Ordinary POD basis, $n = 48$ modes; (Middle) Cotangent Lift POD basis, $n = 84$ modes; (Bottom) Block (q, p) POD basis, $n = 92$ modes.

energy conservation without requiring access to FOM simulation code, making them flexible to deploy and leading to improved performance over generic OpInf techniques in reproductive and predictive problems. Moreover, the technique introduced here has been shown to strictly generalize previous state-of-the-art work on Hamiltonian OpInf methods, reducing to it when the Hamiltonian system in question is canonical, the basis used is a cotangent lift, and the operator to be inferred is block diagonal.

Despite the improvements made here, there are plenty of avenues for future work in the area of Hamiltonian model reduction. First, the gray-box requirement that the nonlinear part of the Hamiltonian system is known can be feasibly removed when this nonlinearity is polynomial, making the Hamiltonian OpInf methods described potentially black-box in this case. Similarly, it would be interesting to apply this technique to systems which have a known conserved quantity but no known Hamiltonian structure, to see if the NC-H-OpInf ROM which preserves this quantity is more accurate and predictively useful than a generic OpInf ROM. Additionally, it is clear that all structure-preserving ROM techniques to date, intrusive or OpInf, are quite sensitive to basis size when problems become large with complex dynamics. It would be useful to have stabilized techniques which produce ROMs with more predictable convergence behavior and which do not destroy the delicate mathematical structure important for long-term behavior of the FOM system. Finally, it would be interesting to extend the techniques mentioned here to quadratic POD bases as well as more general Lie–Poisson variational problems.

Declaration of competing interest

Data availability

The data necessary for reproducing the experiments in this paper can be downloaded from the following github repository: https://github.com/ikalash/HamiltonianOpInf.

Acknowledgments

Reproducibility and software availability

The numerical results presented in Section 5.5 were generated by running the Albany-LCM open-source HPC code, available for download on github at the following URL: https://github.com/sandialabs/LCM. The Albany-LCM code has a strong dependency on Trilinos, available at: https://github.com/trilinos/Trilinos. The following shas for Albany-LCM and Trilinos were used in generating the results herein: `79d3a68bd6176c80ae10e693cd15b1040f6 be10f` and `322132d613777d48b85d70ed95c0ff4a07c8aed0`, respectively. To ensure transparency and reproducibility, we have made available the Albany-LCM input files needed to reproduce our results, as well as configure scripts for Albany-LCM and Trilinos. These input files can be downloaded from the following github repository: https://github.com/ikalash/HamiltonianOpInf. The handwritten Python files and scripts for reproducing the results in Section 5 can be found in this same repository.

Appendix

A.1. Kronecker products and vectorization

We briefly recall some properties of the Kronecker product which are necessary for the results in the body. Interested readers can find more details in, e.g., [92]. Let $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{p \times q}$. The Kronecker product $\mathbf{A} \otimes \mathbf{B} \in \mathbb{R}^{mp \times nq}$ is then the matrix of size $mp \times nq$ whose i, jth block (of size $p \times q$) is given by $(\mathbf{A} \otimes \mathbf{B})^i_j = a^i_j \mathbf{B}$. It is straightforward to show that $\otimes$ is the matricization of the usual tensor product when expressed with respect to a lexicographical ordering of the standard bases for $\mathbb{R}^n \otimes \mathbb{R}^q$ and $\mathbb{R}^m \otimes \mathbb{R}^p$, since $(\mathbf{A} \otimes \mathbf{B}) (\mathbf{x} \otimes \mathbf{y}) = \mathbf{A}\mathbf{x} \otimes \mathbf{B}\mathbf{y}$ for any $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^q$. Moreover, there is a linear vectorization operator "vec" which stacks the columns of a matrix into a long vector, i.e. $A_{ij} = (\text{vec}\,\mathbf{A})_{m(j-1)+i}$. Since vectorization is obviously invertible, this allows for the following computationally convenient reformulation of linear systems with matrix unknowns.

Theorem A.1 (*Vec Trick*). $\text{vec}\,(\mathbf{A}\mathbf{X}\mathbf{B}) = (\mathbf{B}^\mathsf{T} \otimes \mathbf{A})\,\text{vec}\,\mathbf{X}.$

Proof. Let $\mathbf{a}_i, \mathbf{x}_i, \mathbf{b}_i$ denote the ith column of $\mathbf{A}, \mathbf{X}, \mathbf{B}$ respectively. Then, the ith column of $\mathbf{A}\mathbf{X}\mathbf{B}$ is

$$(\mathbf{A}\mathbf{X}\mathbf{B})_i = \mathbf{A}\mathbf{X}\mathbf{b}_i = \mathbf{A}b^j_i \mathbf{x}_j = \left(b^j_i \mathbf{A}\right)\mathbf{x}_j = \left(\mathbf{b}^\mathsf{T}_i \otimes \mathbf{A}\right)\text{vec}\,\mathbf{X}.$$

The conclusion now follows by stacking columns. $\square$

There is also a very concrete (but rather inefficient) way to obtain the transposition matrix $\mathbf{K}$ satisfying $\operatorname{vec}\mathbf{X}^{\mathsf{T}} = \mathbf{K}\operatorname{vec}\mathbf{X}$. While this is true generally for $m \times n$ matrices $\mathbf{X}$, we state the result for square matrices for ease of notation.

Proposition A.1. *Let $\mathbf{E}_{ij} = \mathbf{e}_i\mathbf{e}_j^{\mathsf{T}}$ denote the ijth basis vector for the matrix space $\mathbb{R}^{n\times n}$. Then, we have that*

$$\mathbf{K} = \sum_{ij} \mathbf{E}_{ji}^{\mathsf{T}} \otimes \mathbf{E}_{ji},$$

satisfies $\operatorname{vec}\mathbf{X}^{\mathsf{T}} = \mathbf{K}\operatorname{vec}\mathbf{X}$.

Proof. Given $\mathbf{X} \in \mathbb{R}^{n\times n}$, it follows by the vec trick that

$$\mathbf{K}\operatorname{vec}\mathbf{X} = \left(\sum \mathbf{E}_{ji}^{\mathsf{T}} \otimes \mathbf{E}_{ji}\right)\operatorname{vec}\mathbf{X} = \operatorname{vec}\left(\sum \mathbf{E}_{ji}\mathbf{X}\mathbf{E}_{ji}\right)$$
$$= \operatorname{vec}\left(\sum \mathbf{e}_j x_{ij}\mathbf{e}_i^{\mathsf{T}}\right) = \operatorname{vec}\left(\sum x_{ij}\mathbf{E}_{ji}\right) = \operatorname{vec}\mathbf{X}^{\mathsf{T}}. \quad \square$$

More practically, the following pseudocode is used to generate a sparse matrix representing $\mathbf{K}$.

Algorithm 3 Building the commutation matrix $\mathbf{K}$

Input: Integers $m, n > 0$.
Output: Sparse matrix $\mathbf{K} \in \mathbb{Z}^{mn\times mn}$ satisfying $\operatorname{vec}\mathbf{X}^{\mathsf{T}} = \mathbf{K}\operatorname{vec}\mathbf{X}$ for all $\mathbf{X} \in \mathbb{R}^{m\times n}$.
  1: Let $row = \{1, 2, ..., mn\} \in \mathbb{Z}^{mn}$ be the vector of row indices.
  2: Let $row' \in \mathbb{Z}^{m\times n}$ be defined by reshaping row column-wise.
  3: Let $col \in \mathbb{Z}^{mn}$, the list of column indices, be the row-wise flattening of row'.
  4: **return** Sparse matrix $\mathbf{K}$ with indices (row, col) and entries $\{1, ..., 1\} \in \mathbb{Z}^{mn}$.

A.2. Proofs of results

Here we provide omitted proofs for the results in the body. Note that Einstein summation is assumed throughout, so that any tensor index appearing both "up" and "down" in an expression is implicitly summed over its range.

Proof of Proposition 2.1. This is a straightforward consequence of the fact that OpInf of size n decouples into n^2 scalar minimization problems. To see this, notice that if $\hat{\mathbf{D}}$ solves the OpInf problem of size n and $1 \le i', j' \le n' < K \le n$, $1 \le k \le n$, then

$$\left(\hat{\mathbf{X}}_t\hat{\mathbf{X}}^{\mathsf{T}}\right)_{j'}^{i'} = \hat{D}_k^{i'}\left(\hat{\mathbf{X}}\hat{\mathbf{X}}^{\mathsf{T}}\right)_{j'}^{k} = \hat{D}_{k'}^{i'}\left(\hat{\mathbf{X}}\hat{\mathbf{X}}^{\mathsf{T}}\right)_{j'}^{k'} + \hat{D}_K^{i'}\left(\hat{\mathbf{X}}\hat{\mathbf{X}}^{\mathsf{T}}\right)_{j'}^{K}$$
$$= \hat{D}_{k'}^{i'}\delta_{j'}^{k'}\sigma_{j'}^2 + \hat{D}_K^{i'}\delta_{j'}^K\sigma_{j'}^2 = \hat{D}_{j'}^{i'}\sigma_{j'}^2,$$

where δ denotes the Kronecker delta tensor and the first equality of the second line follows from the fact that, for all $1 \le i, j \le n$,

$$\left(\hat{\mathbf{X}}\hat{\mathbf{X}}^{\mathsf{T}}\right)_j^i = \left\langle\mathbf{X}^{\mathsf{T}}\mathbf{u}_i, \mathbf{X}^{\mathsf{T}}\mathbf{u}_j\right\rangle = \left\langle\mathbf{e}_i, \mathbf{U}^{\mathsf{T}}\mathbf{X}\mathbf{X}^{\mathsf{T}}\mathbf{U}\mathbf{e}_j\right\rangle = \left\langle\mathbf{e}_i, \mathbf{\Sigma}^2\mathbf{e}_j\right\rangle = \delta_j^i\sigma_j^2.$$

Therefore, the minimization problem for each component $\hat{D}_j^i$ has the solution (note the sum on k),

$$\underset{\hat{D}_j^i\in\mathbb{R}}{\operatorname{argmin}}\left|\mathbf{u}_i^{\mathsf{T}}\mathbf{X}_t - \sum_k \hat{D}_k^i\mathbf{u}_k^{\mathsf{T}}\mathbf{X}\right|^2 = \frac{\mathbf{u}_i^{\mathsf{T}}\mathbf{X}_t\mathbf{X}^{\mathsf{T}}\mathbf{u}_j}{\sigma_j^2},$$

showing that each entry of $\hat{\mathbf{D}}$ depends only on the indices i, j. Therefore, the solution $\hat{\mathbf{D}}'$ to the OpInf problem of size $n' < n$ can be extracted from $\hat{\mathbf{D}}$ by extracting the top-left $n' \times n'$ submatrix, as desired. $\quad \square$

Proof of Proposition 3.1. Suppose $\hat{\mathbf{D}}$ is the solution hypothesized in the statement of the Proposition. Then, it follows that

$$\hat{\mathbf{A}}^\mathsf{T}\hat{\mathbf{C}}\hat{\mathbf{B}}^\mathsf{T} \pm \hat{\mathbf{B}}\hat{\mathbf{C}}^\mathsf{T}\hat{\mathbf{A}} = \hat{\mathbf{A}}^\mathsf{T}\hat{\mathbf{A}}\hat{\mathbf{D}}\hat{\mathbf{B}}\hat{\mathbf{B}}^\mathsf{T} \pm \hat{\mathbf{B}}\hat{\mathbf{B}}^\mathsf{T}\hat{\mathbf{D}}^\mathsf{T}\hat{\mathbf{A}}^\mathsf{T}\hat{\mathbf{A}}$$
$$= \hat{\mathbf{A}}^\mathsf{T}\hat{\mathbf{A}}\hat{\mathbf{D}}\hat{\mathbf{B}}\hat{\mathbf{B}}^\mathsf{T} + \hat{\mathbf{B}}\hat{\mathbf{B}}^\mathsf{T}\hat{\mathbf{D}}\hat{\mathbf{A}}^\mathsf{T}\hat{\mathbf{A}}$$

Now, for any $1 \le i, j \le n$, notice that

$$\left(\hat{\mathbf{A}}^\mathsf{T}\hat{\mathbf{C}}\hat{\mathbf{B}}^\mathsf{T} \pm \hat{\mathbf{B}}\hat{\mathbf{C}}^\mathsf{T}\hat{\mathbf{A}}\right)^i_j = \hat{A}^i_i \mathbf{u}_i^\mathsf{T}\mathbf{C}\mathbf{B}^\mathsf{T}\mathbf{u}_j \pm \mathbf{u}_i^\mathsf{T}\mathbf{B}\mathbf{C}^\mathsf{T}\mathbf{u}_j A^j_j.$$

Therefore, the ijth entry of the left-hand side of the optimality condition for $\hat{\mathbf{D}}$ depends only on the basis vectors $\mathbf{u}_i, \mathbf{u}_j$, and we have

$$\hat{\mathbf{A}}'^\mathsf{T}\hat{\mathbf{C}}'\hat{\mathbf{B}}'^\mathsf{T} \pm \hat{\mathbf{B}}'\hat{\mathbf{C}}'^\mathsf{T}\hat{\mathbf{A}}' = \left(\hat{\mathbf{A}}^\mathsf{T}\hat{\mathbf{C}}\hat{\mathbf{B}}^\mathsf{T} \pm \hat{\mathbf{B}}\hat{\mathbf{C}}^\mathsf{T}\hat{\mathbf{A}}\right)'.$$

Similarly, letting $1 \le i', j', k', l' \le n'$ and $1 \le k, l \le n$, it follows that

$$\left(\hat{\mathbf{A}}^\mathsf{T}\hat{\mathbf{C}}\hat{\mathbf{B}}^\mathsf{T} \pm \hat{\mathbf{B}}\hat{\mathbf{C}}^\mathsf{T}\hat{\mathbf{A}}\right)^{i'}_{j'} = \left(\hat{\mathbf{A}}^\mathsf{T}\hat{\mathbf{A}}\right)^{i'}_k \hat{D}^k_l \left(\hat{\mathbf{B}}\hat{\mathbf{B}}^\mathsf{T}\right)^l_{j'} + \left(\hat{\mathbf{B}}\hat{\mathbf{B}}^\mathsf{T}\right)^{i'}_k \hat{D}^k_l \left(\hat{\mathbf{A}}^\mathsf{T}\hat{\mathbf{A}}\right)^l_{j'}$$
$$= \left(\hat{A}^{i'}_{k'}\right)^2 D^{k'}_{l'} \left(\hat{\mathbf{B}}\hat{\mathbf{B}}^\mathsf{T}\right)^{l'}_{j'} + \left(\hat{\mathbf{B}}\hat{\mathbf{B}}^\mathsf{T}\right)^{i'}_{k'} D^{k'}_{l'} \left(\hat{A}^{l'}_{j'}\right)^2,$$

where the second line uses the fact that $\left(\hat{\mathbf{A}}^\mathsf{T}\hat{\mathbf{A}}\right)^{i'}_K = \left(\hat{\mathbf{B}}\hat{\mathbf{B}}^\mathsf{T}\right)^L_{j'} = \left(\hat{\mathbf{B}}\hat{\mathbf{B}}^\mathsf{T}\right)^{i'}_K = \left(\hat{\mathbf{A}}^\mathsf{T}\hat{\mathbf{A}}\right)^L_{j'}$ for all $n' < K, L \le n$. Putting these computations together, this shows that the truncation $\hat{\mathbf{D}}'$ of $\hat{\mathbf{D}}$ satisfies

$$\hat{\mathbf{A}}'^\mathsf{T}\hat{\mathbf{C}}'\hat{\mathbf{B}}'^\mathsf{T} \pm \hat{\mathbf{B}}'\hat{\mathbf{C}}'^\mathsf{T}\hat{\mathbf{A}}' = \hat{\mathbf{A}}'^\mathsf{T}\hat{\mathbf{A}}'\hat{\mathbf{D}}'\bar{\mathbf{B}}\bar{\mathbf{B}}^\mathsf{T} + \bar{\mathbf{B}}\bar{\mathbf{B}}^\mathsf{T}\hat{\mathbf{D}}'\hat{\mathbf{A}}'^\mathsf{T}\hat{\mathbf{A}}',$$

showing that $\hat{\mathbf{D}}'$ is the desired minimizer. $\square$

A.3. Second Hamiltonian formulation of KdV

Another Hamiltonian formulation of the KdV equation is given by the data

$$H(x) = \frac{1}{2}\int_0^l x^2\,ds, \qquad L(x) = \frac{\alpha}{3}\left(x\partial_s + \partial_s(x\cdot)\right) + \rho\partial_s + \nu\partial_{sss},$$

where $\partial_s(x\cdot)y = \partial_s(xy)$. Choosing $\mathbf{A}$ to be the central difference discretization of ∂_s (this was $\mathbf{L}$ in the first formulation) leads to the skew-symmetric discrete operator

$$\mathbf{L}(\mathbf{x}) = \frac{\alpha}{3}\left(\text{Diag}(\mathbf{x})\mathbf{A} + \mathbf{A}\,\text{Diag}(\mathbf{x})\right) + \rho\mathbf{A} + \nu\mathbf{E},$$

where $\mathbf{E}$ is the pentadiagonal circulant matrix representing the central difference discretization of ∂_{sss}, i.e.

$$\mathbf{E} = \frac{1}{2(\Delta x)^3}\begin{pmatrix} 0 & -2 & 1 & & -1 & 2 \\ 2 & \ddots & \ddots & \ddots & & -1 \\ -1 & \ddots & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & \ddots & 1 \\ 1 & & \ddots & \ddots & \ddots & -2 \\ -2 & 1 & & -1 & 2 & 0 \end{pmatrix}.$$

This implies the second Hamiltonian formulation of the KdV system,

$$\dot{\mathbf{x}} = \mathbf{L}(\mathbf{x})\nabla H(\mathbf{x}) = \frac{\alpha}{3}\left(\text{Diag}(\mathbf{x})\mathbf{A}\mathbf{x} + \mathbf{A}\,\text{Diag}(\mathbf{x})\mathbf{x}\right) + \rho\mathbf{A}\mathbf{x} + \nu\mathbf{E}\mathbf{x},$$

which is integrated via AVF to yield the discrete system,

$$\frac{\mathbf{x}^{k+1} - \mathbf{x}^k}{\Delta t} = \mathbf{L}\left(\mathbf{x}^{k+\frac{1}{2}}\right)\mathbf{x}^{k+\frac{1}{2}}.$$
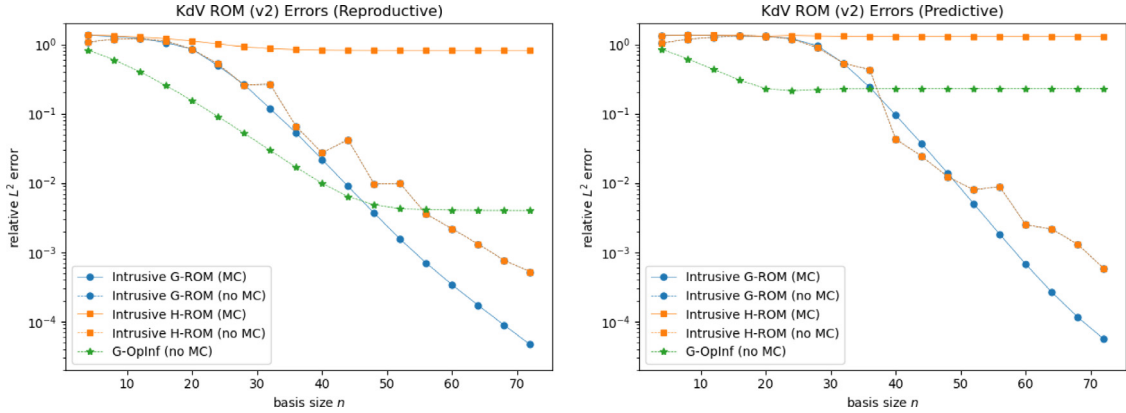
Fig. 32. Relative state errors as a function of basis modes for the ROMs in the KdV equation (V2) example. Left: reproductive case ($T = 20$). Right: predictive case ($T = 100$). "MC" indicates the use of a mean-centered reconstruction.

This leads to the Δt-normalized residual and Jacobian functions,

$$R^k \left(\mathbf{x}^{k+1}\right) = \mathbf{x}^{k+1} - \mathbf{x}^k - \Delta t\, \mathbf{L}\left(\mathbf{x}^{k+\frac{1}{2}}\right) \mathbf{x}^{k+\frac{1}{2}},$$

$$J^k \left(\mathbf{x}^{k+1}\right) = \mathbf{I} - \frac{\Delta t}{2} \left[\frac{2\alpha}{3} \left(\mathrm{Diag}\left(\mathbf{x}^{k+\frac{1}{2}}\right) \mathbf{A} + \mathbf{A}\, \mathrm{Diag}\left(\mathbf{x}^{k+\frac{1}{2}}\right) \right) + \rho \mathbf{A} + \nu \mathbf{E} \right],$$

which are solvable with Newton iterations.

With this, intrusive Galerkin and Hamiltonian ROMs can then be constructed as before. On the other hand, notice that AVF evaluates $\mathbf{L}$ at the midpoint of the discrete trajectory, meaning that Galerkin projection and discretization with AVF no longer commute, since $\mathbf{U}^{\mathsf{T}}\mathbf{L}$ is not a Poisson matrix. However, letting $\tilde{\mathbf{x}} = \mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}}$, Galerkin projection after AVF yields

$$
\begin{aligned}
\frac{\hat{\mathbf{x}}^{k+1} - \hat{\mathbf{x}}^k}{\Delta t} &= \mathbf{U}^{\mathsf{T}}\mathbf{L}\left(\tilde{\mathbf{x}}^{k+\frac{1}{2}}\right) \tilde{\mathbf{x}}^{k+\frac{1}{2}} \\
&= \left(\mathbf{U}^{\mathsf{T}}\mathbf{L}\left(\mathbf{x}_0\right) + \frac{\alpha}{3}\left(\mathbf{U}^{\mathsf{T}}\,\mathrm{Diag}\left(\mathbf{U}\hat{\mathbf{x}}^{k+\frac{1}{2}}\right) \mathbf{A} + \mathbf{U}^{\mathsf{T}}\mathbf{A}\,\mathrm{Diag}\left(\mathbf{U}\hat{\mathbf{x}}^{k+\frac{1}{2}}\right) \right) \right) \left(\mathbf{x}_0 + \mathbf{U}\hat{\mathbf{x}}^{k+\frac{1}{2}} \right) \\
&:= \mathbf{U}^{\mathsf{T}}\mathbf{L}\left(\mathbf{x}_0\right)\mathbf{x}_0 + \left(\mathbf{T}\left(\hat{\mathbf{x}}^{k+\frac{1}{2}}\right)\mathbf{x}_0 + \mathbf{U}^{\mathsf{T}}\mathbf{L}\left(\mathbf{x}_0\right)\mathbf{U}\hat{\mathbf{x}}^{k+\frac{1}{2}} \right) + \hat{\mathbf{T}}\left(\hat{\mathbf{x}}^{k+\frac{1}{2}}\right)\hat{\mathbf{x}}^{k+\frac{1}{2}} \\
&:= \hat{\mathbf{c}} + \hat{\mathbf{C}}\hat{\mathbf{x}}^{k+\frac{1}{2}} + \hat{\mathbf{T}}\left(\hat{\mathbf{x}}^{k+\frac{1}{2}}\right)\hat{\mathbf{x}}^{k+\frac{1}{2}},
\end{aligned}
$$

where $\mathbf{T}, \hat{\mathbf{T}}$ are precomputable order 3 tensors given component-wise by $T^a_{jc} = (\alpha/3)U^a_i \left(U^i_c + U_{jc} \right) A^i_j$ and $\hat{T}^a_{bc} = (\alpha/3)U^a_i \left(U^i_c + U_{jc} \right) A^i_j U^j_b$. Now, a Hamiltonian ROM can be computed in the same way: applying AVF before Hamiltonian projection, it follows that

$$
\begin{aligned}
\frac{\hat{\mathbf{x}}^{k+1} - \hat{\mathbf{x}}^k}{\Delta t} &= \mathbf{U}^{\mathsf{T}}\mathbf{L}\left(\tilde{\mathbf{x}}^{k+\frac{1}{2}}\right) \mathbf{U}\mathbf{U}^{\mathsf{T}}\tilde{\mathbf{x}}^{k+\frac{1}{2}} \\
&= \left(\mathbf{U}^{\mathsf{T}}\mathbf{L}\left(\mathbf{x}_0\right)\mathbf{U} + \frac{\alpha}{3}\left(\mathbf{U}^{\mathsf{T}}\,\mathrm{Diag}\left(\mathbf{U}\hat{\mathbf{x}}^{k+\frac{1}{2}}\right) \mathbf{A}\mathbf{U} + \mathbf{U}^{\mathsf{T}}\mathbf{A}\,\mathrm{Diag}\left(\mathbf{U}\hat{\mathbf{x}}^{k+\frac{1}{2}}\right)\mathbf{U} \right) \right) \left(\mathbf{U}^{\mathsf{T}}\mathbf{x}_0 + \hat{\mathbf{x}}^{k+\frac{1}{2}} \right) \\
&= \mathbf{U}^{\mathsf{T}}\mathbf{L}\left(\mathbf{x}_0\right)\mathbf{U}\mathbf{U}^{\mathsf{T}}\mathbf{x}_0 + \left(\hat{\mathbf{T}}\left(\hat{\mathbf{x}}^{k+\frac{1}{2}}\right)\mathbf{U}^{\mathsf{T}}\mathbf{x}_0 + \mathbf{U}^{\mathsf{T}}\mathbf{L}\left(\mathbf{x}_0\right)\mathbf{U}\hat{\mathbf{x}}^{k+\frac{1}{2}} \right) + \hat{\mathbf{T}}\left(\hat{\mathbf{x}}^{k+\frac{1}{2}}\right)\hat{\mathbf{x}}^{k+\frac{1}{2}} \\
&:= \hat{\mathbf{c}} + \hat{\mathbf{C}}\hat{\mathbf{x}}^{k+\frac{1}{2}} + \hat{\mathbf{T}}\left(\hat{\mathbf{x}}^{k+\frac{1}{2}}\right)\hat{\mathbf{x}}^{k+\frac{1}{2}},
\end{aligned}
$$

where the tensor $\hat{T}$ is identical to before. In either case, these equations are easily solved with Newton iterations, as explained in Section 5.3.

Remark A.2. It is interesting to note that $\nabla H(\mathbf{x}) = \mathbf{x}$ in this formulation, so that its matrix representation $\mathbf{A} = \mathbf{I}$. This has the effect of equalizing the (non mean-centered) H-ROM and G-ROM, since $\hat{\mathbf{L}}\hat{\mathbf{A}} = \widehat{\mathbf{L}\mathbf{A}}$.
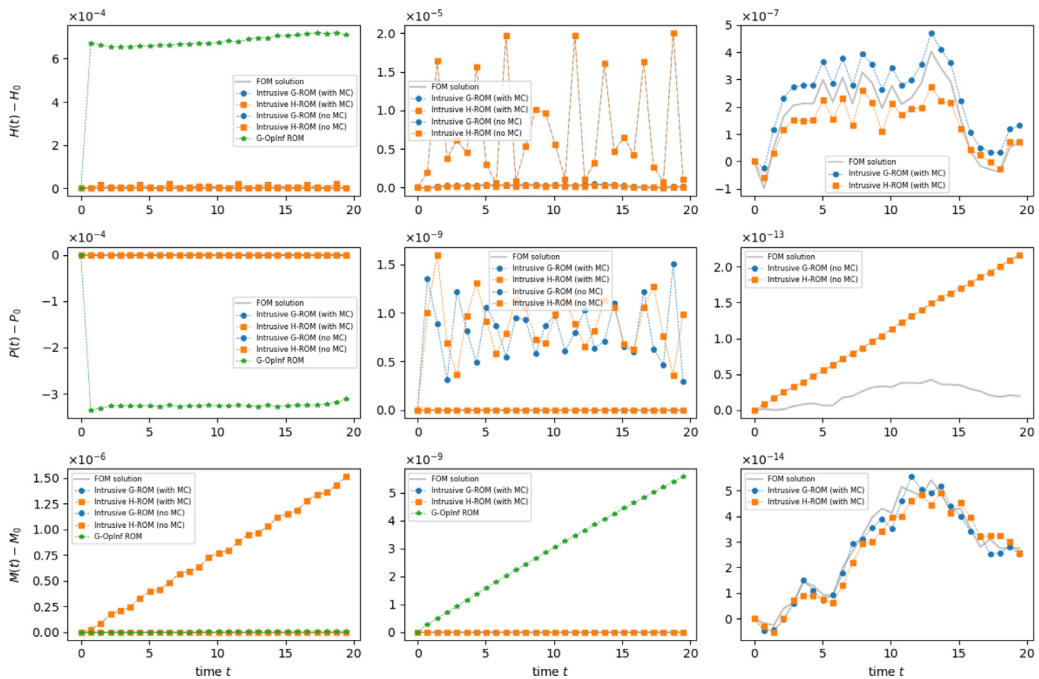
Fig. 33. Errors in conserved quantities for the (mean-centered) ROMs in the KdV (v2) equation example in the predictive case ($T = 20$) when using a POD basis with $n = 72$ modes.

Figs. 32 and 33 show the results of this procedure, alongside a linear G-OpInf ROM for comparison (c.f. Section 2.3). The experimental parameters are identical to those in Section 5.3. It is remarkable that the mean-centered H-ROM does not perform well in this case, despite conserving the first three invariant quantities as well as the mean-centered G-ROM. Note that the naming convention in Fig. 33 follows that of Fig. 20, despite the fact that P is now the Hamiltonian functional.

References

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

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

[3] C. Rowley, T. Colonius, R. Murray, Model reduction for compressible flows using POD and Galerkin projection, Physica D 189 (2004) 115–129.

[4] P. Schmid, Dynamic mode decomposition of numerical and experimental data, J. Fluid Mech. 656 (2010) 5–28.

[5] C. Rowley, Model reduction for fluids using balanced proper orthogonal decomposition, Int. J. Bifurcation Chaos 15 (3) (2005) 997–1013.

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

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

[8] B. Moore, Principal component analysis in linear systems: Controllability, observability, and model reduction, IEEE Trans. Automat. Control 26 (1) (1981).

[9] G. Rozza, Reduced basis approximation and error bounds for potential flows in parametrized geometries, Commun. Comput. Phys. 9 (1) (2011) 1–48.

[10] K. Veroy, A. Patera, Certified real-time solution of the parametrized steady incompressible Navier-Stokes equations: rigorous reduced-bases *a posteriori* error bounds, J. Numer. Meth. Fluids 47 (2005) 773–788.

[11] K. Lee, K.T. Carlberg, Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders, J. Comput. Phys. 404 (2020) 108973.

[12] S. Fresca, G. Gobat, P. Fedeli, A. Frangi, A. Manzoni, Deep learning-based reduced order models for the real-time simulation of the nonlinear dynamics of microstructures, Internat. J. Numer. Methods Engrg. 123 (20) (2022) 4749–4777, http://dx.doi.org/10.1002/nme.7054, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.7054. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.7054.

[13] Y. Kim, Y. Choi, D. Widemann, T. Zohdi, A fast and accurate physics-informed neural network reduced order model with shallow masked autoencoder, J. Comput. Phys. 451 (2022) 110841, http://dx.doi.org/10.1016/j.jcp.2021.110841, URL https://www.sciencedirect.com/science/article/pii/S0021999121007361.

[14] A. Gruber, M. Gunzburger, L. Ju, Z. Wang, A comparison of neural network architectures for data-driven reduced-order modeling, Comput. Methods Appl. Mech. Engrg. 393 (2022) 114764, http://dx.doi.org/10.1016/j.cma.2022.114764, URL https://www.sciencedirect.com/science/article/pii/S004578252200113X.

[15] F. Romor, G. Stabile, G. Rozza, Non-linear manifold reduced-order models with convolutional autoencoders and reduced over-collocation method, J. Sci. Comput. 94 (74) (2023) 1–39.

[16] H. Sharma, H. Mu, P. Buchfink, R. Geelen, S. Glas, B. Kramer, Symplectic model reduction of Hamiltonian systems using data-driven quadratic manifolds, 2023, arXiv:2305.15490.

[17] J.L. Barnett, C. Farhat, Y. Maday, Neural-network-augmented projection-based model order reduction for mitigating the Kolmogorov barrier to reducibility of CFD models, 2022, arXiv:2212.08939.

[18] J. Barnett, C. Farhat, Quadratic approximation manifold for mitigating the Kolmogorov barrier in nonlinear projection-based model order reduction, J. Comput. Phys. 464 (2022) 111348, http://dx.doi.org/10.1016/j.jcp.2022.111348, URL https://www.sciencedirect.com/science/article/pii/S0021999122004107.

[19] R. Geelen, S. Wright, K. Willcox, Operator inference for non-intrusive model reduction with quadratic manifolds, Comput. Methods Appl. Mech. Engrg. 403 (2023) 115717, http://dx.doi.org/10.1016/j.cma.2022.115717, URL https://www.sciencedirect.com/science/article/pii/S0045782522006727.

[20] A. Pinkus, n-Widths in Approximation Theory, in: A Series of Modern Surveys in Mathematics, vol. 7, 1985.

[21] B. Peherstorfer, K. Willcox, Data-driven operator inference for nonintrusive projection-based model reduction, Comput. Methods Appl. Mech. Engrg. 306 (2016) 196–215, http://dx.doi.org/10.1016/j.cma.2016.03.025, URL https://www.sciencedirect.com/science/article/pii/S0045782516301104.

[22] O. Ghattas, K. Willcox, Learning physics-based models from data: perspectives from inverse problems and model reduction, Acta Numer. 30 (2021) 445–554, http://dx.doi.org/10.1017/S0962492921000064.

[23] P. Benner, P. Goyal, B. Kramer, B. Peherstorfer, K. Willcox, Operator inference for non-intrusive model reduction of systems with non-polynomial nonlinear terms, Comput. Methods Appl. Mech. Engrg. 372 (2020) 113433, http://dx.doi.org/10.1016/j.cma.2020.113433, URL https://www.sciencedirect.com/science/article/pii/S0045782520306186.

[24] T. Bui-Thanh, K. Willcox, O. Ghattas, B. van Bloemen Waanders, Goal-oriented, model-constrained optimization for reduction of large-scale systems, J. Comput. Phys. 224 (2) (2007) 880–896, http://dx.doi.org/10.1016/j.jcp.2006.10.026, URL https://www.sciencedirect.com/science/article/pii/S0021999106005535.

[25] M.F. Barone, I. Kalashnikova, D.J. Segalman, H.K. Thornquist, Stable Galerkin reduced order models for linearized compressible flow, J. Comput. Phys. 228 (6) (2009) 1932–1946, http://dx.doi.org/10.1016/j.jcp.2008.11.015, URL https://www.sciencedirect.com/science/article/pii/S0021999108006098.

[26] I. Kalashnikova, S. Arunajatesan, M. Barone, B. van Bloemen Waanders, J. Fike, Reduced Order Modeling for Prediction and Control of Large-Scale Systems, Tech. rep., Sandia National Laboratories, Albuquerque, NM, 2014.

[27] C.W. Rowley, T. Colonius, R.M. Murray, Model reduction for compressible flows using POD and Galerkin projection, Physica D 189 (1) (2004) 115–129, http://dx.doi.org/10.1016/j.physd.2003.03.001, URL https://www.sciencedirect.com/science/article/pii/S0167278903003841.

[28] G. Serre, P. Lafon, X. Gloerfelt, C. Bailly, Reliable reduced-order models for time-dependent linearized Euler equations, J. Comput. Phys. 231 (15) (2012) 5176–5194, http://dx.doi.org/10.1016/j.jcp.2012.04.019, URL https://www.sciencedirect.com/science/article/pii/S0021999112002008.

[29] I. Kalashnikova, M. Barone, Stable and efficient Galerkin reduced order models for non-linear fluid flow, in: AIAA-2011-3110, 6th AIAA Theoretical Fluid Mechanics Conference, Honolulu, HI, 2011.

[30] J. Chan, Entropy stable reduced order modeling of nonlinear conservation laws, J. Comput. Phys. 423 (2020) 109789, http://dx.doi.org/10.1016/j.jcp.2020.109789, URL https://www.sciencedirect.com/science/article/pii/S0021999120305635.

[31] E.J. Parish, F. Rizzi, On the impact of dimensionally-consistent and physics-based inner products for POD-Galerkin and least-squares model reduction of compressible flows, 2022, arXiv:2203.16492.

[32] M. Balajewicz, I. Tezaur, E. Dowell, Minimal subspace rotation on the stiefel manifold for stabilization and enhancement of projection-based reduced order models for the compressible Navier–Stokes equations, J. Comput. Phys. 321 (2016) 224–241, http://dx.doi.org/10.1016/j.jcp.2016.05.037, URL https://www.sciencedirect.com/science/article/pii/S0021999116301826.

[33] I. Kalashnikova, B. van Bloemen Waanders, S. Arunajatesan, M. Barone, Stabilization of projection-based reduced order models for linear time-invariant systems via optimization-based eigenvalue reassignment, Comput. Methods Appl. Mech. Engrg. 272 (2014) 251–270, http://dx.doi.org/10.1016/j.cma.2014.01.011, URL https://www.sciencedirect.com/science/article/pii/S0045782514000164.

[34] E. Rezaian, M. Wei, A global eigenvalue reassignment method for the stabilization of nonlinear reduced-order models, Internat. J. Numer. Methods Engrg. 122 (10) (2021) 2393–2416, http://dx.doi.org/10.1002/nme.6625, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.6625. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.6625.

[35] R.B. Klein, B. Sanderse, Structure-preserving hyper-reduction and temporal localization for reduced order models of incompressible flows, 2023, arXiv:2304.09229.

[36] K. Carlberg, Y. Choi, S. Sargsyan, Conservative model reduction for finite-volume models, J. Comput. Phys. 371 (2018) 280–314, http://dx.doi.org/10.1016/j.jcp.2018.05.019, URL https://www.sciencedirect.com/science/article/pii/S002199911830319X.

[37] K. Carlberg, C. Farhat, C. Bou-Mosleh, Efficient non-linear model reduction via a least-squares Petrov–Galerkin projection and compressive tensor approximations, Internat. J. Numer. Methods Engrg. 86 (2) (2011) 155–181.

[38] H.K.E. Rosenberger, B. Sanderse, Momentum-conserving ROMs for the incompressible Navier-Stokes equations, 2022, arXiv:2208.09360.

[39] M. Mohebujjaman, L. Rebholz, T. Iliescu, Physically constrained data-driven correction for reduced-order modeling of fluid flows, Internat. J. Numer. Methods Fluids 89 (3) (2019) 103–122, http://dx.doi.org/10.1002/fld.4684, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/fld.4684. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/fld.4684.

[40] H. Sharma, Z. Wang, B. Kramer, Hamiltonian operator inference: Physics-preserving learning of reduced-order models for canonical Hamiltonian systems, Physica D 431 (2022) 133122.

[41] H. Sharma, B. Kramer, Preserving Lagrangian structure in data-driven reduced-order modeling of large-scale dynamical systems, 2022, arXiv:2203.06361.

[42] S. Lall, P. Krysl, J.E. Marsden, Structure-preserving model reduction for mechanical systems, Physica D 184 (1) (2003) 304–318, http://dx.doi.org/10.1016/S0167-2789(03)00227-6, URL https://www.sciencedirect.com/science/article/pii/S0167278903002276. Complexity and Nonlinearity in Physical Systems – A Special Issue to Honor Alan Newell.

[43] K. Carlberg, R. Tuminaro, P. Boggs, Preserving Lagrangian structure in nonlinear model reduction with application to structural dynamics, SIAM J. Sci. Comput. 37 (2) (2015) B153–B184, http://dx.doi.org/10.1137/140959602, arXiv:https://doi.org/10.1137/140959602.

[44] S. Chaturantabut, D.C. Sorensen, Discrete empirical interpolation for nonlinear model reduction, in: Proceedings of the 48h IEEE Conference on Decision and Control (CDC) Held Jointly with 2009 28th Chinese Control Conference, 2009, pp. 4316–4321, http://dx.doi.org/10.1109/CDC.2009.5400045.

[45] R. Everson, L. Sirovich, Karhunen–Loève procedure for gappy data, J. Opt. Soc. Amer. A 12 (8) (1995) 1657–1664, http://dx.doi.org/10.1364/JOSAA.12.001657, URL https://opg.optica.org/josaa/abstract.cfm?URI=josaa-12-8-1657.

[46] C. Farhat, T. Chapman, P. Avery, Structure-preserving, stability, and accuracy properties of the energy-conserving sampling and weighting method for the hyper reduction of nonlinear finite element dynamic models, Internat. J. Numer. Methods Engrg. 102 (5) (2015) 1077–1110, http://dx.doi.org/10.1002/nme.4820, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.4820. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.4820.

[47] L. Peng, K. Mohseni, Symplectic model reduction of Hamiltonian systems, 2014, http://dx.doi.org/10.48550/ARXIV.1407.6118, URL https://arxiv.org/abs/1407.6118.

[48] J.S. Hesthaven, C. Pagliantini, N. Ripamonti, Structure-preserving model order reduction of Hamiltonian systems, 2021, http://dx.doi.org/10.48550/ARXIV.2109.12367, URL https://arxiv.org/abs/2109.12367.

[49] P. Buchfink, S. Glas, B. Haasdonk, Optimal bases for symplectic model order reduction of canonizable linear Hamiltonian systems*, IFAC-PapersOnLine 55 (20) (2022) 463–468, http://dx.doi.org/10.1016/j.ifacol.2022.09.138, URL https://www.sciencedirect.com/science/article/pii/S2405896322013398. 10th Vienna International Conference on Mathematical Modelling MATHMOD 2022.

[50] B.M. Afkham, J.S. Hesthaven, Structure preserving model reduction of parametric Hamiltonian systems, SIAM J. Sci. Comput. 39 (6) (2017) A2616–A2644, http://dx.doi.org/10.1137/17M1111991, arXiv:https://doi.org/10.1137/17M1111991.

[51] B. Maboudi Afkham, J.S. Hesthaven, Structure-preserving model-reduction of dissipative Hamiltonian systems, J. Sci. Comput. 81 (1) (2019) 3–21, http://dx.doi.org/10.1007/s10915-018-0653-6.

[52] C. Pagliantini, F. Vismara, Gradient-preserving hyper-reduction of nonlinear dynamical systems via discrete empirical interpolation, 2022, arXiv:2206.01792.

[53] Y. Gong, Q. Wang, Z. Wang, Structure-preserving Galerkin POD reduced-order modeling of Hamiltonian systems, Comput. Methods Appl. Mech. Engrg. 315 (2017) 780–798, http://dx.doi.org/10.1016/j.cma.2016.11.016, URL https://www.sciencedirect.com/science/article/pii/S0045782516308477.

[54] K. Sockwell, Mass Conserving Hamiltonian-Structure Preserving Reduced Order Modeling for the Rotating Shallow Water Equations Discretized by a Mimetic Spatial Scheme (Ph.D. thesis), Florida State University, 2019.

[55] H. Sharma, B. Kramer, Preserving Lagrangian structure in data-driven reduced-order modeling of large-scale mechanical systems, 2022, arXiv:2203.06361.

[56] P. Buchfink, S. Glas, B. Haasdonk, Symplectic model reduction of Hamiltonian systems on nonlinear manifolds, 2021, http://dx.doi.org/10.48550/ARXIV.2112.10815, URL https://arxiv.org/abs/2112.10815.

[57] S. Chaturantabut, C. Beattie, S. Gugercin, Structure-preserving model reduction for nonlinear port-Hamiltonian systems, SIAM J. Sci. Comput. 38 (5) (2016) B837–B865, http://dx.doi.org/10.1137/15M1055085, arXiv:https://doi.org/10.1137/15M1055085.

[58] A. Gruber, M. Gunzburger, L. Ju, Z. Wang, Energetically consistent model reduction for metriplectic systems, Comput. Methods Appl. Mech. Engrg. 404 (2023) 115709, http://dx.doi.org/10.1016/j.cma.2022.115709, URL https://www.sciencedirect.com/science/article/pii/S0045782522006648.

[59] E. Qian, B. Kramer, B. Peherstorfer, K. Willcox, Lift & learn: Physics-informed machine learning for large-scale nonlinear dynamical systems, Physica D 406 (2020) 132401, http://dx.doi.org/10.1016/j.physd.2020.132401, URL https://www.sciencedirect.com/science/article/pii/S0167278919307651.

[60] P. Khodabakhshi, K.E. Willcox, Non-intrusive data-driven model reduction for differential–algebraic equations derived from lifting transformations, Comput. Methods Appl. Mech. Engrg. 389 (2022) 114296, http://dx.doi.org/10.1016/j.cma.2021.114296, URL https://www.sciencedirect.com/science/article/pii/S0045782521005910.

[61] B. Kramer, K. Willcox, Nonlinear model order reduction via lifting transformations and proper orthogonal decomposition, AIAA J. 57 (6) (2019) 2297–2307, http://dx.doi.org/10.2514/1.J057791.

[62] S.A. McQuarrie, C. Huang, K.E. Willcox, Data-driven reduced-order models via regularised operator inference for a single-injector combustion process, J. R. Soc. N.Z. 51 (2) (2021) 194–211, http://dx.doi.org/10.1080/03036758.2020.1863237, arXiv:https://doi.org/10.1080/03036758.2020.1863237.

[63] R. Swischuk, B. Kramer, C. Huang, K. Willcox, Learning physics-based reduced-order models for a single-injector combustion process, AIAA J. 58 (6) (2020) 2658–2672, http://dx.doi.org/10.2514/1.J058943, arXiv:https://doi.org/10.2514/1.J058943.

[64] B.G. Zastrow, A. Chaudhuri, K.E. Willcox, A.S. Ashley, M.C. Henson, Data-driven model reduction via operator inference for coupled aeroelastic flutter. http://dx.doi.org/10.2514/6.2023-0330 URL https://arc.aiaa.org/doi/abs/10.2514/6.2023-0330. arXiv:https://arc.aiaa.org/doi/pdf/10.2514/6.2023-0330.

[65] O. Issan, B. Kramer, Predicting solar wind streams from the inner-heliosphere to earth via shifted operator inference, J. Comput. Phys. 473 (2023) 111689, http://dx.doi.org/10.1016/j.jcp.2022.111689, URL https://www.sciencedirect.com/science/article/pii/S0021999122007525.

[66] Y. Filanova, I.P. Duff, P. Goyal, P. Benner, An operator inference oriented approach for mechanical systems, 2022, arXiv:2210.07710.

[67] M. Guo, S.A. McQuarrie, K.E. Willcox, Bayesian operator inference for data-driven reduced-order modeling, Comput. Methods Appl. Mech. Engrg. 402 (2022) 115336, http://dx.doi.org/10.1016/j.cma.2022.115336, URL https://www.sciencedirect.com/science/article/pii/S0045782522004273. A Special Issue in Honor of the Lifetime Achievements of J. Tinsley Oden.

[68] N. Aretz, K. Willcox, Structure-exploiting learning for nested operator inference, in: Presentation at SIAM Conference on Computational Science & Engineering, 2023.

[69] B. Kramer, K. Willcox, in: C. Beattie, P. Benner, M. Embree, S. Gugercin, S. Lefteriu (Eds.), Balanced Truncation Model Reduction for Lifted Nonlinear Systems, Springer International Publishing, Cham, 2022, pp. 157–174.

[70] E. Celledoni, V. Grimm, R.I. McLachlan, D.I. McLaren, D. O'Neale, B. Owren, G.R.W. Quispel, Preserving energy resp. dissipation in numerical PDEs using the "average vector field" method, J. Comput. Phys. 231 (20) (2012/08/15/) 6770–6789, http://dx.doi.org/10.1016/j.jcp.2012.06.022, URL https://www.sciencedirect.com/science/article/pii/S0021999112003373.

[71] H. Touchette, Legendre-fenchel transforms in a nutshell, 2005, URL http://www.maths.qmul.ac.uk/~ht/archive/lfth2.pdf.

[72] Y. Nutku, On a new class of completely integrable nonlinear wave equations. I. infinitely many conservation laws, J. Math. Phys. 26 (6) (1985) 1237–1242, http://dx.doi.org/10.1063/1.526530, arXiv:https://doi.org/10.1063/1.526530.

[73] Y. Nutku, Lagrangian approach to integrable systems yields new symplectic structures for KdV, in: NATO Advanced Research Workshop on Integrable Hierarchies and Modern Physical Theories (NATO ARW - UIC 2000), 2000, arXiv:hep-th/0011052.

[74] J.H. Tu, C.W. Rowley, D.M. Luchtenburg, S.L. Brunton, J.N. Kutz, On dynamic mode decomposition: Theory and applications, J. Comput. Dyn. 1 (2) (2014) 391–421.

[75] P.J. Schmid, Dynamic mode decomposition and its variants, Annu. Rev. Fluid Mech. 54 (2022) 225–254.

[76] K. Kunisch, S. Volkwein, Galerkin proper orthogonal decomposition methods for parabolic problems, Numer. Math. 90 (1) (2001) 117–148, http://dx.doi.org/10.1007/s002110100282.

[77] Y. Gong, Q. Wang, Z. Wang, Structure-preserving Galerkin POD reduced-order modeling of Hamiltonian systems, Comput. Methods Appl. Mech. Engrg. 315 (2017) 780–798, http://dx.doi.org/10.1016/j.cma.2016.11.016, URL https://www.sciencedirect.com/science/article/pii/S0045782516308477.

[78] P. Lindsay, J. Fike, I. Tezaur, K. Carlberg, Preconditioned least-squares Petrov–Galerkin reduced order models, Internat. J. Numer. Methods Engrg. 123 (20) (2022) 4809–4843, http://dx.doi.org/10.1002/nme.7056, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.7056. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.7056.

[79] J. Conneryd, A. Lassen, Explicita symplektiska integratorer för icke-separabla Hamiltonianer i molekyldynamik, 2019.

[80] M. Tao, Explicit symplectic approximation of nonseparable Hamiltonians: Algorithm and long time performance, Phys. Rev. E 94 (2016) 043303.

[81] B. Karasözen, G. Şimşek, Energy preserving integration of bi-Hamiltonian partial differential equations, Appl. Math. Lett. 26 (12) (2013) 1125–1133, http://dx.doi.org/10.1016/j.aml.2013.06.005, URL https://www.sciencedirect.com/science/article/pii/S0893965913001924.

[82] B. Peherstorfer, Sampling low-dimensional Markovian dynamics for preasymptotically recovering reduced models from data with operator inference, SIAM J. Sci. Comput. 42 (5) (2020) A3489–A3515, http://dx.doi.org/10.1137/19M1292448.

[83] D.H. Peregrine, Calculations of the development of an undular bore, J. Fluid Mech. 25 (2) (1966) 321–330.

[84] T.B. Benjamin, J.L. Bona, J.J. Mahony, Model equations for long waves in nonlinear dispersive systems, Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 272 (1220) (1972) 47–78, http://dx.doi.org/10.1098/rsta.1972.0032, URL https://royalsocietypublishing.org/doi/abs/10.1098/rsta.1972.0032 arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.1972.0032.

[85] L.C. Evans, Partial Differential Equations, Vol. 19, American Mathematical Soc., 2010.

[86] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, İ. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, SciPy 1.0: Fundamental algorithms for scientific computing in python, Nature Methods 17 (2020) 261–272, http://dx.doi.org/10.1038/s41592-019-0686-2.

[87] J.E. Marsden, T. Ratiu, Introduction to Mechanics and Symmetry: A Basic Exposition of Classical Mechanical Systems, Springer, 1998.

[88] A. Salinger, R. Bartlett, A. Bradley, Q. Chen, I. Demeshko, X. Gao, G. Hansen, A. Mota, R. Muller, E. Nielsen, J. Ostien, R. Pawlowski, M. Perego, E. Phipps, W. Sun, I. Tezaur, Albany: Using agile components to develop a flexible, generic multiphysics analysis code, Int. J. Multiscale Comput. Eng. 14 (2016) 415–438.

[89] A. Mota, I. Tezaur, C. Alleman, The Schwarz alternating method in solid mechanics, Comput. Methods Appl. Mech. Engrg. 319 (2017) 19–51, http://dx.doi.org/10.1016/j.cma.2017.02.006, URL https://www.sciencedirect.com/science/article/pii/S0045782516310982.

[90] A. Mota, I. Tezaur, G. Phlipot, The Schwarz alternating method for transient solid dynamics, Internat. J. Numer. Methods Engrg. 123 (21) (2022) 5036–5071, http://dx.doi.org/10.1002/nme.6982, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.6982 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.6982.

[91] S.A. McQuarrie, C. Huang, K.E. Willcox, Data-driven reduced-order models via regularised operator inference for a single-injector combustion process, J. R. Soc. N.Z. 51 (2) (2021/04/03) 194–211, http://dx.doi.org/10.1080/03036758.2020.1863237.

[92] C.F. Van Loan, The ubiquitous kronecker product, J. Comput. Appl. Math. 123 (1–2) (2000) 85–100.