RESEARCH ARTICLE | MARCH 12 2024

Learning physics-based reduced-order models from data using nonlinear manifolds

Special Collection: Nonlinear Model Reduction From Equations and Data

Rudy Geelen ; Laura Balzano ; Stephen Wright ; Karen Willcox

Chaos 34, 033122 (2024)
https://doi.org/10.1063/5.0170105

Chaos ARTICLE pubs.aip.org/aip/cha

Learning physics-based reduced-order models from data using nonlinear manifolds

Cite as: Chaos 34, 033122 (2024); doi: 10.1063/5.0170105
Submitted: 31 July 2023 · Accepted: 15 February 2024 ·
Published Online: 12 March 2024

View Online Export Citation CrossMark

Rudy Geelen,[1,a)] Laura Balzano,[2] Stephen Wright,[3] and Karen Willcox[1]

AFFILIATIONS

[1]Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, Texas 78712, USA
[2]Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, Michigan 48109, USA
[3]Computer Sciences Department, University of Wisconsin, Madison, Wisconsin 53706, USA

Note: This paper is part of the Focus Issue on Nonlinear Model Reduction From Equations and Data.
[a)]**Author to whom correspondence should be addressed:** rudy.geelen@austin.utexas.edu

ABSTRACT

We present a novel method for learning reduced-order models of dynamical systems using nonlinear manifolds. First, we learn the manifold by identifying nonlinear structure in the data through a general representation learning problem. The proposed approach is driven by embeddings of low-order polynomial form. A projection onto the nonlinear manifold reveals the algebraic structure of the reduced-space system that governs the problem of interest. The matrix operators of the reduced-order model are then inferred from the data using operator inference. Numerical experiments on a number of nonlinear problems demonstrate the generalizability of the methodology and the increase in accuracy that can be obtained over reduced-order modeling methods that employ a linear subspace approximation.

Model reduction rests on the fundamental assumption that system states in complex, physics-based models can often be represented with a smaller number of variables without a significant loss of information. The identification of such intrinsic, low-dimensional structure in these problems and the subsequent inference of projection-based reduced-order models lies at the heart of this paper. We treat the construction of nonlinear state approximations of polynomial form as a general representation learning problem. By leveraging data-driven operator inference, we can then learn reduced-order models directly from available snapshot data. These models are physics-informed in that their algebraic structure is dictated by the original, high-dimensional problem. The proposed nonlinear model reduction method is interpretable and effective for reducing large-scale, dynamical-system models.

I. INTRODUCTION AND BACKGROUND

Projection-based model reduction is comprised of a family of methods that build approximations of complex physics-based models with (generally speaking) orders-of-magnitude reduction in computational complexity. Through identification of inherent low-dimensional structure, the cost of many computational tasks can be lowered significantly. Model reduction makes tractable many applications in control, uncertainty quantification, optimal experimental design, and inverse problems.[1] The key idea behind many model reduction methods is to identify a low-dimensional representation for a set of training snapshots by applying data compression. The computation of high-dimensional states is then replaced with identification of the coefficients of a basis expansion in the reduced subspace. As the effectiveness of the reduction step hinges on the ability to find a sufficiently accurate reduced-dimensional representation of full-state vectors, the task of identifying and learning such a representation is crucial to model reduction theory and methods.

Traditionally, linear approaches such as the Proper Orthogonal Decomposition (POD) are the method of choice in problems with high-dimensional state spaces associated with physics-based data and modeling.[2–4] There are several model-reduction approaches based on POD, such as dynamic mode decomposition, balanced POD, the reduced basis method, POD-based discrete empirical interpolation, and data-driven operator inference.[5–15] While all such methods have their own strengths and limitations, their formulation in terms of linear dimension reduction principles can lead to difficulties in problems for which the Kolmogorov N-width decreases slowly with increasing N. The Kolmogorov N-width is a measure for the worst-case error that might arise from the projection of a solution manifold onto a linear subspace of dimension N.[16]

Nonlinear methods with machine learning techniques at their core have received a surge in attention over the past few years for adoption in reduced-order modeling applications.[17–21] We here focus on methods that build on the concept of nonlinear *embeddings*.[22–27] More specifically, this paper develops a methodology in which we learn a reduced, nonlinear representation of the data, after which the system dynamics can be expressed and learned in terms of this new representation. We first reveal the low-dimensional manifold structure of a dynamical system by fitting a nonlinear basis expansion to the observed data. (The task of characterizing and representing this inherent structure is often referred to as *representation learning* in the machine learning community.[28]) The construction of nonlinear state representations is achieved here by enriching linear approximations with low-order polynomial terms. We then summarize two different approaches for learning the manifold geometry from data. The first is based on the POD, while the second builds on alternating minimization techniques.[26] Once the nonlinear manifold can be represented in the form of a basis expansion, a projection of the PDE model onto the manifold reveals the algebraic structure of the reduced-order model for the problem of interest. Data-driven operator inference (OpInf) is then used to construct projection-based reduced-order models from data in a non-intrusive fashion.[14,15]

A major benefit of our proposed methodology is that the system states are expressed through a compact *modal* representation. The use of modal decomposition and analysis techniques has paved the way to scalable nonlinear model reduction methods that favor interpretability and promote physical intuition. Modal representations have a long history in approximations of physics-based models, stretching back to early work in normal mode representations in structures and reduced basis representations.[29–31] In many applications, the modes have physical significance in that they represent dominant physical modal responses.[32–35] Our proposed methodology follows this template in that the latent space coordinates are the multipliers (i.e., modal coordinates) of a nonlinear combination of basis vectors with physical interpretation. An immediate consequence of this construction is that the coordinates now also represent the *evolution* of the dynamical system on a low-dimensional manifold. Because the proposed techniques are directly applicable to discrete-time observed data, the reduced model constructed in this fashion has the form of a system of nonlinear ordinary differential equations (ODEs), which can be solved effectively by modern ODE solvers.

The proposed approach shares conceptual parallels with the seminal work from Kevrekidis and co-workers on the so-called approximate inertial manifolds.[36–39] The theory of inertial manifolds states that infinite-dimensional system of partial different equations (PDEs) may be described accurately in their long term behavior by finite-dimensional systems. The existence of inertial manifolds has been established for various PDE systems in computational physics.[36,40–42] Approximate inertial manifold calculations are carried out as follows. One first performs a Galerkin approximation in the derivation of low-dimensional reduced systems, after which the higher-order modes are represented by means of the lower-order modes (a process sometimes referred to as slaving). Of particular relevance to this paper is the idea suggested in Deane *et al.*:[43] "*A particularly interesting direction is to combine the approximate inertial manifold method with the POD eigenmode hierarchy, and approximate the solution component on the higher POD modes as a function of its components on the lower, more energetic ones.*"

An outline of the remainder of the paper follows. Section II discusses the construction of nonlinear state approximations through the lens of representation learning. We demonstrate carefully how conventional linear state approximations can be enriched with polynomial terms and how the unknown basis matrices, coefficient matrix, and representation of the data in a low-dimensional coordinate system can be determined in a principled manner. Section III works with these nonlinear state approximations to derive the algebraic structure of the corresponding reduced-order models and shows how physics-based reduced-order models may be learned from data using the OpInf methodology. We then provide numerical evidence for the effectiveness of manifold-based OpInf approaches in Sec. IV using representative numerical experiments. Some conclusions and future research directions are presented in Sec. V.

II. LEARNING NONLINEAR MANIFOLDS

This section outlines our method for learning nonlinear manifolds and presents an illustrative example. Section II A discusses the general representation learning problem for constructing nonlinear state approximations in problems with high-dimensional state spaces. The numerical procedures introduced by Geelen *et al.*[26] for solving this learning problem are summarized in Sec. II B. In Sec. II C, a geometric interpretation of the method is illustrated by means of a simple three-dimensional example. A Jupyter notebook outlining the computational steps for this example is available at https://github.com/geelenr/nl_manifolds.

A. A general representation learning problem

Our focus is on data generated from complex PDE models that represent the governing laws of nature. A training data set is comprised of a set of snapshots, each snapshot being a sample of the high-fidelity state representing a particular condition of the physical system. We denote each snapshot by $\mathbf{s}_j \in \mathbb{R}^n$ for $n \in \mathbb{N}$ and construct a snapshot matrix $\mathbf{S} \in \mathbb{R}^{n\times k}$ from k such snapshots,

$$\mathbf{S} := \begin{pmatrix} | & | & & | \\ \mathbf{s}_1 & \mathbf{s}_2 & \dots & \mathbf{s}_k \\ | & | & & | \end{pmatrix}. \tag{1}$$

We make use of a reference state, $\mathbf{s}_{\text{ref}} \in \mathbb{R}^n$, and denote by $\mathbf{S}_{\text{ref}}$ the $n \times k$ reference matrix each of whose columns is $\mathbf{s}_{\text{ref}}$.

To approximate the high-dimensional state $\mathbf{s}(t) \in \mathbb{R}^n$, we seek low-dimensional approximations $\mathbf{\Gamma} : \mathbb{R}^r \mapsto \mathbb{R}^n$ such that

$$\mathbf{s}(t) \approx \mathbf{\Gamma}(\hat{\mathbf{s}}(t)), \tag{2}$$

where $t \in \mathbb{R}$ is some parameter on which the state depends (frequently, time). The vector $\hat{\mathbf{s}}(t) \in \mathbb{R}^r$ denotes the reduced state coordinate vector of dimension r. The transformation $\mathbf{\Gamma}$ constitutes a nonlinear mapping from the reduced-state coordinate system to the original, high-dimensional state space. We consider the following specific nonlinear modal basis structure for $\boldsymbol{\Gamma}$:

$$\mathbf{s}(t) \approx \boldsymbol{\Gamma}(\hat{\mathbf{s}}(t)) := \mathbf{s}_{\text{ref}} + \underbrace{\mathbf{V}\hat{\mathbf{s}}(t)}_{\text{linear}} + \underbrace{\overline{\mathbf{V}}\boldsymbol{\Xi}\,\mathbf{g}(\hat{\mathbf{s}}(t))}_{\text{nonlinear}}, \tag{3}$$

where $\mathbf{V} = [\mathbf{v}_1 \mid \dots \mid \mathbf{v}_r] \in \mathbb{R}^{n\times r}$ and $\overline{\mathbf{V}} = [\overline{\mathbf{v}}_1 \mid \dots \mid \overline{\mathbf{v}}_q] \in \mathbb{R}^{n\times q}$ are a pair of basis matrices. The matrix $\boldsymbol{\Xi} \in \mathbb{R}^{q\times(p-1)r}$ is a coefficient matrix that controls the weighting of the basis functions contained in $\overline{\mathbf{V}}$. The vector $\mathbf{g}(\hat{\mathbf{s}}(t)) \in \mathbb{R}^{(p-1)r}$ has the form

$$\mathbf{g}(\hat{\mathbf{s}}(t)) = \begin{pmatrix} \hat{\mathbf{s}}^2(t) \\ \hat{\mathbf{s}}^3(t) \\ \vdots \\ \hat{\mathbf{s}}^p(t) \end{pmatrix}, \tag{4}$$

where each $\hat{\mathbf{s}}^j(t) \in \mathbb{R}^r$ consists of the jth power of the components of $\hat{\mathbf{s}}(t)$, that is, $\hat{\mathbf{s}}^j(t) = [\hat{s}_1(t)^j, \hat{s}_2(t)^j, \dots, \hat{s}_r(t)^j]$. While one could instead employ Kronecker products of the reduced-state vector $\hat{\mathbf{s}}(t)$ (in contrast to the proposed element-wise formulation without the cross terms), the number of terms in $\mathbf{g}(\hat{\mathbf{s}}(t))$ would grow exponentially in such a construction. Keeping the dimension of $\mathbf{g}(\hat{\mathbf{s}}(t))$ relatively small is of particular importance in learning physics-based reduced-order models, to be discussed in Sec. III.

The representation learning problem to construct a nonlinear state approximation of the form (3) is now posed as a constrained optimization problem,[26]

$$\min_{\mathbf{V},\overline{\mathbf{V}},\boldsymbol{\Xi},\hat{\mathbf{S}}} \left(F(\mathbf{V}, \overline{\mathbf{V}}, \boldsymbol{\Xi}, \hat{\mathbf{S}}) + \frac{\gamma}{2}\|\boldsymbol{\Xi}\|_F^2 \right)$$
$$\text{such that } \begin{pmatrix}\mathbf{V} & \overline{\mathbf{V}}\end{pmatrix} \in \mathcal{V}_{n,(r+q)}, \tag{5}$$

where $\hat{\mathbf{S}} := (\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2, \dots, \hat{\mathbf{s}}_k) \in \mathbb{R}^{r\times k}$ is the reduced-state representation of the given system states $\mathbf{s}_j$ for $j = 1, \dots, k$; the objective function term F is defined as

$$\begin{aligned} F(\mathbf{V}, \overline{\mathbf{V}}, \boldsymbol{\Xi}, \hat{\mathbf{S}}) &= \frac{1}{2}\sum_{j=1}^{k} \left\| \mathbf{s}_j - \boldsymbol{\Gamma}(\hat{\mathbf{s}}_j) \right\|_2^2 \\ &= \frac{1}{2}\sum_{j=1}^{k} \left\| \mathbf{s}_j - \mathbf{s}_{\text{ref}} - \begin{pmatrix}\mathbf{V} & \overline{\mathbf{V}}\end{pmatrix} \begin{pmatrix} \hat{\mathbf{s}}_j \\ \boldsymbol{\Xi}\,\mathbf{g}(\hat{\mathbf{s}}_j) \end{pmatrix} \right\|_2^2. \end{aligned} \tag{6}$$

Frobenius norm regularization involving $\boldsymbol{\Xi}$ is used to avoid overfitting to the training data; $\gamma \geq 0$ is a regularization parameter; and $\mathcal{V}_{n,(r+q)}$ is the Stiefel manifold, defined as the set of matrices in $\mathbb{R}^{n\times(r+q)}$ with orthonormal columns, that is,

$$\mathcal{V}_{n,(r+q)} = \left\{ \begin{pmatrix}\mathbf{V} & \overline{\mathbf{V}}\end{pmatrix} \in \mathbb{R}^{n\times(r+q)} : \begin{pmatrix}\mathbf{V} & \overline{\mathbf{V}}\end{pmatrix}^\top \begin{pmatrix}\mathbf{V} & \overline{\mathbf{V}}\end{pmatrix} = \mathbf{I}_{r+q} \right\}, \tag{7}$$

where $\mathbf{I}_{r+q}$ is the $\mathbb{R}^{(r+q)\times(r+q)}$ identity matrix.

It should be noted that an expressive approximation of the form (3) has the tendency to overfit noise or anomalous behavior in the data. In general, the more parameters in the model, the higher the likelihood that it will overfit. Specifically, approximations with only low-order polynomial degree p (as we advocate here) are less prone to overfitting in comparison to higher order polynomial methods or alternative black-box regression methods. The addition of Frobenius regularization in (5) also helps to mitigate overfitting.

ALGORITHM 1. POD-based representation learning.

Input: Snapshot matrix $\mathbf{S} \in \mathbb{R}^{n\times k}$, reference state $\mathbf{s}_{\text{ref}} \in \mathbb{R}^n$, regularization $\gamma \in \mathbb{R}^+$, polynomial order $p \in \mathbb{N}_{\geq 2}$.
Output: Basis matrices $\mathbf{V} \in \mathbb{R}^{n\times r}$ and $\overline{\mathbf{V}} \in \mathbb{R}^{n\times q}$, coefficient matrix $\boldsymbol{\Xi} \in \mathbb{R}^{q\times(p-1)r}$, projected snapshot data $\hat{\mathbf{S}} \in \mathbb{R}^{r\times k}$.
Step 1: Fixing an orthogonal set of basis vectors

$$\mathbf{S} - \mathbf{S}_{\text{ref}} = \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Psi}^\top. \tag{8}$$

$(\mathbf{V}, \overline{\mathbf{V}}) \leftarrow$ the $r+q$ leading left singular vectors of $\mathbf{S} - \mathbf{S}_{\text{ref}}$
Step 2: Compute snapshot representation in POD coordinates

$$\hat{\mathbf{S}} = \mathbf{V}^\top(\mathbf{S} - \mathbf{S}_{\text{ref}}). \tag{9}$$

Step 3: Fixing $\mathbf{V}$, $\overline{\mathbf{V}}$, and $\hat{\mathbf{S}}$ at their values defined above, compute the coefficient matrix $\boldsymbol{\Xi}$ by solving a linear least-squares problem,

$$\min_{\boldsymbol{\Xi}} \left(F(\mathbf{V}, \overline{\mathbf{V}}, \boldsymbol{\Xi}, \hat{\mathbf{S}}) + \frac{\gamma}{2}\|\boldsymbol{\Xi}\|_F^2 \right). \tag{10}$$

B. Computing the basis expansion

This section summarizes the two methodologies from Geelen *et al.*[26] for finding a numerical approximation to the solution of the representation learning problem (5) and (6): the POD-based and alternating-minimization-based methods. To build approximations of the form (3), these approaches make informed choices on the basis matrices $\mathbf{V}$ and $\overline{\mathbf{V}}$, the coefficient matrix $\boldsymbol{\Xi}$, and the reduced-state representation of the data $\hat{\mathbf{S}}$.

In the POD-based representation learning method (see Algorithm 1), the columns of $\mathbf{V}$ are chosen to be the POD basis vectors, that is, the left singular vectors of $\mathbf{S} - \mathbf{S}_{\text{ref}}$ corresponding to the r largest singular values. The columns of the basis matrix $\overline{\mathbf{V}}$ are chosen to be the left singular vectors corresponding to the next q largest singular values. This choice of matrices $\{\mathbf{V}, \overline{\mathbf{V}}\}$ satisfies constraint (7) by virtue of the orthogonality property of singular vectors. Representation of snapshots in the POD coordinates can be calculated by means of an orthogonal projection. The coefficient matrix $\boldsymbol{\Xi}$ is obtained from a linear least-squares problem.

The alternating minimization (AM) representation learning approach (Algorithm 2) proceeds as follows. Initial guesses are required for the projected snapshot data matrix $\hat{\mathbf{S}}$ and the coefficient matrix $\boldsymbol{\Xi}$. Then, the objective in (5) is successively minimized for the three blocks of variables in turn—first $\boldsymbol{\Omega} = \begin{pmatrix}\mathbf{V} & \overline{\mathbf{V}}\end{pmatrix}$, then $\boldsymbol{\Xi}$, then $\hat{\mathbf{S}}$—with the pattern repeating until a convergence criterion is satisfied. Generally speaking, three-block alternating minimization

ALGORITHM 2. Alternating minimization-based representation learning.

Input: Snapshot matrix $\mathbf{S} \in \mathbb{R}^{n\times k}$, reference state $\mathbf{s}_{\text{ref}} \in \mathbb{R}^n$, regularization $\gamma \in \mathbb{R}^+$, polynomial order $p \in \mathbb{N}_{\geq 2}$, stopping criterion.
Output: Basis matrices $\mathbf{V} \in \mathbb{R}^{n\times r}$ and $\overline{\mathbf{V}} \in \mathbb{R}^{n\times q}$, coefficient matrix $\boldsymbol{\Xi} \in \mathbb{R}^{q\times(p-1)r}$, projected snapshot data $\hat{\mathbf{S}} \in \mathbb{R}^{r\times k}$.
while stopping criterion not satisfied **do**
Step 1: Orthogonal Procrustes: Compute the basis vectors $\boldsymbol{\Omega} := (\mathbf{V} \quad \overline{\mathbf{V}})$ by solving

$$\min_{\boldsymbol{\Omega}} \frac{1}{2}\left\| \mathbf{S} - \mathbf{S}_{\text{ref}} - \boldsymbol{\Omega}\begin{pmatrix}\hat{\mathbf{S}} \\ \boldsymbol{\Xi}\,\mathbf{g}(\hat{\mathbf{S}})\end{pmatrix}\right\|_F^2 \quad \text{such that } \boldsymbol{\Omega}^\top\boldsymbol{\Omega} = \mathbf{I}_{r+q}. \tag{11}$$

Step 2: Compute the coefficient matrix $\boldsymbol{\Xi}$ from a linear least-squares problem:

$$\min_{\boldsymbol{\Xi}} \left(F(\mathbf{V}, \overline{\mathbf{V}}, \boldsymbol{\Xi}, \hat{\mathbf{S}}) + \frac{\gamma}{2}\,\|\boldsymbol{\Xi}\|_F^2 \right) \tag{12}$$

Step 3: Project data onto the nonlinear manifold by solving

$$\min_{\hat{\mathbf{s}}_j} \frac{1}{2}\left\| \mathbf{s}_j - \mathbf{s}_{\text{ref}} - \boldsymbol{\Omega}\begin{pmatrix}\hat{\mathbf{s}}_j \\ \boldsymbol{\Xi}\,\mathbf{g}(\hat{\mathbf{s}}_j)\end{pmatrix}\right\|_2^2, \; j = 1, 2, \ldots, k. \tag{13}$$

end while

schemes have no theoretical guarantees of convergence.[44] However, in this context, and many others, they converge on practical instances.[26]

The minimization with respect to $\boldsymbol{\Omega}$ is a standard problem known as the orthogonal Procrustes problem, and it can be solved via singular value decomposition.[45] The minimization with respect to $\boldsymbol{\Xi}$ is a linear least squares problem, as in Algorithm 1. The minimization with respect to $\hat{\mathbf{S}}$ decomposes into k separate problems, each of which has a single reduced state $\hat{\mathbf{s}}_j$ as its variable.

C. Orthogonal subspace transformations: An illustrative example

We demonstrate the nonlinear representation (3) by means of a small numerical example. Consider a manifold in a three-dimensional Euclidian space parametrized by the vector $\mathbf{s}(x, y) = (x, y, \sin(x)\cos(y))^\top \in \mathbb{R}^3$ for $x, y \in [0, 4]$. A dataset is built by sampling $\mathbf{s}(x, y)$ uniformly over the domain with grid spacings $\Delta x = \Delta y = 0.1$. The resulting data matrix $\mathbf{S}$ has dimension 3×1681. After defining $\mathbf{s}_{\text{ref}}$ to be the column-averaged mean of the data matrix, we build approximations to the nonlinear manifold of dimension $r = 2$. By choosing a polynomial embedding of degree $p = 3$ in (4), the nonlinear state approximation (3) of the jth data sample becomes

$$\mathbf{s}_j \approx \mathbf{s}_{\text{ref}} + \underbrace{\mathbf{v}_1\hat{s}_{1,j} + \mathbf{v}_2\hat{s}_{2,j}}_{\text{linear}} + \underbrace{\overline{\mathbf{v}}_1\left(\Xi_1\hat{s}_{1,j}^2 + \Xi_2\hat{s}_{2,j}^2 + \Xi_3\hat{s}_{2,j}^3 + \Xi_4\hat{s}_{2,j}^3\right)}_{\text{nonlinear}}, \tag{14}$$

where the basis vectors $\mathbf{v}_1, \mathbf{v}_2, \overline{\mathbf{v}}_1$ form an orthogonal set. The modal coefficients are given by $\hat{s}_{1,j}$ and $\hat{s}_{2,j}$. Instead of introducing a *separate* coefficient for the third basis vector $\overline{\mathbf{v}}_1$, we express its coefficient in terms of the coefficients of the first two basis vectors $\hat{s}_{1,j}$ and $\hat{s}_{2,j}$. The coefficients $\boldsymbol{\Xi} = (\Xi_1, \Xi_2, \Xi_3, \Xi_4) \in \mathbb{R}^4$ control the weighting of $\overline{\mathbf{v}}_1$.

We now follow the steps from the POD-based representation learning formulation of Algorithm 1. The vectors $\mathbf{v}_1, \mathbf{v}_2, \overline{\mathbf{v}}_1$ are fixed to be the three left singular vectors of the shifted data matrix. Accordingly, $\hat{s}_{1,j}$ and $\hat{s}_{2,j}$ represent the coefficients of expansion in the basis $\mathbf{V}$. The coefficients $\boldsymbol{\Xi}$ are inferred from the data via linear regression; see (10).

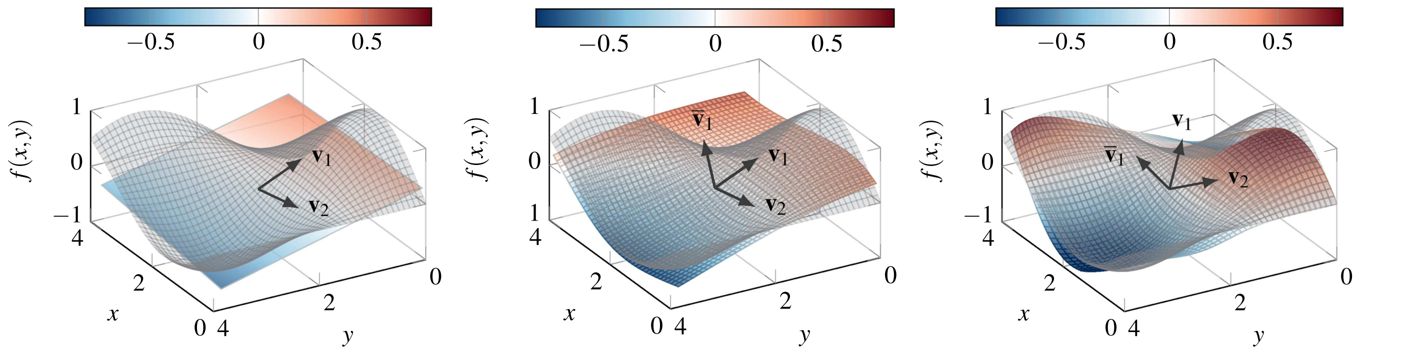
In the alternating minimization-based representation learning approach of Algorithm 2, the basis vectors $\mathbf{v}_1, \mathbf{v}_2, \overline{\mathbf{v}}_1$ and the data representation in the reduced-space coordinate system (through coordinates $\hat{s}_1, \hat{s}_2$) are computed by way of an orthogonal Procrustes problem and a set of $k = 1681$ unconstrained nonlinear optimization problems, respectively. The computation of the coefficient matrix $\boldsymbol{\Xi}$ is the linear least-squares problem (10).

Figure 1 compares the reconstructions of the standard (linear) POD approximation with those that use (14). Linear-subspace POD invokes a large projection error and, in this example, is ill-suited for data reconstruction tasks. The POD-based representation learning approach warps the POD subspace to produce a nonlinear manifold that is slightly closer to the exact solution, see Fig. 1(b). By applying a series of rotations and/or reflections to the POD basis, the alternating minimization-based method finds the preferred directions along which to apply curvature to further reduce the representation error, as shown in Fig. 1(c).

III. LEARNING REDUCED-ORDER MODELS ON NONLINEAR MANIFOLDS

In this section, we show how nonlinear manifold representations of the form (3) can be employed to learn physics-based reduced-order models from data. Assuming that the full-space data $\mathbf{s}(t)$ is generated by solving nonlinear governing equations with particular (widely relevant) structure, we substitute from (3) to obtain the corresponding system in the reduced space. We then describe a process for learning the operators that define the reduced-order model.

Section III A derives the algebraic structure of the underlying reduced-order models through a manifold projection method. In Sec. III B, we propose a manifold-based inference method for constructing reduced-order models from snapshot data.



(a) Traditional POD (linear subspace) (22.6% error)
(b) POD-based representation learning (21.1% error)
(c) AM-based representation learning (10.4% error)

FIG. 1. Comparison of (a) the linear-subspace POD with (b) the POD-based representation learning and (c) alternating minimization-based representation learning approaches for reconstructing, using two modal coefficients, the three-dimensional manifold. The relative state error $\|\mathbf{S} - \Gamma(\hat{\mathbf{S}})\|_F / \|\mathbf{S} - \mathbf{S}_{\text{ref}}\|_F$ is given between parentheses. The gray surface denotes the original three-dimensional manifold, whereas the colored surfaces illustrate the different reconstructions. The black arrows represent the basis vectors. (a) Traditional POD (linear subspace) (22.6% error), (b) POD-based representation learning (21.1% error), and (c) AM-based representation learning (10.4% error).

A. Projection-based model reduction

Consider the following initial-value nonlinear ODE problem:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{s}(t) = \mathbf{f}(\mathbf{s}(t)); \quad \mathbf{s}(0) = \mathbf{s}_0, \tag{15}$$

where as before $\mathbf{s}(t) \in \mathbb{R}^n$ is the system state at time t and $\mathbf{f} : \mathbb{R}^n \mapsto \mathbb{R}^n$ maps the state to its time derivative. In many systems that arise throughout computational engineering and sciences, the operator $\mathbf{f}$ has a certain linear-quadratic form,[15,46,47] allowing us to work with the following special case of (15):

$$\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}t} = \mathbf{A}\mathbf{s} + \mathbf{H}(\mathbf{s} \otimes \mathbf{s}); \quad \mathbf{s}(0) = \mathbf{s}_0, \tag{16}$$

where we omitted the dependence of the system states on t to simplify the notation and $\otimes$ denotes the Kronecker product. [We will refer to (16) henceforth as the full-order model (FOM).] The operators $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{H} \in \mathbb{R}^{n\times n^2}$ denote the FOM operators corresponding to linear and quadratic terms, respectively, in the governing semi-discrete equations.

The use of nonlinear state approximations of the form (3) informs the algebraic structure of the reduced-order analog of (16), requiring that we account for cubic and higher-order interactions between the modal coefficients. Specifically, by introducing (3) into (16), and projecting the residual of the resulting system onto the span of $\mathbf{V}$, we obtain

$$\begin{aligned}\frac{\mathrm{d}\hat{\mathbf{s}}}{\mathrm{d}t} &= \mathbf{V}^\top\mathbf{A}\mathbf{s}_{\text{ref}} + \mathbf{V}^\top\mathbf{A}\mathbf{V}\hat{\mathbf{s}} + \mathbf{V}^\top\mathbf{A}\overline{\mathbf{V}}\boldsymbol{\Xi}\mathbf{g}(\hat{\mathbf{s}}) + \mathbf{V}^\top\mathbf{H}(\mathbf{s}_{\text{ref}} \otimes \mathbf{s}_{\text{ref}})\\ &\quad + \mathbf{V}^\top\mathbf{H}(\mathbf{s}_{\text{ref}} \otimes \mathbf{V}\hat{\mathbf{s}}) + \mathbf{V}^\top\mathbf{H}(\mathbf{s}_{\text{ref}} \otimes \overline{\mathbf{V}}\boldsymbol{\Xi}\mathbf{g}(\hat{\mathbf{s}}))\\ &\quad + \mathbf{V}^\top\mathbf{H}(\mathbf{V}\hat{\mathbf{s}} \otimes \mathbf{s}_{\text{ref}}) + \mathbf{V}^\top\mathbf{H}(\mathbf{V}\hat{\mathbf{s}} \otimes \mathbf{V}\hat{\mathbf{s}})\\ &\quad + \mathbf{V}^\top\mathbf{H}(\mathbf{V}\hat{\mathbf{s}} \otimes \overline{\mathbf{V}}\boldsymbol{\Xi}\mathbf{g}(\hat{\mathbf{s}})) + \mathbf{V}^\top\mathbf{H}(\overline{\mathbf{V}}\boldsymbol{\Xi}\mathbf{g}(\hat{\mathbf{s}}) \otimes \mathbf{s}_{\text{ref}})\\ &\quad + \mathbf{V}^\top\mathbf{H}(\overline{\mathbf{V}}\boldsymbol{\Xi}\mathbf{g}(\hat{\mathbf{s}}) \otimes \mathbf{V}\hat{\mathbf{s}}) + \mathbf{V}^\top\mathbf{H}(\overline{\mathbf{V}}\boldsymbol{\Xi}\mathbf{g}(\hat{\mathbf{s}}) \otimes \overline{\mathbf{V}}\boldsymbol{\Xi}\mathbf{g}(\hat{\mathbf{s}})),\end{aligned} \tag{17}$$

with $\hat{\mathbf{s}}(0) = \hat{\mathbf{s}}_0$ the representation of the initial condition $\mathbf{s}_0$ in the low-dimensional coordinate system. An alternative approach would be to choose a nonlinear, state-dependent projection of the state equations, which comes at the cost of increased algebraic complexity in the reduced-order models.[48]

By adapting the argument of Sec. II B from a previous work[24] to this case, we can show that when $\mathbf{V}$ and $\hat{\mathbf{S}}$ are obtained from an SVD of $\mathbf{S} - \mathbf{S}_{\text{ref}}$ (as in standard linear POD), $\boldsymbol{\Xi}$ is fixed at some reasonable value, and we minimize the objective function in (5) over $\overline{\mathbf{V}}$ alone, the orthogonality property $\mathbf{V}^\top\overline{\mathbf{V}} = \mathbf{0}$ will be satisfied automatically by the minimizing value of $\overline{\mathbf{V}}$. Although this property cannot be guaranteed to hold when we optimize *simultaneously* with respect to $\mathbf{V}$, $\overline{\mathbf{V}}$, $\hat{\mathbf{S}}$, and $\boldsymbol{\Xi}$, as in (5), much remains to be gained in terms of simplification of our inferred reduced-order model. Specifically, by enforcing the orthogonality property in (5) the left-hand side of (17) simplifies to a time derivative applied to the reduced state vector $\hat{\mathbf{s}}(t)$. A lack of orthogonality, on the other hand, would result in an expression that involves the time derivative of the polynomial function $\mathbf{g}(\hat{\mathbf{s}}(t))$. The added nonlinearity would cause the evaluation of the data-driven models to be more mathematically cumbersome.

The right-hand side of (17) contains polynomial nonlinear terms up to order $2p$. In practice, this structure poses a significant challenge from the implementation viewpoint: explicitly computing the different projected operators becomes cumbersome and requires explicit access to the full-order operators $\mathbf{A}$ and $\mathbf{H}$. Rather than operating on (17) directly, we expose its polynomial structure by using the mixed-product property of Kronecker products and grouping

TABLE I. Dimension $d(r, p)$ as a function of the reduced basis dimension, r and the degree of the polynomial embeddings, p.

	$p=2$	$p=3$	$p=4$
$r=2$	7	16	27
$r=4$	26	64	114
$r=6$	57	144	261
$r=8$	100	256	468
$r=10$	155	400	735

the constant, linear, quadratic, and higher-order terms as follows:

$$\frac{\mathrm{d}\hat{\mathbf{s}}}{\mathrm{d}t} = \hat{\mathbf{c}} + \hat{\mathbf{A}}\hat{\mathbf{s}} + \hat{\mathbf{H}}(\hat{\mathbf{s}} \otimes \hat{\mathbf{s}}) + \hat{\mathbf{P}}\hat{\mathbf{g}}(\hat{\mathbf{s}}), \tag{18}$$

where $\hat{\mathbf{c}} \in \mathbb{R}^r$, $\hat{\mathbf{A}} \in \mathbb{R}^{r\times r}$, $\hat{\mathbf{H}} \in \mathbb{R}^{r\times r^2}$, and $\hat{\mathbf{P}} \in \mathbb{R}^{r\times d(r,p)}$ are the reduced matrix operators. The operator $\hat{\mathbf{P}}$ accounts for the higher-order interactions between the modal coefficients in the reduced-order model. These interactions are captured by the vector $\hat{\mathbf{g}}(\hat{\mathbf{s}})$ [which is a subvector of $\mathbf{g}(\hat{\mathbf{s}})$ defined in (4)] and consist of monomials from degree three to degree $2p$. The total number of unique coefficients in $\hat{\mathbf{g}}(\hat{\mathbf{s}})$ scales as $d(r,p) \sim \mathcal{O}(p^2r^2)$. Example III.1 provides an illustration of the structure of $\hat{\mathbf{g}}(\hat{\mathbf{s}})$ and the scaling of $d(r,p)$ in reduced-order models of the form (18). A technique for approximating the reduced-order operators $\hat{\mathbf{c}}, \hat{\mathbf{A}}, \hat{\mathbf{H}}, \hat{\mathbf{P}}$ is presented in Sec. III B.

Example III.1. For illustration purposes, consider a state approximation of dimensionality $r=2$ and polynomial degree $p=3$ given by

$$\mathbf{s} \approx \mathbf{s}_{\mathrm{ref}} + \mathbf{V}(\hat{s}_1, \hat{s}_2)^\top + \overline{\mathbf{V}}\boldsymbol{\Xi}\left(\hat{s}_1^2, \hat{s}_2^2, \hat{s}_1^3, \hat{s}_2^3\right)^\top, \tag{19}$$

where $\mathbf{V} \in \mathbb{R}^{n\times 2}$ and $\overline{\mathbf{V}} \in \mathbb{R}^{n\times q}$ are the basis matrices and $\boldsymbol{\Xi} \in \mathbb{R}^{q\times 4}$ is the coefficient matrix, calculated as described in Sec. II. The modal coefficients $\hat{s}_1, \hat{s}_2$ are the only unknowns. By substituting from (19) into the linear-quadratic full-order model (16) and following the derivation from Sec. III A, we obtain a reduced model of the form (18). This model then accounts for the following nonlinear interactions among the modal coefficients:

$$\begin{aligned} \hat{\mathbf{s}} &= \left(\hat{s}_1, \hat{s}_2\right)^\top \in \mathbb{R}^2, \\ \hat{\mathbf{s}} \otimes \hat{\mathbf{s}} &= \left(\hat{s}_1^2, \hat{s}_1\hat{s}_2, \hat{s}_2^2\right)^\top \in \mathbb{R}^3, \\ \hat{\mathbf{g}}(\hat{\mathbf{s}}) &= \left(\hat{s}_1^3, \hat{s}_1\hat{s}_2^2, \hat{s}_1^2\hat{s}_2, \hat{s}_2^3, \hat{s}_1^4, \hat{s}_1\hat{s}_2^3, \hat{s}_1^3\hat{s}_2, \hat{s}_1^2\hat{s}_2^2, \hat{s}_2^4, \ldots \right. \\ &\qquad \left. \hat{s}_1^5, \hat{s}_1^2\hat{s}_1^3, \hat{s}_1^3\hat{s}_2^2, \hat{s}_2^5, \hat{s}_1^6, \hat{s}_1^3\hat{s}_2^3, \hat{s}_2^6\right)^\top \in \mathbb{R}^{d(r,p)=16}, \end{aligned}$$

where $\hat{\mathbf{s}}$, $\hat{\mathbf{s}} \otimes \hat{\mathbf{s}}$, and $\hat{\mathbf{g}}(\hat{\mathbf{s}})$ contain monomials of the modal coefficients of first, second, and higher-order degree, respectively. Using elementwise powers of the reduced-state vector $\hat{\mathbf{s}}$ [see (4)] ensures that the number of entries in $\hat{\mathbf{g}}(\hat{\mathbf{s}})$ remains tractable. Table I lists $d(r,p)$ the number of terms contained in $\hat{\mathbf{s}}$ as a function of the reduced basis dimension r and the degree of the polynomial embeddings p.

B. Learning physics-based reduced-order models from data

We employ the data-driven operator inference (OpInf) method for learning the low-dimensional dynamical system (18) from time-domain simulation data.[14] While traditional linear-subspace POD lies at the heart of the formulation of Peherstorfer and Willcox,[14] OpInf can be extended to the nonlinear manifold setting, as demonstrated in Geelen *et al.* for linear systems.[24] Here, we consider reduction of nonlinear systems. The OpInf methodology finds the reduced matrix operators $\hat{\mathbf{c}}, \hat{\mathbf{A}}, \hat{\mathbf{H}}, \hat{\mathbf{P}}$ that define the reduced model that best matches the projected snapshot data, in the following sense of regularized least squares:

$$\begin{aligned} \left(\hat{\mathbf{c}}, \hat{\mathbf{A}}, \hat{\mathbf{H}}, \hat{\mathbf{P}}\right) = \mathrm{argmin}_{\tilde{\mathbf{c}},\tilde{\mathbf{A}},\tilde{\mathbf{H}},\tilde{\mathbf{P}}} \Big(& J(\tilde{\mathbf{c}}, \tilde{\mathbf{A}}, \tilde{\mathbf{H}}, \tilde{\mathbf{P}}) \\ & + \frac{\lambda_1}{2}\left(\|\tilde{\mathbf{c}}\|_2^2 + \|\tilde{\mathbf{A}}\|_F^2\right) + \frac{\lambda_2}{2}\|\tilde{\mathbf{H}}\|_F^2 + \frac{\lambda_3}{2}\|\tilde{\mathbf{P}}\|_F^2 \Big), \end{aligned} \tag{20}$$

where the function $J(\tilde{\mathbf{c}}, \tilde{\mathbf{A}}, \tilde{\mathbf{H}}, \tilde{\mathbf{P}})$ is defined to be

$$\sum_{j=1}^{k} \left\| \tilde{\mathbf{c}} + \tilde{\mathbf{A}}\hat{\mathbf{s}}_j + \tilde{\mathbf{H}}(\hat{\mathbf{s}}_j \otimes \hat{\mathbf{s}}_j) + \tilde{\mathbf{P}}\hat{\mathbf{g}}(\hat{\mathbf{s}}_j) - \frac{\mathrm{d}\hat{\mathbf{s}}_j}{\mathrm{d}t} \right\|_2^2, \tag{21}$$

while the nonnegative scalars λ_i with $i = 1, 2, 3$ are Tikhonov regularization parameters that promote stability in the inferred reduced-order models and inhibit the overfitting of the system operators to potentially noisy data.[49] Generally speaking, good choices of the regularization parameters λ_i tend to be different as they are coefficients of terms with different scales. The time derivatives in the objective function are typically estimated numerically using finite difference approximations. The optimization problem (20) decouples into r independent linear least-squares problems.[14]

Algorithm 3 presents the steps of the linear-subspace OpInf approach for quadratic systems,[14,49] while the workflow of the proposed nonlinear manifold-based OpInf methodology is summarized in Algorithm 4. We use the acronyms MPOD-OpInf and MAM-OpInf to distinguish between the nonlinear manifold OpInf approaches based on Algorithms 1 and 2, respectively. These methodologies differ in the manner in which the projected snapshot data are computed and thus also in their subsequent reconstructions in the original state space. This is due to differences in the adopted low-dimensional basis, coefficient matrix, and reduced-state data representation; see Sec. II.

IV. NUMERICAL EXPERIMENTS

In this section, we discuss the application of the OpInf model reduction methods described in Sec. III to several dynamical systems. We compare the OpInf approach from Peherstorfer and Willcox[14] (Algorithm 3) and the nonlinear-manifold-based OpInf approaches proposed above: MPOD-OpInf and MAM-OpInf (see Algorithm 4). We report numerical results for benchmark problems involving the Allen–Cahn equation, the Korteweg–de Vries equation, and a cylinder flow problem.

ALGORITHM 3. The standard (POD-based) OpInf methodology for quadratic systems.[14,49]

Input: Snapshot matrix $\mathbf{S} \in \mathbb{R}^{n\times k}$, reference state $\mathbf{s}_{\text{ref}} \in \mathbb{R}^n$
Output: Reduced operators $\hat{\mathbf{c}}$, $\hat{\mathbf{A}}$, $\hat{\mathbf{H}}$
1: Compute SVD of $\mathbf{S} - \mathbf{S}_{\text{ref}}$
2: $\mathbf{V} \leftarrow$ The r leading left singular vectors of the $\mathbf{S} - \mathbf{S}_{\text{ref}}$
3: $\hat{\mathbf{S}} \leftarrow \mathbf{V}^\top(\mathbf{S} - \mathbf{S}_{\text{ref}})$ {Project snapshot data onto POD subspace}
4: Approximate $\frac{\mathrm{d}}{\mathrm{d}t}\hat{\mathbf{S}}$ from $\hat{\mathbf{S}}$ {Time derivative approximation}
5: Choose λ_1, λ_2 {Hyperparameter optimization}
6: Find reduced operators $\hat{\mathbf{c}}$, $\hat{\mathbf{A}}$, $\hat{\mathbf{H}}$ through OpInf regression

A. Practical considerations

The dimensionality of a reduced-order model is typically informed by the representation error of the training data used in its construction. For nonlinear approximations of the form (3), we compute the metric[24]

$$\varepsilon_r = \frac{\|\mathbf{V}\hat{\mathbf{S}} + \overline{\mathbf{V}}\boldsymbol{\Xi}\mathbf{g}(\hat{\mathbf{S}})\|_F^2}{\|\mathbf{S} - \mathbf{S}_{\text{ref}}\|_F^2}. \tag{22}$$

When $\overline{\mathbf{V}}$ is zero (as in the linear-subspace OpInf approach), we recover the well-known expression

$$\varepsilon_r = \sum_{j=1}^{r}\sigma_j^2 \Big/ \sum_{j=1}^{k}\sigma_j^2, \tag{23}$$

where σ_j denotes the jth singular value of the mean-centered snapshot matrix $\mathbf{S} - \mathbf{S}_{\text{ref}}$. This indicator is commonly referred to as the (cumulative) snapshot energy captured by the basis. The dimensionality of the reduced-order models, r, and the number of orthogonal basis vectors in the nonlinear part of the approximation, q, are user-specified. In the following, these values are chosen based on the singular value decay of the shifted snapshot matrices. The polynomial order, p, of approximation (3) is chosen, in accordance with r and q, based on the ability to represent data in the reduced-state coordinate system to sufficient accuracy.

The primary error metric used in numerical experiments is the relative error in the states, namely, $\|\mathbf{S} - \boldsymbol{\Gamma}(\hat{\mathbf{S}})\|_F/\|\mathbf{S} - \mathbf{S}_{\text{ref}}\|_F$. Initial guesses for $\mathbf{V}, \overline{\mathbf{V}}, \boldsymbol{\Xi}, \hat{\mathbf{S}}$ in the alternating minimization method from Algorithm 2 are obtained from the POD-based representation learning method of Algorithm 1. The iterative process from Algorithm 2 is terminated when the change in the relative snapshot energy ε_r (22) on consecutive iterations falls below 10^{-3}. Note that this termination criterion for the alternating minimization scheme is not indicative of the snapshot energy captured by a given number of basis vectors and reduced-state coefficients. One might change this threshold based on model size or desired optimization tolerance. The function tolerance for the nonlinear least-squares solver used in solving (10) is set to 10^{-9}. It is noted that the MAM-OpInf approach calls for a representation of the initial condition in the reduced-state coordinate system. This involves computing r coefficients through a nonlinear least-squares problem which is carried out at the start of each reduced model evaluation.

The regularization parameters λ_i with $i = 1, 2, 3$ in (20) and the regularization parameter γ in representation learning problem (5) are calibrated through a grid search conducted over a predetermined range of candidate values. We seek the parameter combination that minimizes the relative state error over the available training data.[49] This requires that the reduced-order model be evaluated for every parameter combination. We also note that γ does not (explicitly) show up in OpInf problem (20) as it appears only upon computing a representation of the data in the original state-space through (3). The time derivatives of the projected snapshot data in the OpInf regression problems are estimated via a fourth-order finite difference approximation.

B. The Allen–Cahn equation

While the Allen–Cahn model was originally conceived to describe the motion of anti-phase boundaries in metallic alloys,[50] it has become prototypical for describing phase separation and interfacial dynamics in many application domains. The equation has also been studied in the context of model reduction.[51,52] We consider the Allen–Cahn equation

$$\partial_t s = \kappa\,\partial_x^2 s + s - s^3 \tag{24}$$

in the domain $x \in [-1, 1]$ with Dirichlet boundary conditions $s(-1, t) = -1; s(1, t) = 1$ and initial condition

$$s(x, 0) = \mu x + (1 - \mu)\sin(-1.5\pi x),$$

ALGORITHM 4. Proposed nonlinear manifold-based OpInf method for quadratic systems.

Input: Snapshot matrix $\mathbf{S} \in \mathbb{R}^{n\times k}$, reference state $\mathbf{s}_{\text{ref}} \in \mathbb{R}^n$, regularization parameter $\gamma \in \mathbb{R}^+$, polynomial order $p \in \mathbb{N}_{\geq 2}$, stopping criterion
Output: Reduced operators $\hat{\mathbf{c}}$, $\hat{\mathbf{A}}$, $\hat{\mathbf{H}}$, $\hat{\mathbf{P}}$
1: Compute SVD of $\mathbf{S} - \mathbf{S}_{\text{ref}}$
2: $r, q \leftarrow$ Choose the number of left singular vectors to be used in basis matrices $\mathbf{V}$ and $\overline{\mathbf{V}}$
3: $\boldsymbol{\Gamma}(\mathbf{V}, \overline{\mathbf{V}}, \boldsymbol{\Xi}, \hat{\mathbf{S}}) \leftarrow$ Approximate the solution to general representation learning problem (5) using Algorithm 1 or 2
4: Approximate $\frac{\mathrm{d}}{\mathrm{d}t}\hat{\mathbf{S}}$ from $\hat{\mathbf{S}}$ {Time derivative approximation}
5: Choose $\lambda_1, \lambda_2, \lambda_3$ {Hyperparameter optimization}
6: Solve OpInf regression problem (20) to find $\hat{\mathbf{c}}$, $\hat{\mathbf{A}}$, $\hat{\mathbf{H}}$, $\hat{\mathbf{P}}$.

in which the parameter μ varies uniformly on the range $[0.5, 0.6]$. The operators ∂_x and ∂_t in (24) denote partial differentiation with respect to space and time, respectively. The interface parameter $\kappa \in \mathbb{R}^+$ is positive constant which represents the thickness of the interface that separates the two phases.

While the Allen–Cahn Eq. (24) is characterized by linear and cubic terms, adding an auxiliary variable yields system dynamics with the desired quadratic model structure (16). We apply the transformation[15,53]

$$\mathcal{T} : (s) \mapsto \begin{pmatrix} s \\ s^2 \end{pmatrix} \equiv \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}.$$

The lifted system is then given by

$$\begin{aligned} \partial_t w_1 &= \varepsilon \partial_x^2 w_1 + w_1 + w_1 w_2, \\ \partial_t w_2 &= 2 w_1 \partial_t w_1 \\ &= 2 w_1 \left(\varepsilon \partial_x^2 w_1 + w_1 + w_1 w_2\right) \\ &= 2 \varepsilon w_1 \partial_x^2 w_1 + 2 w_2 + 2 (w_2)^2, \end{aligned} \tag{25}$$

which contains only quadratic nonlinear dependencies on the state. It is also important to note that no approximations are invoked in the process of lifting (24) to (25).

State data are computed on a uniform spatial grid consisting of $n = 512$ grid points. The state snapshot data are generated from three simulations of the Allen–Cahn model, corresponding to the parameters $\mu = [0.50, 0.55, 0.60]$. For testing, ten more trajectories are generated with parameter μ drawn uniformly at random from the interval $[0.5, 0.6]$. The data are recorded every 0.1 time units up to time $T = 60$, yielding 600 snapshots per trajectory, for a total of 1800. The lifted snapshot matrix is centered by the mean initial condition across the test parameters.

Figure 2 shows the decay of the normalized singular values, where normalized means that the first normalized singular value equals 1. The first two POD modes contain 97.7% of the energy in the lifted state data, but a total of $r + q = 20$ POD modes are needed to drive the projection error in the training data below 10^{-5}. We consider two-equation reduced-order models constructed from OpInf methodologies. For the MPOD-OpInf formulation, the basis matrix $\mathbf{V}$ contains the first $r = 2$ POD modes, with the remaining $q = 20 - r = 18$ POD modes captured in $\overline{\mathbf{V}}$. The regularization parameter in the representation learning problem in (5) is chosen to be $\gamma = 10^{-2}$. Value for the regularization parameters λ_i, $i = 1, 2, 3$ in the OpInf problems are found by minimizing the relative state error across all the training parameters μ.

Figure 3 shows the reconstructed trajectories at a random test parameter $\mu = 0.5127$ for the linear-subspace OpInf formulation and its manifold-based counterpart using fourth-order polynomial embeddings. The MPOD-OpInf model represents the phase separation process over time more accurately, in both the material phases and the interface dynamics. Pointwise errors in the reconstructions of the original state data for the nonlinear manifold models are shown in Fig. 4. We note from these figures that increasing the degree of the polynomial embeddings can improve the predictive capabilities. (Performance of a linear-subspace OpInf reduced-order model is shown for reference.) The relative state errors for the

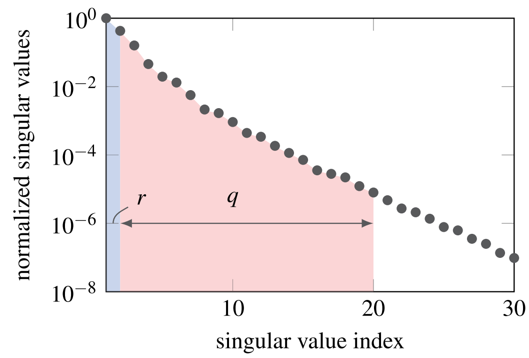


FIG. 2. Normalized singular values of the centered snapshot matrix for the Allen–Cahn problem. The blue and red areas denote the singular values whose corresponding left singular vectors are columns in $\mathbf{V}$ and $\overline{\mathbf{V}}$, respectively, in the MPOD-OpInf formulation.

reduced-order models are tabulated in Table II. It can be seen that the MPOD-OpInf formulation outperforms the linear-subspace OpInf formulation in the training regime as well as in the predictive setting.

C. The Korteweg–de Vries equation

The second numerical experiment is concerned with traveling wave physics.[54] We consider a single propagating soliton in a one-dimensional domain with periodic boundary conditions. The evolution of the wave field s in the space-time domain $[-\pi, \pi] \times [0, T]$ is obtained from the Korteweg–de Vries equation

$$\partial_t s = -\alpha s \partial_x s - \beta \partial_x^3 s. \tag{26}$$

The initial condition is given by $s_0(x) = 1 + 24 \operatorname{sech}^2\left(\sqrt{8} x\right)$. We use an equidistant computational grid consisting of 256 evenly spaced points in space. State data are saved every 0.0002 time units. We choose a final time $T = 1$ and model constants $\alpha = 4$ and $\beta = 1$.

To learn the nonlinear manifolds and train our data-driven reduced-order models, 1001 snapshots of the solution are collected uniformly across the time interval $t \in [0, 0.2]$. The snapshot matrix under consideration is centered by its column-averaged (thus time-averaged) mean value; the decay of its singular values is shown in Fig. 5. The dynamics of the system can be captured well with only 14 modes capturing 99.3% of the cumulative snapshot energy. The values of r (the number of columns of $\mathbf{V}$) and q (the number of columns

TABLE II. Median of relative state error (22) in the training and test problems across the parameters for the Allen–Cahn problem.

	Training	Testing
Linear-subspace OpInf	3.599×10^{-1}	3.424×10^{-1}
MPOD-OpInf ($p = 2$)	1.823×10^{-1}	1.552×10^{-1}
MPOD-OpInf ($p = 3$)	5.091×10^{-2}	4.368×10^{-2}
MPOD-OpInf ($p = 4$)	2.567×10^{-2}	2.552×10^{-2}

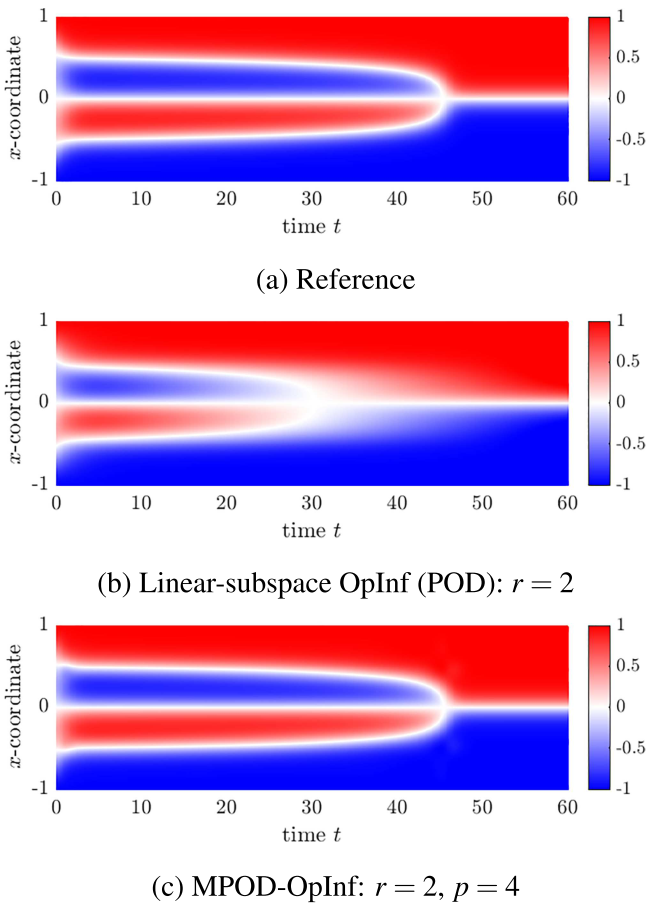


FIG. 3. Comparison of the reference solution at parameter value $\mu = 0.5127$ (top) with the reconstructions from the two-equation OpInf (middle) and MPOD-OpInf (bottom) models for the Allen–Cahn model. (a) Reference, (b) linear-subspace OpInf (POD): $r = 2$, (c) MPOD-OpInf: $r = 2$, $p = 4$.

of $\overline{\mathbf{V}}$) are then chosen so that $r + q = 14$. The regularization parameter in the representation learning problem in (5) is chosen so that $\gamma = 10^{-3}$. We now consider the performance of the reduced-order models in both the training regime and the predictive setting.

Visual comparisons of the reference solution and the solutions produced by the OpInf models are shown in Fig. 6. The MPOD-OpInf and MAM-OpInf models use polynomial embeddings of degree $p = 2$. Even at $r = 5$ basis vectors, the space-time evolution of the propagating soliton is relatively well captured. The OpInf, MPOD-OpInf, and MAM-OpInf models account for 73.6%, 82.0%, and 93.1%, respectively, of the cumulative energy (22) in the system. The relative state errors for these formulations, over the window of training, equals 51.4%, 42.5%, and 30.0%. Predictably, the addition of quadratic terms in the state approximation yields an increase in accuracy. By accounting for orthogonal transformations of the basis, as induced by the alternating minimization procedure, the contrast in accuracy becomes more pronounced, as

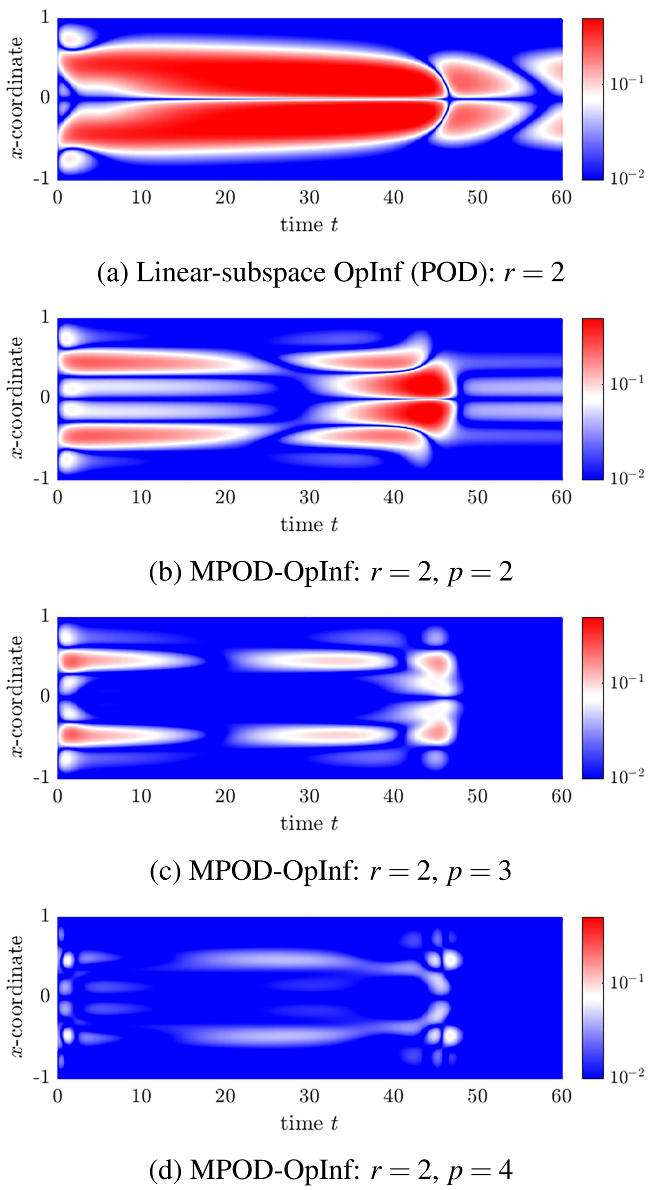


FIG. 4. Pointwise error in the reconstructions for the test trajectory at parameter value $\mu = 0.5127$ for the Allen–Cahn model. (a) Linear-subspace OpInf (POD): $r = 2$, (b) MPOD-OpInf: $r = 2$, $p = 2$, (c) MPOD-OpInf: $r = 2$, $p = 3$, and (d) MPOD-OpInf: $r = 2$, $p = 4$.

the MAM-OpInf model is more accurate than the MPOD-OpInf variant (despite these two models having the exact same online computational expense).

Outside the range of their training data (that is, $t \in [0.2, 1]$), the relative state errors are 52.9%, 43.3%, and 35.1%, for the OpInf, MPOD-OpInf, and MAM-OpInf models, respectively. This experiment demonstrates the potential of both the MPOD-OpInf

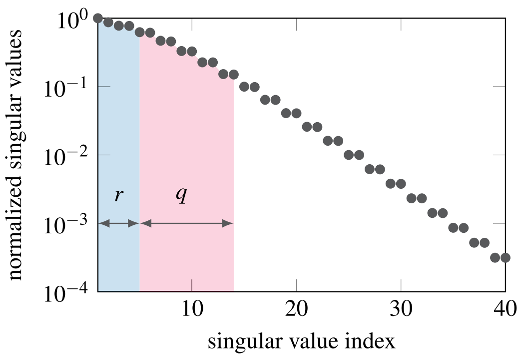


FIG. 5. Normalized singular values of the mean-subtracted snapshot matrix for the Korteweg–de Vries problem. The blue and red areas denote the singular values whose corresponding left singular vectors are columns in $\mathbf{V}$ and $\overline{\mathbf{V}}$, respectively, in the MPOD-OpInf formulation.

and MAM-OpInf methods to outperform linear-subspace OpInf methods in the training and (especially) in the predictive regime. These results are corroborated by comparing the solution snapshots at the end of training ($t = 0.2$) and at final time ($t = 1$) for the different methods (see Fig. 7). While there is some error associated with the prediction of the soliton's exact spatial location, the MPOD-OpInf and MAM-OpInf formulations are better suited for capturing the soliton's representation over time.

The first four basis vectors (corresponding to the dominant left singular vectors) for state approximations of dimension $r = 5$ of order $p = 2$ are shown in Fig. 8. While both the MPOD-OpInf and MAM-OpInf approaches produce an orthogonal set of vectors, the alternating minimization approach can be seen to incorporate some of the small-scale solution features into the dominant modes.

When we repeat the experiment but with the dimension of the reduced-order model increased to $r = 16$ (choosing $r + q = 25$), the state error for the OpInf, MPOD-Opinf, and MAM-OpInf models in the range of the training data drops further to 5.7%, 4.3%, and 2.4%, respectively (see Fig. 9). However, if the models are integrated to final time T, the error increases rapidly. The state errors across the time interval $[0.2, 1]$ are 30.4%, 48.2%, and 76.4%, respectively, for the three OpInf variants. Figure 10 shows that these errors can be attributed to inaccurate predictions of the soliton's spatial location. These results demonstrate an important trade-off between choosing r, the dimension of the linear subspace described by basis $\mathbf{V}$, and q, the number of basis vectors in the basis $\overline{\mathbf{V}}$. With $r = 16$, the linear subspace captures 99.7% of the snapshot energy. Enriching the approximation by adding the $\overline{\mathbf{V}}$ manifold terms using the next $q = 9$ singular vectors leads to 99.85% snapshot energy being captured by the polynomial manifold representation (i.e., an increase of only 0.15%). In this case, the components of the basis $\overline{\mathbf{V}}$ correspond to singular vectors with near-zero singular values and the additional terms provide little benefit—in fact, in this example they lead to overfitting and a decline in reduced model predictive performance.

We now shift attention to the computational cost of integrating reduced models of the form (18). We monitor the accuracy of the model in the training regime, as given by the representation error

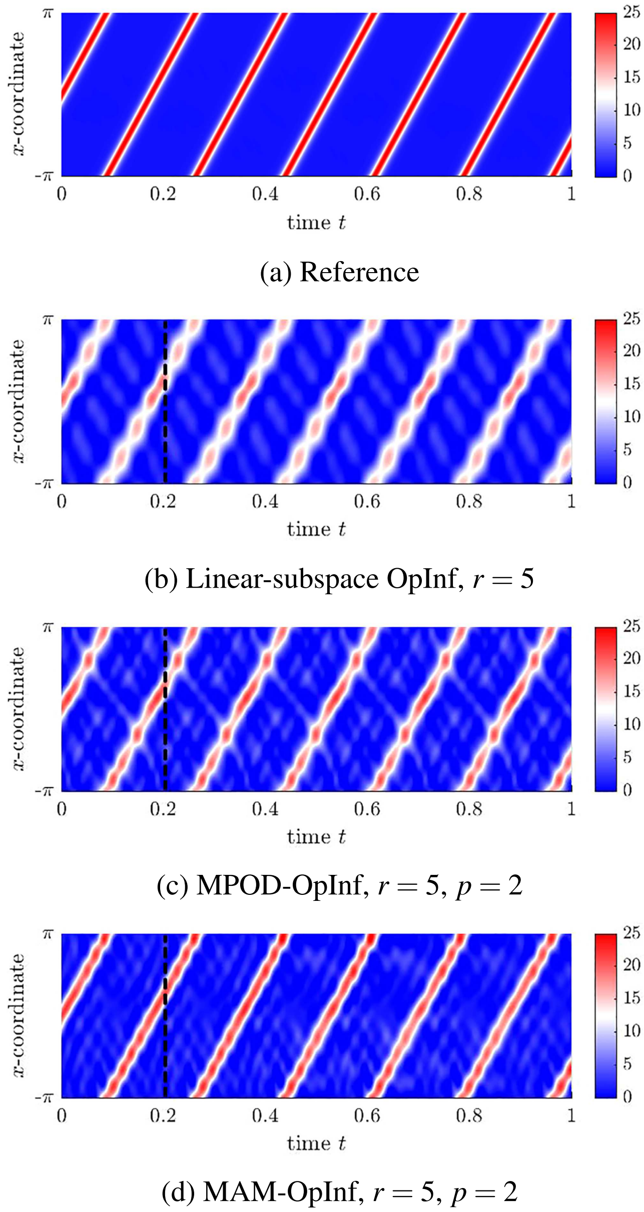


FIG. 6. Plots of the reference solution and OpInf-produced predictions for the Korteweg–de Vries equation over the time window $t \in [0, T]$ at a reduced basis dimension of $r = 5$. The end of the training window is indicated by the dashed line. (a) Reference, (b) linear-subspace OpInf, $r = 5$, (c) MPOD-OpInf, $r = 5$, $p = 2$, and (d) MAM-OpInf, $r = 5$, $p = 2$.

of the training data (22), as a function of the length of the reduced-representation vector $[\hat{\mathbf{s}}^\top, (\hat{\mathbf{s}} \otimes \hat{\mathbf{s}})^\top, \hat{\mathbf{g}}(\hat{\mathbf{s}})^\top]^\top$, which equals $r + r(r+1)/2 + d(r, p)$. The results are summarized in Fig. 11 for the POD-based and AM-based manifold formulations. A higher dimension for the reduced-order model leads to both more expensive

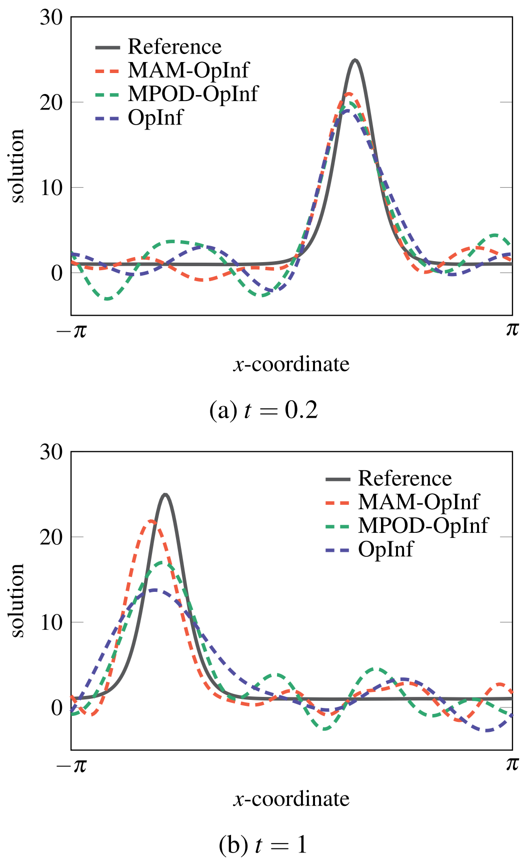


FIG. 7. Reference and OpInf solution snapshots from the Korteweg–de Vries experiment at the end of the training regime ($t = 0.2$) and predictions for the final time T. The plots are given for models of dimension $r = 5$. (a) $t = 0.2$ and (b) $t = 1$.

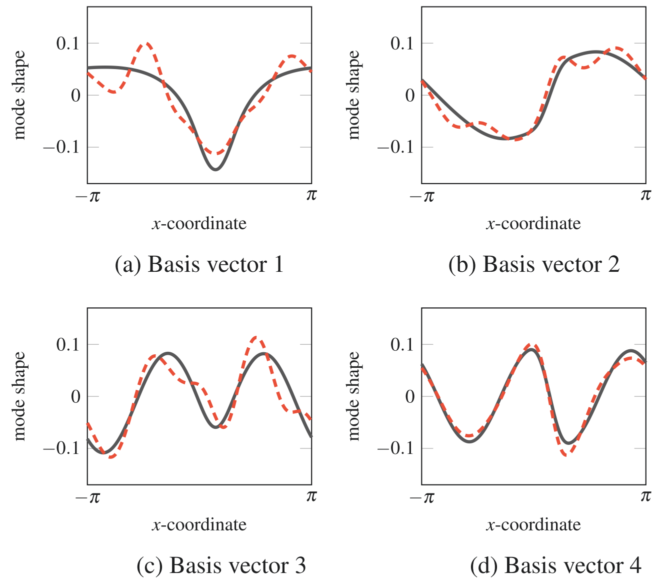


FIG. 8. First four basis vectors computed using POD (black curves) and those obtained by means of orthogonal Procrustes problem (10) in Algorithm 2 (red curves) for the Korteweg–de Vries problem. For the latter curves, the modes are computed for models of size $r = 5$ with quadratic embeddings ($p = 2$). (a) Basis vector 1, (b) Basis vector 2, (c) Basis vector 3, and (d) Basis vector 4.

computation and increased accuracy. However, for models of the same dimensionality (for example, $r = 4$), it can be seen that the total number of terms grows rapidly with the degree of the polynomial embeddings. Note that this analysis pertains only to *online* computational costs. (Reduced-order modeling of large-scale dynamical systems typically invokes a high cost during the offline phase, which is performed only once.)

D. Incompressible Navier–Stokes—Flow past a cylinder

We now apply our techniques to the well-investigated problem of two-dimensional transient flow past a circular cylinder. We focus on the configuration with Reynolds number $Re = 100$, a value that is above the critical Reynolds number for the onset of the two-dimensional vortex shedding. The fluid flow is governed by the incompressible Navier–Stokes equations

$$\partial_t \mathbf{u} + \nabla \cdot (\mathbf{u} \otimes \mathbf{u}) = \nabla p + Re^{-1} \Delta \mathbf{u}, \qquad \nabla \cdot \mathbf{u} = 0. \tag{27}$$

The velocity vector is given by $\mathbf{u} = (u, v)^\top$, where u and v are the components in the x and y directions, respectively. Pressure is denoted by p. We integrate the model over time interval $t \in [0, 8]$. Problem setup, geometry, and parameters are taken from the DFG 2D-3 benchmark in the FeatFlow benchmark suite.[55] In the model reduction experiments that follow, we did not explicitly account for the pressure term. This omission is known to be valid for both the transient and periodic regime of the flow.[43,56]

We collected 200 snapshots of a periodic reference simulation at $Re = 100$ in the interval $t \in [4, 5]$ and store each snapshot as a column vector with 292 678 entries. As usual, the snapshot matrix $\mathbf{S}$ is centered by its column-averaged mean value $\mathbf{S}_{\text{ref}}$, and the orthogonal basis vectors are computed by means of the POD. Because the cylinder flow example is periodic, the POD modes can be grouped in pairs $(\mathbf{v}_1, \mathbf{v}_2)$, $(\mathbf{v}_3, \mathbf{v}_4)$, $(\mathbf{v}_5, \mathbf{v}_6)$, $(\mathbf{v}_7, \mathbf{v}_8)$. Figure 12 displays the computed mean flow and the first POD mode for each pair.

We choose the reference state in (3) to represent the mean flow. The flow dynamics can be captured well with only eight modes capturing 99.89% of the snapshot energy. However, physical and mathematical system reduction approaches have revealed that only two modes are *actual* degrees of freedom of the system; the remaining ones are completely dependent on these two.[56] This insight can also be obtained from a nonlinear correlation analysis.[57] Although the POD analysis indicates that eight POD modes should be considered for accurate flow reconstructions, we use instead the proposed nonlinear model reduction framework for learning dynamical-system models that respect the problem's intrinsic dimensionality of 2. Although we could also compute an

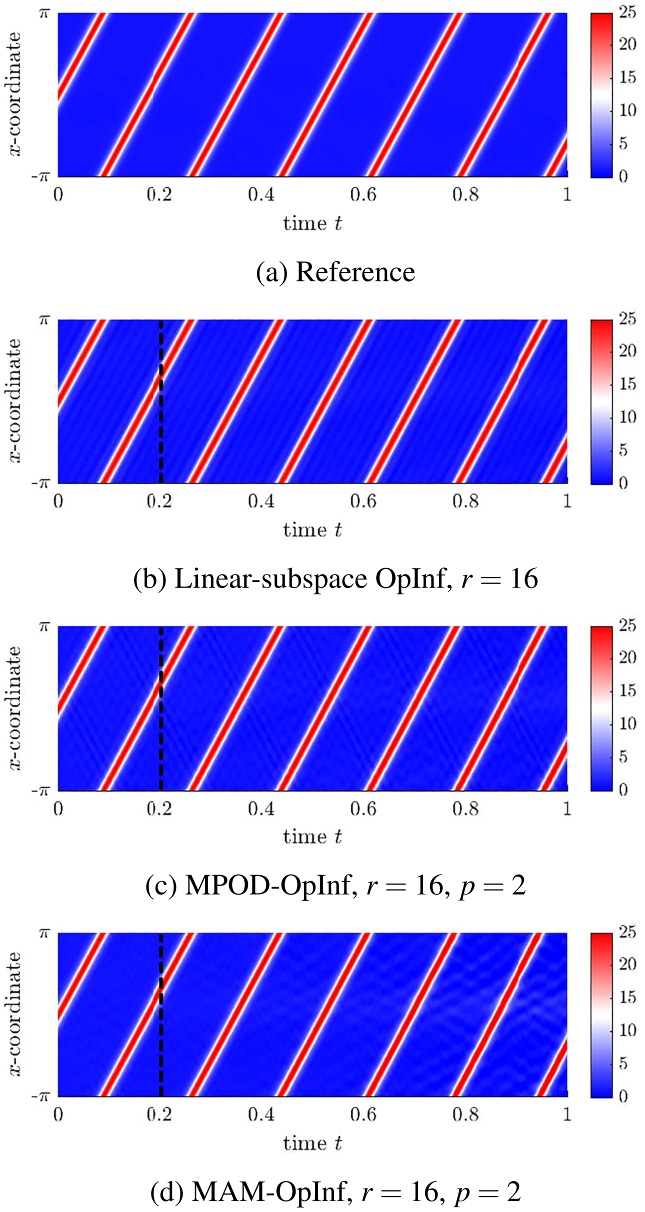


FIG. 9. Plots of the reference solution and OpInf-produced predictions for the Korteweg–de Vries equation over the time window $t \in [0, T]$ at a reduced basis dimension of $r = 16$. The end of the training window is indicated by the dashed line. (a) Reference, (b) linear-subspace OpInf, $r = 16$, (c) MPOD-OpInf, $r = 16$, $p = 2$, and (d) MAM-OpInf, $r = 16$, $p = 2$.

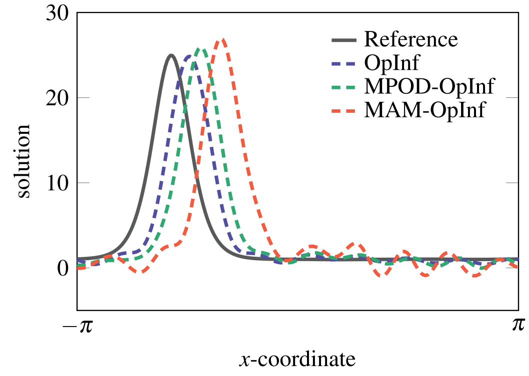


FIG. 10. Reference and OpInf solution snapshots from the Korteweg–de Vries experiment at final time $t = 1$. The plots are given for models of dimension $r = 16$.

orthogonal set of basis vectors through an orthogonal Procrustes problem, as in the alternating minimization-based representation learning problem (see Algorithm 2), the advantages accruing from orthogonal transformations of a POD subspace were found to be negligible: The Procrustes modes were found to be virtually indistinguishable from the ones computed using POD. We thus focus exclusively on the MPOD-OpInf formulation.

Figure 13(a) shows that an OpInf model of size $r = 2$ is unable to capture the periodic nature of the fluid flow. The state error for this model across the training window is 42.9%, while with all eight modes, the error drops to 7.4%. While the model does an excellent job of capturing the transient dynamics in the training regime $t \in [4, 5]$, it fails soon after exiting the training window. For the MPOD-OpInf model, we consider only a reduced basis dimension of $r = 2$, as informed by physical intuition. This means that the remaining six POD modes are contained in the basis matrix associated with the nonlinear part of the state approximation (3), that is, $\overline{\mathbf{V}}$. These approximations are built from quadratic embeddings. (Although high-order embeddings were considered for this

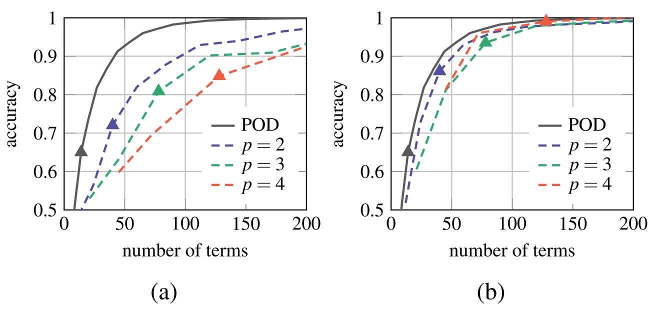


FIG. 11. Cost-accuracy assessment in evaluating reduced-order model (18) for the Korteweg–de Vries problem using (a) the POD-based manifold formulation (Algorithm 1) and (b) the alternating minimization-based manifold formulation (Algorithm 2) as a function of the degree of the polynomial embeddings, p. The results for the different models at a reduced dimensionality of $r = 4$ are indicated by the triangles ($\triangle$).

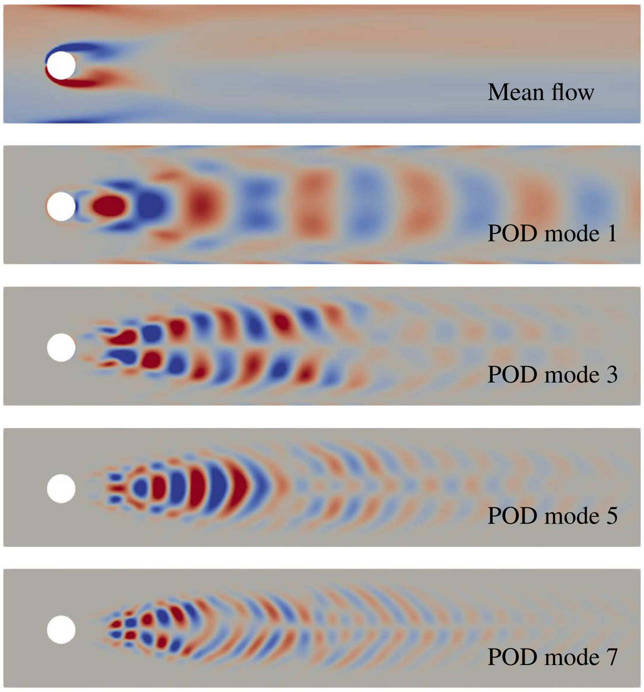


FIG. 12. Mean flow (top) and the dominant odd-numbered POD modes in the cylinder flow problem. We show the vorticity computed from the velocity fields.

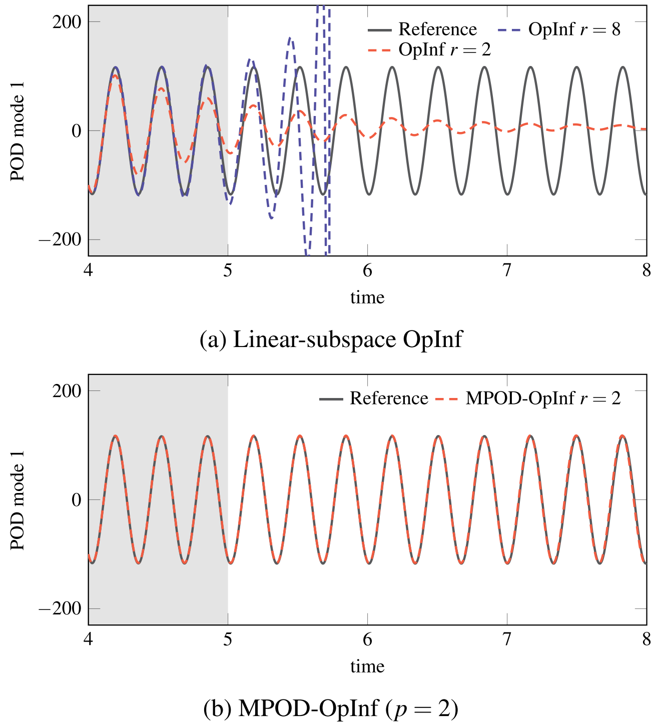


FIG. 13. Comparison of the amplitudes of the first POD mode over time in the OpInf (top) and the MPOD-OpInf (bottom) models in the cylinder flow problem. The manifold models used quadratic embeddings. The gray shaded region highlights the window over which flow snapshots have been collected for training. (a) Linear-subspace OpInf and (b) MPOD-OpInf ($p = 2$).

problem, a polynomial degree of $p = 2$ was found to be sufficient for learning accurate reduced-order models.) The regularization parameter for the representation learning problem was set to $\gamma = 10^{-2}$. The training error for the MPOD-OpInf was found to be 18.6%, which is, as expected, larger than the eight-equation OpInf model. However, the inferred two-equation MPOD-OpInf model, which has a quadratic term in the nonlinear part of the state approximation, was found to be stable well outside of the training regime [see Fig. 13(b)]. The modal amplitudes of the original simulation model are found when the flow data are projected onto the eight POD modes. A comparison in the phase space of the first two coefficients, shown in Fig. 14, finds the MPOD-OpInf model to be accurate and stable with respect to the flow data. Finally, Fig. 15 shows a reconstruction of the flow field in the original state space as predicted at time $t = 8$ compared to the reference solution at the same time step. While some of the finer-scale flow features are not resolved fully, the overall flow dynamics are predicted accurately.

It should be noted that reduced-order models with excellent predictive performance can be obtained by other means. For instance, approaches in which the models are equipped with *linear* structure (such as the dynamic mode decomposition[5–8]) are reported to work well for capturing periodic vortex shedding in the cylinder flow problem.[58] In the work from Baddoo *et al.*,[58] for instance, the data were truncated to the first 15 POD modes. It remains unclear to what extent dynamic mode decomposition and its variants can issue efficient and accurate predictions at or near the *true* dimensionality of two.

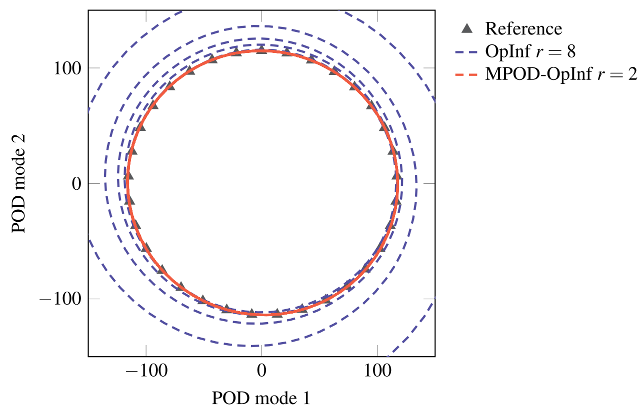


FIG. 14. Limit cycles observed in the full simulation model (△), the eight-equation OpInf model (blue curve), and the two-equation MPOD-OpInf model (red curve) for the cylinder flow problem.

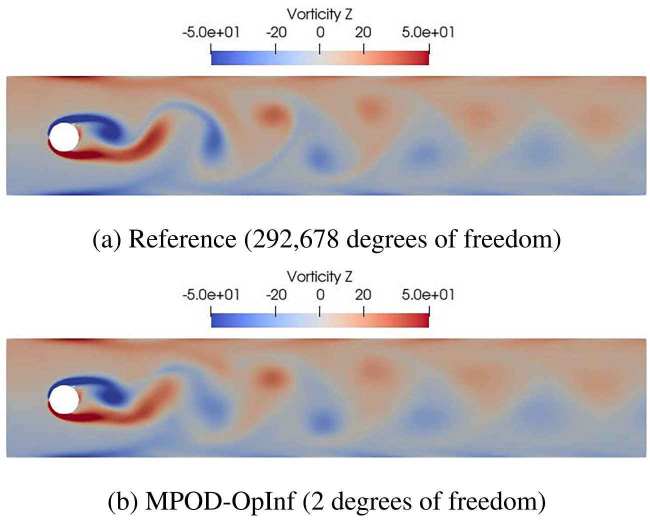


FIG. 15. Flow field (vorticity) predicted at $t = 8$ with a two-equation MPOD-OpInf model using quadratic embeddings (that is, $r = 2$; $p = 2$) for the cylinder flow problem. (a) Reference (292 678 degrees of freedom) and (b) MPOD-OpInf (2 degrees of freedom).

V. CONCLUSION AND DISCUSSION

We have presented a general framework for nonlinear model reduction of large-scale physical systems. We draw on recently developed techniques for constructing nonlinear manifolds of polynomial structure via representation learning in the form of two different learning approaches. First, a POD-based version of the approach is intuitive due to its connection to conventional POD. Second, if one is willing to depart from the interpretable nature of POD methods, alternating minimization techniques can boost model accuracy by means of better approximations to the solution of the general representation learning problem. We then turn to the issue of learning reduced-order models from data. By projecting PDE systems onto the nonlinear manifold, we can identify the algebraic structure of the projection-based reduced-order model. This process calls for careful consideration of the structure of (1) the high-dimensional, physical system and (2) the nonlinear state approximation of choice. The non-intrusive OpInf method was used for learning models directly from time-domain simulation data. Coupling of the two different representation learning approaches with the OpInf framework leads to a set of methods referred to as POD-based manifold OpInf (MPOD-OpInf) and alternating-minimization-based manifold OpInf (MAM-OpInf).

We applied this methodology to the Allen–Cahn equation, the Korteweg–de Vries equation, and the incompressible Navier–Stokes equation. In all numerical experiments, we found the proposed OpInf approaches to be able to circumvent the limitations of linear-subspace OpInf that are due to its use of linear state approximations. The polynomial manifold constructions provide the most benefit in situations where the linear subspace does not accurately represent the full dynamics of the training data. In these situations, the manifold acts as a closure term that accounts for the effects of modes truncated from the linear subspace. The increased accuracy enabled by a nonlinear compression of the state data does not point to computational speedups: The reduced dimensionality comes at the cost of increased algebraic complexity (and thus computational burden) for the manifold-based reduced models. Although the results from Sec. IV D imply that model robustness and predictive performance are important additional considerations to be made, further investigation is needed for better understanding the tradeoffs between dimensionality and complexity.

Further improvements in OpInf reduced-order models may be possible if constraints are introduced to enforce particular mathematical properties of the dynamical system. For example, some classes of problems can be expressed using Hamiltonian or Lagrangian formalisms.[59] Biasing OpInf models toward such structure may enable more accurate long-time predictions far outside the training time interval and will be addressed in future work. In another research direction, data-driven OpInf could be combined with the dynamic training via roll outs of neural ordinary differential equations.[60] This OpInf formulation should be more robust against perturbations in the data because the whole predicted trajectory is considered in the training loss rather than a single time step.

ACKNOWLEDGMENTS

This work has been supported in part by the U.S. Department of Energy AEOLUS MMICC center under Award No. DE-SC0019303, program manager W. Spotz, and by the AFOSR MURI on physics-based machine learning, Award No. FA9550-21-1-0084, program manager F. Fahroo. Stephen Wright was also supported by the National Science Foundation (NSF) under Award Nos. DMS 2023239 and CCF 2224213. Laura Balzano was supported by ARO YIP under Award No. W911NF1910027 and NSF CAREER under award No. CCF-1845076. Karen Willcox would also like to thank Yvon Maday and Albert Cohen for several helpful conversations during the 2022 Leçons Jacques-Louis Lions in Paris, France.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Rudy Geelen: Conceptualization (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Laura Balzano:** Conceptualization (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Stephen Wright:** Conceptualization (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Karen Willcox:** Conceptualization (equal); Funding acquisition (lead); Methodology (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

A Jupyter notebook outlining the representation learning problem for inferring, from data, nonlinear state approximations of the form (3) for the problem from Sec. II C is available at https://github.com/geelenr/nl_manifolds. The notebook features both the POD and alternating minimization-based formulations from Algorithms 1 and 2. The data used in numerical experiments from Sec. IV B and IV C are available upon reasonable request from the authors. The FEniCSx computing platform is used to solve Eq. (23) through their tutorial example.[61,62]

REFERENCES

[1] O. Ghattas and K. Willcox, "Learning physics-based models from data: Perspectives from inverse problems and model reduction," Acta Numer. **30**, 445–554 (2021).
[2] J. L. Lumley, "The structures of inhomogeneous turbulent flow," in *Atmospheric Turbulence and Radio Wave Propagation* (Nauka, Moscow, 1967), pp. 166–178.
[3] L. Sirovich, "Turbulence and the dynamics of coherent structures. I. Coherent structures," Q. Appl. Math. **45**, 561–571 (1987).
[4] P. Holmes, J. L. Lumley, G. Berkooz, and C. W. Rowley, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge University Press, Cambridge, 1996).
[5] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," J. Fluid Mech. **656**, 5–28 (2010).
[6] J. H. Tu, "Dynamic mode decomposition: Theory and applications," Ph.D. thesis (Princeton University, 2013).
[7] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems* (SIAM, 2016).
[8] P. J. Schmid, "Dynamic mode decomposition and its variants," Annu. Rev. Fluid Mech. **54**, 225–254 (2022).
[9] K. Willcox and J. Peraire, "Balanced model reduction via the proper orthogonal decomposition," AIAA J. **40**, 2323–2330 (2002).
[10] C. W. Rowley, "Model reduction for fluids, using balanced proper orthogonal decomposition," Int. J. Bifurcat. Chaos **15**, 997–1013 (2005).
[11] A. Quarteroni, A. Manzoni, and F. Negri, "Reduced basis methods for partial differential equations: An introduction," in *UNITEXT* (Springer International Publishing, 2015).
[12] K. Veroy and A. T. Patera, "Certified real-time solution of the parametrized steady incompressible Navier–Stokes equations: Rigorous reduced-basis a posteriori error bounds," Int. J. Numer. Methods Fluids **47**, 773–788 (2005).
[13] S. Chaturantabut and D. C. Sorensen, "Nonlinear model reduction via discrete empirical interpolation," SIAM J. Sci. Comput. **32**, 2737–2764 (2010).
[14] B. Peherstorfer and K. Willcox, "Data-driven operator inference for nonintrusive projection-based model reduction," Comput. Methods Appl. Mech. Eng. **306**, 196–215 (2016).
[15] E. Qian, B. Kramer, B. Peherstorfer, and K. Willcox, "Lift & learn: Physics-informed machine learning for large-scale nonlinear dynamical systems," Physica D **406**, 132401 (2020).
[16] A. Kolmogoroff, "Uber die beste Annaherung von Funktionen einer gegebenen Funktionenklasse," Ann. Math. **37**, 107–110 (1936).
[17] K. Lee and K. T. Carlberg, "Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders," J. Comput. Phys. **404**, 108973 (2020).
[18] Z. Y. Wan, P. Vlachas, P. Koumoutsakos, and T. Sapsis, "Data-assisted reduced-order modeling of extreme events in complex dynamical systems," PLoS One **13**, e0197704 (2018).
[19] S. Pan and K. Duraisamy, "Data-driven discovery of closure models," SIAM J. Appl. Dyn. Syst. **17**, 2381–2413 (2018).
[20] Y. Kim, Y. Choi, D. Widemann, and T. Zohdi, "A fast and accurate physics-informed neural network reduced order model with shallow masked autoencoder," J. Comput. Phys. **451**, 110841 (2022).
[21] S. Fresca and A. Manzoni, "POD-DL-ROM: Enhancing deep learning-based reduced order models for nonlinear parametrized PDEs by proper orthogonal decomposition," Comput. Methods Appl. Mech. Eng. **388**, 114181 (2022).
[22] S. Jain, P. Tiso, J. B. Rutzmoser, and D. J. Rixen, "A quadratic manifold for model order reduction of nonlinear structural dynamics," Comput. Struct. **188**, 80–94 (2017).
[23] J. Barnett and C. Farhat, "Quadratic approximation manifold for mitigating the Kolmogorov barrier in nonlinear projection-based model order reduction," J. Comput. Phys. **464**, 111348 (2022).
[24] R. Geelen, S. Wright, and K. Willcox, "Operator inference for non-intrusive model reduction with quadratic manifolds," Comput. Methods Appl. Mech. Eng. **403**, 115717 (2023).
[25] J. Axås, M. Cenedese, and G. Haller, "Fast data-driven model reduction for nonlinear dynamical systems," Nonlinear Dyn. **111**, 7941–7957 (2023).
[26] R. Geelen, L. Balzano, and K. Willcox, "Learning latent representations in high-dimensional state spaces using polynomial manifold constructions," arXiv:2306.13748 (2023).
[27] A. Kalur, P. Mortimer, J. Sirohi, R. Geelen, and K. E. Willcox, "Data-driven closures for the dynamic mode decomposition using quadratic manifolds," in *AIAA AVIATION 2023 Forum* (AIAA, 2023), p. 4352.
[28] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1798–1828 (2013).
[29] R. R. Craig and M. C. C. Bampton, "Coupling of substructures for dynamic analyses," AIAA J. **6**, 1313–1319 (1968).
[30] G. Kerschen, M. Peeters, J. Golinval, and A. Vakakis, "Nonlinear normal modes, Part I: A useful framework for the structural dynamicist," Mech. Syst. Signal Process. **23**, 170–194 (2009).
[31] A. K. Noor and J. M. Peters, "Reduced basis technique for nonlinear analysis of structures," AIAA J. **18**, 455–462 (1980).
[32] E. H. Dowell, "Eigenmode analysis in unsteady aerodynamics-reduced-order models," AIAA J. **34**, 1578–1583 (1996).
[33] E. H. Dowell, K. C. Hall, and M. C. Romanowski, "Eigenmode analysis in unsteady aerodynamics: Reduced order models," Appl. Mech. Rev. **50**, 371–386 (1997).
[34] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, "Spectral analysis of nonlinear flows," J. Fluid Mech. **641**, 115–127 (2009).
[35] I. Mezić, "Analysis of fluid flows via spectral properties of the Koopman operator," Annu. Rev. Fluid Mech. **45**, 357–378 (2013).
[36] C. Foias, M. Jolly, I. Kevrekidis, G. R. Sell, and E. Titi, "On the computation of inertial manifolds," Phys. Lett. A **131**, 433–436 (1988).
[37] M. Jolly, I. Kevrekidis, and E. Titi, "Approximate inertial manifolds for the Kuramoto-Sivashinsky equation: Analysis and computations," Physica D **44**, 38–60 (1990).
[38] M. E. Johnson, M. S. Jolly, and I. G. Kevrekidis, "Two-dimensional invariant manifolds and global bifurcations: Some approximation and visualization studies," Numer. Algorithms **14**, 125–140 (1997).
[39] M. D. Graham and I. G. Kevrekidis, "Alternative approaches to the Karhunen-Loeve decomposition for model reduction and data analysis," Comput. Chem. Eng. **20**, 495–506 (1996).
[40] P. Constantin, P. S. Constantin, C. Foias, B. Nicolaenko, and R. Temam, *Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations* (Springer Science & Business Media, 1989), Vol. 70.
[41] B. Nicolaenko, B. Scheurer, and R. Temam, "Some global dynamical properties of a class of pattern formation equations," Commun. Partial Differ. Equ. **14**, 245–297 (1989).
[42] J. Mallet-Paret and G. R. Sell, "Inertial manifolds for reaction diffusion equations in higher space dimensions," J. Am. Math. Soc. **1**, 805–866 (1988).
[43] A. Deane, I. Kevrekidis, G. E. Karniadakis, and S. Orszag, "Low-dimensional models for complex geometry flows: Application to grooved channels and circular cylinders," Phys. Fluids A: Fluid Dyn. **3**, 2337–2354 (1991).
[44] C. Chen, B. He, Y. Ye, and X. Yuan, "The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent," Math. Program. **155**, 57–79 (2016).
[45] P. H. Schönemann, "A generalized solution of the orthogonal procrustes problem," Psychometrika **31**, 1–10 (1966).
[46] C. Gu, "QLMOR: A projection-based nonlinear model order reduction approach using quadratic-linear representation of nonlinear systems," IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. **30**, 1307–1320 (2011).
[47] B. Kramer and K. E. Willcox, "Nonlinear model order reduction via lifting transformations and proper orthogonal decomposition," AIAA J. **57**, 2297–2307 (2019).
[48] P. Benner, P. Goyal, J. Heiland, and I. Pontes Duff, "A quadratic decoder approach to nonintrusive reduced-order modeling of nonlinear dynamical systems," PAMM **23**, e202200049 (2023).

[49] S. A. McQuarrie, C. Huang, and K. E. Willcox, "Data-driven reduced-order models via regularised operator inference for a single-injector combustion process," J. R. Soc. New Zealand **51**, 194–211 (2021).

[50] S. M. Allen and J. W. Cahn, "A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening," Acta Metall. **27**, 1085–1095 (1979).

[51] H. Song, L. Jiang, and Q. Li, "A reduced order method for Allen–Cahn equations," J. Comput. Appl. Math. **292**, 213–229 (2016).

[52] C. Dechanubeksa and S. Chaturantabut, "An application of a modified gappy proper orthogonal decomposition on complexity reduction of Allen-Cahn equation," Algorithms **13**, 148 (2020).

[53] P. Benner and T. Breiten, "Two-sided projection methods for nonlinear model order reduction," SIAM J. Sci. Comput. **37**, B239–B260 (2015).

[54] A. Mendible, S. L. Brunton, A. Y. Aravkin, W. Lowrie, and J. N. Kutz, "Dimensionality reduction and reduced-order modeling for traveling wave physics," Theor. Comput. Fluid Dyn. **34**, 385–400 (2020).

[55] See https://www.mathematik.tu-dortmund.de/~featflow/en/benchmarks/ff_benchmarks.html for more details on the FEATFLOW benchmark problems.

[56] B. R. Noack, K. Afanasiev, M. Morzyński, G. Tadmor, and F. Thiele, "A hierarchy of low-dimensional models for the transient and post-transient cylinder wake," J. Fluid Mech. **497**, 335–363 (2003).

[57] J.-C. Loiseau, S. L. Brunton, and B. R. Noack, *From the POD-Galerkin Method to Sparse Manifold Models* (De Gruyter, Berlin, 2021), Chap. 9, pp. 279–320.

[58] P. J. Baddoo, B. Herrmann, B. J. McKeon, J. Nathan Kutz, and S. L. Brunton, "Physics-informed dynamic mode decomposition," Proc. R. Soc. A **479**, 20220576 (2023).

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

[60] W. I. T. Uy, D. Hartmann, and B. Peherstorfer, "Operator inference with roll outs for learning reduced models from scarce and low-quality data," Comput. Math. Appl. **145**, 224–239 (2023).

[61] M. Alnæs *et al.*, The FEniCS project version 1.5. *Archive of Numerical Software* **3** (2015).

[62] A. Logg, K.-A. Mardal, and G. Wells, *Automated Solution of Differential Equations by the Finite Element Method: The FEniCS Book* (Springer Science & Business Media, 2012), Vol. 84.