Localized non-intrusive reduced-order modeling in the operator inference framework

Rudy Geelen* Karen Willcox*

November 5, 2021

Abstract

This paper presents data-driven learning of localized reduced models. Instead of a global reduced basis, the approach employs multiple local approximation subspaces. This localization permits adaptation of a reduced model to local dynamics, thereby keeping the reduced dimension small. This is particularly important for reduced models of nonlinear systems of partial differential equations, where the solution may be characterized by different physical regimes or exhibit high sensitivity to parameter variations. The contribution of this paper is a non-intrusive approach that learns the localized reduced model from snapshot data using operator inference. In the offline phase, the approach partitions the state space into subregions and solves a regression problem to determine localized reduced operators. During the online phase, a local basis is chosen adaptively based on the current system state. The non-intrusive nature of localized operator inference makes the method accessible, portable, and applicable to a broad range of scientific problems, including those that use proprietary or legacy high-fidelity codes. We demonstrate the potential for achieving large computational speedups while maintaining good accuracy for a Burgers' equation governing shock propagation in a one-dimensional domain and a phase-field problem governed by the Cahn-Hilliard equation.

1 Introduction

The complexity of today's models and numerical simulations in computational science and engineering has made the acceleration of model evaluations essential in many design and optimization applications. Often, the trajectory of full-model states can be approximated in a low-dimensional subspace whose dimension is orders of magnitude lower than the dimension of the high-fidelity model itself. With the use of *localization* techniques, this dimension can be reduced even further by constructing a set of local subspaces which are each adapted to a particular region of the solution space. Existing localization-based model reduction methods are powerful, yet they are intrusive, requiring access to (and often modifications of) the high-fidelity model source code. There is a growing recognition of the importance of *non-intrusive* methods. These approaches aim to strike a good balance between ease of implementation in situations where an intrusive reduced model is unavailable, and exploiting knowledge of the underlying problem to recover certain convergence and stability guarantees [1]. This paper presents a new non-intrusive localized model reduction approach that derives localized reduced space operators using the operator inference framework [2]. Our approach applies to nonlinear partial differential equations (PDEs) with low-order polynomial nonlinear terms and is fully driven by simulation data.

Projection-based reduced-order modeling identifies an appropriate low-dimensional subspace onto which the governing equations are projected. This defines a reduced-order model (ROM) that has many fewer degrees of freedom than the original full-order model (FOM). A popular choice for constructing the reduced-order basis vectors is the proper orthogonal decomposition (POD), which constructs a basis from representative solutions (referred to as *snapshots*) of the system taken at various locations of the state space [3, 4, 5]. The POD basis is comprised of the leading principal components of the snapshot data. Often, the dynamics of the underlying equations can be represented by a small number of POD modes. In other situations, the

*Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX (rudy.geelen@austin.utexas.edu, kwillcox@oden.utexas.edu, https://kiwi.oden.utexas.edu)

solution of the governing PDEs lies in a manifold that cannot be covered effectively by a single ROM without significantly increasing its dimension, thus countering the computational efficiency gains. Linear dimensionality reduction is not always suitable for problems whose underlying governing equations are characterized by local solution features that evolve with time (see for example [6, 7, 8]). This includes phase-field models in materials science and engineering, the driving application behind the present study, which are notoriously challenging to solve with traditional model order reduction techniques [9].

To address this challenge, localization-based techniques have been developed in which *multiple* local approximation spaces are tailored to different regions of the state-space. The construction of multiple local reduced-order models can offer substantial improvements in the efficiency of applications that otherwise require an impractical number of POD modes. Broadly speaking, local subspaces can be defined in the parameter space [10, 11, 12, 13, 14], in time [15, 16], or in or state-space [17, 14, 18, 19, 20]. Physics-informed clustering strategies form the basis of another set of approaches for the construction of localized ROMs. The goal of physics-informed clustering is to identify clusters that are appropriate for model reduction purposes, through physical quantities corresponding to the underlying problem. Clustering directly in the solution space can be achieved by grouping snapshots into a set of clusters using unsupervised learning methods according to a predefined similarity metric. A variation of this can also be found in the projection error-based local ROM [21], where scale-invariances of the underlying model are captured by using a true projection error as a dissimilarity criterion instead of conventional Euclidean distances. In [22] a POD analysis of the numerical data is followed by the clustering procedure to reduce the computational and input load of the cluster analysis in the POD subspace. In the parameter-based version of the localized discrete empirical interpolation method (DEIM) from [14], a clustering approach is used to define clusters in the solution space by means of local DEIM residuals. More recently, the concept of dictionary-based ROM-nets is introduced, where deep neural networks recommend a suitable local ROM from a dictionary [23]. It was found that the particular physics-informed clustering analysis has a strong impact on the performance of the local reduced models, after which the authors propose a novel dissimilarity measure for clustering purposes [24].

Despite the demonstrated effectiveness of localization approaches in the field of reduced-order modeling, existing approaches are intrusive in nature: they require access to the high-fidelity code in order to compute the projected reduced-order operators. This limits the applicability of the approaches, since source-code access and modification is often impractical or impossible. The contribution of our work is to combine the aforementioned state-of-the-art localization techniques with a non-intrusive model reduction formulation, leading to a flexible framework for reduction of large-scale nonlinear problems. We note that the nonintrusivity property does not imply that the simulation process is considered to be a black-box: this work develops a localization approach based on the operator inference framework [2] and embeds knowledge of the governing equations, their boundary/initial conditions, and their parametric nature. The additional dimension reduction achieved through localization has the added benefit of addressing the numerical ill-conditioning of the regression problem in operator inference approaches for non-intrusive model reduction [1]. Our work differs from nonlinear system identification techniques that seek to discover the underlying physics of a dynamical system from data [25, 26], since we target here the learning of a ROM for problems where the governing equations are known and the goal is to approximate a trusted high-fidelity simulation. We also distinguish between the proposed approach and *calibration* methods in fluid mechanics and turbulence modeling, which aim to improve the accuracy of POD-based reduced-order models through the solution of optimization problems [27, 28]. The interested reader is referred to [29] for a comparison of various methods.

The remainder of this paper is organized as follows. In Section 2 the problem setting of model reduction for nonlinear dynamical systems is described and the baseline operator inference method is discussed. Section 3 presents the new localized operator inference approach, discussing how local reduced-order bases are introduced within the scope of operator inference by utilizing data clustering and classification techniques. Its application to two nonlinear problems is discussed in Section 4. Finally, concluding remarks are offered in Section 5.

2 Non-intrusive reduced-order modeling with operator inference

This section begins with introducing the high-dimensional dynamical systems of interest. We then briefly summarize the baseline non-intrusive operator inference approach [2], which is the main building block for our new localized method, in which reduced models are learned purely from numerical training data generated by solving the high-dimensional dynamical system. The section concludes with some important numerical considerations of operator inference approaches.

2.1 Dynamical system with high-dimensional states

This work considers problems governed by systems of nonlinear PDEs. Consider the governing equations of the system of interest written, after spatial discretization, in semi-discrete form. This results in a high-dimensional system of ordinary differential equations (ODEs), since the order of the spatial discretization is typically at least thousands to millions. We consider a particular form of the semi-discrete equations with a nonlinear term of the second order in the state,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{s}(t) = \mathbf{C} + \mathbf{A}\mathbf{s}(t) + \mathbf{H}\left(\mathbf{s}(t) \otimes \mathbf{s}(t)\right) + \mathbf{B}\mathbf{u}(t); \quad \mathbf{s}(0) = \mathbf{s}_0. \tag{1}$$

Here, the state dimension is $n \in \mathbb{N}$ and the semi-discrete state at time $t \in [0, T]$ is $\mathbf{s}(t) \in \mathbb{R}^n$, with specified initial condition $\mathbf{s}_0 \in \mathbb{R}^n$. Throughout this paper (1) will be referred to as the *full-order* model (FOM). We note its polynomial structure: the vector $\mathbf{C} \in \mathbb{R}^n$ are the constant terms of the system of differential equations; and the matrix operators $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{H} \in \mathbb{R}^{n\times n^2}$ are the terms that are linear and quadratic in the state $\mathbf{s}(t)$, respectively. There are $m \in \mathbb{N}$ inputs contained in the input vector $\mathbf{u} \in \mathbb{R}^m$ that enter linearly via the input matrix $\mathbf{B} \in \mathbb{R}^{n\times m}$, typically encoding boundary conditions or forcing terms. Here, $\otimes$ denotes the Kronecker product, following the notation from [30], which for a column vector $\mathbf{s} = [s_1, s_2, ..., s_n]^\top$ is given by $\mathbf{s} \otimes \mathbf{s} = \left[s_1^2, s_1 s_2, \ldots, s_1 s_n, s_2 s_1, s_2^2, \ldots, s_2 s_n, \ldots, s_n^2\right]^\top \in \mathbb{R}^{n^2}$. When a polynomial structure is not an attribute of the underlying governing equations, variable transformations and auxiliary variables may be employed to expose structure that is amenable to projection-based approximation. It was shown in [31] that many nonlinear dynamical systems may be lifted to the quadratic form of (1). Similarly, it has been shown in [32, 33, 34, 35] that a number of PDE models that appear in complex engineering systems can be lifted to this quadratic form.

2.2 Learning low-dimensional reduced models via operator inference

Consider the task of reducing high-dimensional model (1). A projection-based ROM can be obtained by restricting the state to lie in a low-dimensional subspace spanned by a set of basis vectors through, for instance, the POD. The POD basis is computed empirically from training data in the form of solution snapshots and applies to both linear and nonlinear systems. To compute the POD basis, consider a set of k snapshots, $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_k$, which are state solutions computed at different time instances and/or parameter values. We define the snapshot matrix $\mathbf{S} \in \mathbb{R}^{n\times k}$ whose jth column is the snapshot $\mathbf{s}_j \in \mathbb{R}^n$. The POD basis vectors are the left singular vectors of $\mathbf{S}$ corresponding to the r largest singular values. They populate the columns of the basis matrix $\mathbf{V}$. The column space of $\mathbf{V}$ defines an r-dimensional subspace of the full state space $\mathbb{R}^n$, where, generally speaking, $r \ll n$. The POD basis is orthonormal, that is $\mathbf{V}^\top\mathbf{V} = \mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix of dimension $r \times r$. Approximating the high-dimensional state $\mathbf{s}$ in the low-dimensional basis $\mathbf{V}$, we write

$$\mathbf{s}(t) \approx \mathbf{V}\widehat{\mathbf{s}}(t), \tag{2}$$

where $\widehat{\mathbf{s}} \in \mathbb{R}^r$ is the r-dimensional reduced state.

Classical model reduction techniques use an intrusive process to compute a set of reduced matrix operators via projections of the FOM operators, which typically involves intrusive queries to the FOM source code. In contrast, the operator inference approach from [2] is non-intrusive in nature and permits the reduced operators to be learned from data without the need for access to FOM operators. We define the operator inference ROM of Eq. (1) as

$$\frac{\mathrm{d}}{\mathrm{d}t}\widehat{\mathbf{s}}(t) = \widehat{\mathbf{C}} + \widehat{\mathbf{A}}\widehat{\mathbf{s}}(t) + \widehat{\mathbf{H}}\left(\widehat{\mathbf{s}}(t) \otimes \widehat{\mathbf{s}}(t)\right) + \widehat{\mathbf{B}}\mathbf{u}(t); \quad \widehat{\mathbf{s}}(0) = \mathbf{V}^\top\mathbf{s}_0, \tag{3}$$

where $\widehat{\mathbf{C}} \in \mathbb{R}^r$, $\widehat{\mathbf{A}} \in \mathbb{R}^{r\times r}$, $\widehat{\mathbf{H}} \in \mathbb{R}^{r\times r^2}$, and $\widehat{\mathbf{B}} \in \mathbb{R}^{r\times m}$ are the reduced operators determined via a data-driven regression problem.

In the following least-squares learning problem we will work with a quadratic system as in (1) and (3). However, operator inference readily extends to systems with cubic and higher-order polynomial terms, noting that as the polynomial order of the system increases, so too does the number of coefficients to be inferred in the corresponding operator inference regression problem. Given k reduced state snapshots, the corresponding reduced-order time derivative data, and a postulated model form, the operator inference approach formulates a regression problem to find the reduced matrix operators that yield the reduced model that best matches the given projected snapshot data. For a model with quadratic form, this leads to the following minimization problem:

$$\left(\widehat{\mathbf{C}}, \widehat{\mathbf{A}}, \widehat{\mathbf{H}}, \widehat{\mathbf{B}}\right) = \underset{\widehat{\mathbf{C}},\widehat{\mathbf{A}},\widehat{\mathbf{H}},\widehat{\mathbf{B}}}{\arg\min} \sum_{j=1}^{k} \left\| \widehat{\mathbf{C}} + \widehat{\mathbf{A}}\widehat{\mathbf{s}}_j + \widehat{\mathbf{H}}\left(\widehat{\mathbf{s}}_j \otimes \widehat{\mathbf{s}}_j\right) + \widehat{\mathbf{B}}\mathbf{u}_j - \dot{\widehat{\mathbf{s}}}_j \right\|_2^2, \tag{4}$$

with $\dot{\widehat{\mathbf{s}}}_j$ the representation, in the POD coordinates, of the time derivative of the jth snapshot, and $\mathbf{u}_j$ the input data $\mathbf{u}_j := \mathbf{u}(t_j)$ at time t_j. It has been shown in [2] that Eq. (4) decouples into r independent linear least-squares problems that solve for each row of $\widehat{\mathbf{C}}, \widehat{\mathbf{A}}, \widehat{\mathbf{H}}$ and $\widehat{\mathbf{B}}$ independently. Ref. [2] also shows that as the time step $\Delta t \to 0$ and reduced space dimension $r \to n$, the inferred operators of (4) will converge to those obtained from intrusive projection.

2.3 Numerical considerations

While the least-squares optimization problem is linear in the coefficients of the unknown reduced operators, the number of coefficients to be inferred grows exponentially with higher-order polynomial terms in the inference problem. For a ROM in which the system dynamics are quadratic, inferred using k state and time derivative pairs, we require $k > 1 + r + r(r+1)/2 + m$ to avoid an underdetermined least-squares problem [36, §5.3]. In addition, the snapshot data must be sufficiently rich to yield a well conditioned problem (4). For systems with higher-order polynomial state dependence the number of coefficients to be inferred in each least-squares problem grows even more rapidly. For a fixed number of basis functions r, this leads to an increase of training costs in the offline stage of the approach. It also becomes increasingly difficult to avoid conditioning issues, even for a full-rank linear least squares problem (4). A large condition number typically originates from linear dependencies of the states at different time steps and/or parameter variations and can introduce significant errors into the numerical solutions. In [2] a data collection strategy for avoiding ill-conditioning was proposed.

Operator inference problems can also be noisy due to error in the numerically estimated time derivatives, model misspecification (e.g., if the system is not truly quadratic), and unresolved system dynamics due to the truncated POD modes. The ROM constructed through Eq. (4) can thus suffer from overfitting the operators to the data, resulting in poor performance. To this end, [37] developed a regularized operator inference formulation that balances residual minimization with a Tikhonov regularization term. This yields a regularized operator inference problem of the form

$$\begin{aligned}\left(\widehat{\mathbf{C}}, \widehat{\mathbf{A}}, \widehat{\mathbf{H}}, \widehat{\mathbf{B}}\right) = \underset{\widehat{\mathbf{C}},\widehat{\mathbf{A}},\widehat{\mathbf{H}},\widehat{\mathbf{B}}}{\arg\min} &\sum_{j=1}^{k} \left\| \widehat{\mathbf{C}} + \widehat{\mathbf{A}}\widehat{\mathbf{s}}_j + \widehat{\mathbf{H}}\left(\widehat{\mathbf{s}}_j \otimes \widehat{\mathbf{s}}_j\right) + \widehat{\mathbf{B}}\mathbf{u}_j - \dot{\widehat{\mathbf{s}}}_j \right\|_2^2 \\ &+ \lambda_1 \left(\|\widehat{\mathbf{C}}\|_2^2 + \|\widehat{\mathbf{A}}\|_F^2 + \|\widehat{\mathbf{B}}\|_F^2 \right) + \lambda_2 \|\widehat{\mathbf{H}}\|_F^2,\end{aligned} \tag{5}$$

where $\lambda_1, \lambda_2 \geq 0$ are scalar regularization hyperparameters and $\|\cdot\|_F$ denotes the Frobenius norm. The values of these hyperparameters are chosen in an outer optimization loop that also accounts for reduced model stability, as discussed in [37]. We refer to [38] for a recent discussion on the stability of learned models with operator inference. In that work, the authors propose a regularizer that explicitly leverages the quadratic model form to penalize unstable models. It is shown that additional physical insights, in the form of structure of the linear dynamics, can be imposed on the operator-inference models via constraints.

3 Operator inference using local reduced-order bases

The operator inference approach from [2] employs a single *global* basis with span encompassing the desired range of dynamics. Ensuring that a reduced-order representation is sufficiently rich to yield accurate predictions over the desired range of parameter values is often challenging: the performance of a projection-based ROM hinges on the ability of the underlying subspaces to accurately represent nonlinear trajectories in the state space. Rather than constructing a single linear subspace to approximate the underlying differential equations, we compute several local subspaces which are each tailored to a particular region of the solution space. The distinguishing feature in the proposed approach is that the set of local reduced models are now learned via operator inference in completely non-intrusive fashion. We refer to this approach as *localized operator inference*.

In the next subsection we introduce our localized operator inference method in step-by-step fashion. We then specify the individual offline and online components of the proposed approach used in the learning problem. A cost analysis of the construction of ROMs with localized operator inference concludes the section.

3.1 Localized operator inference

Classical model reduction follows a decomposition of computational tasks into two phases: (1) the offline phase where the reduced model is derived from the high-fidelity model, and (2) the online phase where the reduced model is deployed to compute, in the low-dimensional manifold, the solution that best captures the dynamics of the high-fidelity solution. The offline stage of localized operator inference approach consists of the following consecutive steps:

1. **Data collection & clustering** – The training data is collected and partitioned into groups of kinematically similar observations using unsupervised learning methods. The clustering algorithm returns n_p partitions of the data matrix.

2. **Training the classifier** – We train the classifier. Its purpose is to recommend a good local ROM with respect to a predefined indicator.

3. **Learning a set of cluster-specific ROMs** – Using the operator inference approach from Section 2 we learn a set of cluster-specific system operators, one associated with each snapshot partition p. This leads to a set of n_p local ROMs that operate independently of one another.

The output of the offline stage contains a set of cluster-specific ROMs and the classifier. The online phase of the localized operator inference method consists of the following steps:

4. **Selection of ROM** – The indicator is computed and, using the classifier constructed in the offline stage, the most appropriate local ROM is identified.

5. **ROM evaluation** – We evaluate the learned ROM using the reduced model operators corresponding to the selected local ROM.

The workflow of the localized operator inference method is sketched in Figure 1. We now specify the different aspects of the offline and online stages of the proposed approach.

3.2 Localized operator inference: the offline phase

3.2.1 Data collection & clustering

In the operator inference framework, the goal is to learn a ROM of the form (3) directly from data. Let $\mathbf{s}_1, \ldots, \mathbf{s}_k$ be the solutions of the FOM (1) at time steps $t_1, \ldots, t_k$ computed with a time stepping scheme and initial condition $\mathbf{s}_0$. The inputs $\mathbf{u}(t_0), \ldots, \mathbf{u}(t_k)$ are sampled at the same time steps. Then, the snapshot and time derivative data matrices and input trajectory are defined as

$$\mathbf{S} = \begin{bmatrix} | & | & & | \\ \mathbf{s}_0 & \mathbf{s}_1 & \cdots & \mathbf{s}_k \\ | & | & & | \end{bmatrix}, \quad \dot{\mathbf{S}} = \begin{bmatrix} | & | & & | \\ \dot{\mathbf{s}}_0 & \dot{\mathbf{s}}_1 & \cdots & \dot{\mathbf{s}}_k \\ | & | & & | \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} | & | & & | \\ \mathbf{u}(t_0) & \mathbf{u}(t_1) & \cdots & \mathbf{u}(t_k) \\ | & | & & | \end{bmatrix}. \tag{6}$$

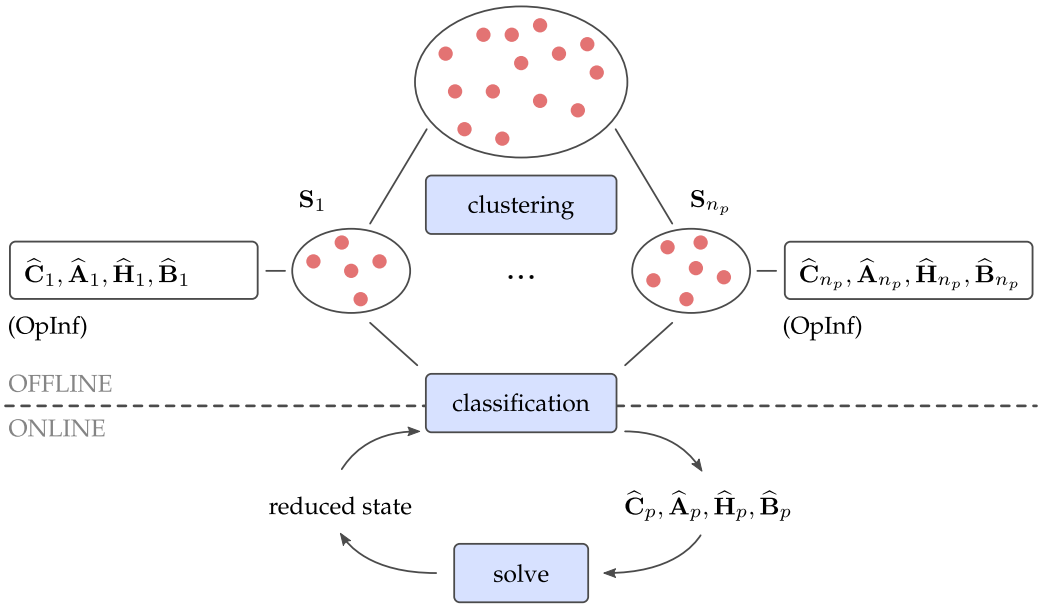


Figure 1: Localized operator inference workflow. In the offline phase a clustering approach is employed to partition the snapshot data into n_p clusters. A cluster-specific set of reduced state operators is then inferred using the operator inference approach from [2]. A classifier is trained to learn the mapping between the states and their appropriate partition p which, during the online stage, selects the optimal local ROM with respect to the current state of the system.

We now consider the partitioning of state and time derivative pairs in the data set. For high-dimensional data, effects attributed to the curse of dimensionality hinder the successful application of many traditional clustering algorithms. Most notably, the assessment of similarity using distance is no longer a meaningful criterion for distinguishing between data objects that should be part of the same cluster or that should be separated into distinct clusters [39]. Instead of directly clustering the high-dimensional snapshots, we follow the procedure from [22] in first projecting the snapshot data onto a lower dimension subspace. Such an approach improves the stability of clustering approaches, because we cluster points in a low-dimensional subspace and not in the high-dimensional space, and it also enables the handling of large numbers of clusters [14]. For this clustering step, we compute a global POD basis denoted $\overline{\mathbf{V}} \in \mathbb{R}^{n \times r_G}$, where r_G is the dimension of this global basis. We obtain the representation of each snapshot $\mathbf{s}_j$ in these global POD coordinates as $\widetilde{\mathbf{s}}_j = \overline{\mathbf{V}}^\top \mathbf{s}_j$, or

$$\widetilde{\mathbf{S}} = \overline{\mathbf{V}}^\top \mathbf{S}, \tag{7}$$

where the tilde denotes the reduced-dimension state representation in the global POD basis $\overline{\mathbf{V}}$. Note that these global POD coordinates are used only for the clustering and classification steps, not for defining the reduced models themselves. The global basis dimension $r_G \ll n$ is chosen such that $\widetilde{\mathbf{S}}$ contains only the most important information about the snapshots. It is noted that only the snapshot matrix, and not the time derivative data matrix, is expressed in the POD coordinate system as we choose to drive the clustering algorithm by similarities in the states.

The clustering algorithm is tasked with partitioning the snapshot data in $\widetilde{\mathbf{S}}$ with respect to a certain criterion. The input to the clustering method is the set $\widetilde{\mathbf{S}}$, and the output is a partition $\{\widetilde{\mathbf{S}}_1, \widetilde{\mathbf{S}}_2, \ldots, \widetilde{\mathbf{S}}_{n_p}\}$, denoting n_p individual clusters. The k-means algorithm is a standard clustering algorithm that has been applied successfully to many different applications in the context of the unsupervised learning methods for model reduction [17, 18, 20, 21, 22]. It is a *hard* clustering algorithm: each data point is assigned to exactly one cluster. While these methods are widely used in practice, they ignore any measure of data similarity across clusters. In this work, a *fuzzy* c-means (FCM) algorithm is adopted as the clustering method [40, 41, 42]. In contrast to k-means and other hard clustering methods, data is no longer divided into distinct

clusters. Each data point can potentially belong to multiple clusters as a result of the varying degree of membership. By iteratively updating the cluster centers and the membership grades for each data point, the FCM algorithm moves the cluster centers to the correct location within a data set. It returns a list of n_p cluster centers, $\{\mathbf{c}_1, \ldots, \mathbf{c}_{n_p}\}$, and a partition matrix, $\boldsymbol{\mu}$, where each element expresses the degree to which reduced snapshot $\widetilde{\mathbf{s}}_j$ belongs to cluster $\mathbf{c}_p$. The algorithm minimizes an objective function that represents the distance from any given data point to a cluster center weighted by the membership of that data point in the cluster

$$\sum_{j=1}^{k}\sum_{p=1}^{n_p} \mu_{jp}^{q} \left\|\widetilde{\mathbf{s}}_j - \mathbf{c}_p\right\|^2, \tag{8}$$

where

$$\mu_{jp} = \frac{1}{\displaystyle\sum_{l=1}^{n_p} \left(\frac{\|\widetilde{\mathbf{s}}_j - \mathbf{c}_p\|}{\|\widetilde{\mathbf{s}}_j - \mathbf{c}_l\|}\right)^{\frac{2}{q-1}}}. \tag{9}$$

The fuzzy partition matrix exponent, controlling the softness of the overlap of the reduced snapshot clusters, is given by $q > 1$ and the total number of reduced snapshots equals k. The cluster centers are computed as

$$\mathbf{c}_p = \sum_{j=1}^{k} \mu_{jp}^{q} \widetilde{\mathbf{s}}_j \Bigg/ \sum_{j=1}^{k} \mu_{jp}^{q}, \tag{10}$$

which are points in $\mathbb{R}^{r_G}$. Generally speaking, the total number of data points assigned to each cluster will vary. In the context of localization approaches for model reduction, FCM eliminates the need for auxiliary algorithms to introduce overlap between the generated state clusters to avoid the formation of gaps between the generated clusters. Such techniques promote the continuity of the approximation in the transitional regions of the state space [17, 18]. A data point can be assigned to a given cluster solely through its degree of membership. This is a straightforward task as, for a given projected snapshot, $\widetilde{\mathbf{s}}_i$, the sum of the membership values for all clusters is one.

3.2.2 Training the classifier

The use of localized ROMs introduces the need for a model selection method that identifies the most suitable ROM. When model selection during the online phase is not straightforward one can train a classifier to learn the model selection task and enable fast model recommendation [14, 23, 43]. Having obtained a partition of $\widetilde{\mathbf{S}}$, we train the classifier $\Theta : \mathcal{Z} \to \{1, 2, \ldots, n_p\}$ to select the most appropriate local ROM with respect to a predefined indicator $\mathbf{z} \in \mathcal{Z}$. In machine learning terms this ubiquitous task is called classification [44]. The task of constructing the indicator $\mathbf{z}$ is also called feature *extraction* or feature *selection* [45]. It can be viewed as applying a mask to a high-dimensional vector to obtain a low-dimensional vector containing the most relevant information. The purpose of an indicator is to describe the high-dimensional state well enough to decide which local ROM should be used. For this reason we use the projected snapshot data partitions, $\widetilde{\mathbf{S}} = \{\widetilde{\mathbf{S}}_1, \widetilde{\mathbf{S}}_2, \ldots, \widetilde{\mathbf{S}}_P\}$, whose columns denote the solution snapshots represented in the global POD coordinate system.

Many classification methods are available in the literature. Here we employ a nearest neighbor classifier with an Euclidean distance metric. The classifier $\Theta : \mathcal{Z} \to \{1, 2, \ldots, n_p\}$ with $\mathcal{Z} \in \mathbb{R}^{r_G}$ is trained on the reduced snapshot data $\widetilde{\mathbf{s}}_1, \ldots, \widetilde{\mathbf{s}}_k$ from the clustering algorithm and their assigned cluster index p.

3.2.3 Learning a set of cluster-specific ROMs

The proposed method learns a set of cluster-specific reduced models by constructing multiple local bases throughout the parameter space. We consider again the quadratic FOM, as in (1). Given a partitioning of $\widetilde{\mathbf{S}}$, and thus of $\mathbf{S}$, we compute the thin singular value decomposition in every pth snapshot cluster via

$$\mathbf{S}_p = \boldsymbol{\Phi}_p \boldsymbol{\Sigma}_p \boldsymbol{\Psi}_p^{\top}; \quad p = 1, \ldots, n_p, \tag{11}$$

where $\boldsymbol{\Phi}_p \in \mathbf{R}^{n \times k_p}, \boldsymbol{\Sigma}_p \in \mathbf{R}^{k_p \times k_p}$, and $\boldsymbol{\Psi}_p \in \mathbf{R}^{k_p \times k_p}$. We let k_p denote the number of snapshots collected in $\mathbf{S}_p$. The singular values $\sigma_{p,1} \geq \sigma_{p,2} \geq \cdots \geq \sigma_{p,k_p} \geq 0$ in $\boldsymbol{\Sigma}_p$, with $\sigma_{p,j}$ denoting the jth singular value of

snapshot cluster p, are given in non-increasing order. The local POD basis vectors $\mathbf{V}_p$ are the first r_p left singular vectors of $\mathbf{S}_p$. The size of the POD basis is typically chosen using the singular values as a guideline. In the context of localized ROMs, this means that one could use different local basis sizes for each cluster.

Note that, in practice, one must be careful about the scaling of the snapshots in order to obtain an adequate basis. Centering the data can be achieved through mean-subtraction, for instance. This is particularly important when dealing with coupled problems in which the different physical variables are measured by different scales [1].

In the operator inference framework, the goal is to learn a ROM for a postulated model form directly from data [2]. In the proposed localized operator inference approach, we will learn n_p localized ROMs. For each $p = 1, \ldots, n_p$, one starts with the trajectories of $\mathbf{s}(t)$ and $\dot{\mathbf{s}}(t)$ projected onto the local basis matrix $\mathbf{V}_p \in \mathbb{R}^{n \times r_p}$:

$$\widehat{\mathbf{S}}_p = \mathbf{V}_p^\top \mathbf{S}_p, \quad \widehat{\dot{\mathbf{S}}}_p = \mathbf{V}_p^\top \dot{\mathbf{S}}_p; \quad p = 1, \ldots, n_p. \tag{12}$$

The localized operator inference optimization problem to compute the localized reduced operators $\widehat{\mathbf{C}}_p, \widehat{\mathbf{A}}_p, \widehat{\mathbf{H}}_p$ and $\widehat{\mathbf{B}}_p$ from the above projected data is:

$$\begin{aligned}\left(\widehat{\mathbf{C}}_p, \widehat{\mathbf{A}}_p, \widehat{\mathbf{H}}_p, \widehat{\mathbf{B}}_p\right) = \underset{\widehat{\mathbf{C}}_p, \widehat{\mathbf{A}}_p, \widehat{\mathbf{H}}_p, \widehat{\mathbf{B}}_p}{\arg\min} \Bigg(& \left\| \widehat{\mathbf{C}}_p \mathbf{1}_{k_p} + \widehat{\mathbf{A}}_p \widehat{\mathbf{S}}_p + \widehat{\mathbf{H}}_p \left(\widehat{\mathbf{S}}_p \odot \widehat{\mathbf{S}}_p \right) + \widehat{\mathbf{B}}_p \mathbf{U}_p - \widehat{\dot{\mathbf{S}}}_p \right\|_F^2 \\ & + \lambda_{1,p} \left(\|\widehat{\mathbf{C}}_p\|_F^2 + \|\widehat{\mathbf{A}}_p\|_F^2 + \|\widehat{\mathbf{B}}_p\|_F^2 \right) + \lambda_{2,p} \|\widehat{\mathbf{H}}_p\|_F^2 \Bigg); \quad p = 1, \ldots, n_p,\end{aligned} \tag{13}$$

associated with the pth snapshot cluster, where $\mathbf{1}_{k_p} \in \mathbb{R}^{k_p}$ is a k_p-length row vector with all entries set to unity, operator $\odot$ denotes the Khatri-Rao product of two matrices (also known as column-wise Kronecker product [30]), and $\lambda_{1,p}, \lambda_{2,p}$ are the regularization hyperparameters chosen for each cluster p. Solving the cluster-specific regression problem, (13), is not only important for keeping the reduced-state dimension low (leading to faster online solves), it also guards against ill-conditioning of the operator inference regression problem with a growing number of POD modes. Algorithm 1 summarizes the proposed localized operator inference learning approach.

The ideal set of regularization parameters balance the minimization between the data fit and the regularisation to produce a ROM that minimizes some error metric over the full time domain. Various strategies can be devised for choosing these scalar hyperparameters. In the context of localized operator inference, we employ the following strategy: for every trajectory in the training set, an optimal set of hyperparameters is computed through a grid-search technique. These hyperparameters are chosen to be constant across the learning problems for inferring the local reduced state operators, i.e. $\lambda_{1,p} = \lambda_1, \lambda_{2,p} = \lambda_2$. For parameter values not included in the training set, the regularization parameter for learning problem (13) is chosen corresponding to the closest neighboring samples in the training set. This approach was found to lead to stable ROMs.

3.3 Localized operator inference: the online phase

3.3.1 Selection of ROM

In the online phase of the localized operator inference method we evaluate the classifier to decide which local ROM should be used given an indicator $\mathbf{z}$. Given a local reduced state $\widehat{\mathbf{s}}_p$, associated with local ROM p, the indicator is computed as

$$\mathbf{z} := \widetilde{\mathbf{s}} = \overline{\mathbf{V}}^\top \mathbf{V}_p \widehat{\mathbf{s}}_p \in \mathbb{R}^{r_G} \tag{14}$$

during the evaluation of the ROM. Cluster-related reduced-order quantities $\{\overline{\mathbf{V}}^\top \mathbf{V}_p\}_{p=1}^{n_p} \in \mathbb{R}^{r_G \times r_p}$ may be pre-computed since low computational cost is crucial during the online phase. We then evaluate the nearest neighbor classifier for the ROM selection.

Algorithm 1 Non-intrusive localized operator inference using Tikhonov regularization.

Input: partitions of the snapshot data $\{\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_{n_p}\}$, time derivative data $\{\dot{\mathbf{S}}_1, \dot{\mathbf{S}}_2, \ldots, \dot{\mathbf{S}}_{n_p}\}$, input data $\{\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_{n_p}\}$, and a user-specified energy threshold κ

Output: n_p sets of reduced model operators $\{\widehat{\mathbf{C}}_p, \widehat{\mathbf{A}}_p, \widehat{\mathbf{H}}_p, \widehat{\mathbf{B}}_p\}_{p=1}^{n_p}$

for $p \in \{1, 2, \ldots, n_p\}$, one for each data cluster **do**

  Compute the SVD of $\mathbf{S}_p$

  $r_p \leftarrow$ choose r_p such that $\dfrac{\sum_{i=1}^{r_p} \sigma_{p,i}^2}{\sum_{i=1}^{k_p} \sigma_{p,i}^2} > \kappa$ ▷ Dimension of local POD basis

  $\mathbf{V}_p \leftarrow$ the r_p leading left singular vectors of $\mathbf{S}_p$ ▷ Compute the local POD basis from the snapshots

  $\widehat{\mathbf{S}}_p \leftarrow \mathbf{V}_p^\top \mathbf{S}_p$ ▷ Project snapshots onto the r_p-dimensional POD subspace

  $\dot{\widehat{\mathbf{S}}}_p \leftarrow \mathbf{V}_p^\top \dot{\mathbf{S}}_p$ ▷ Projected time derivatives of the snapshots

  $\{\lambda_{1,p}, \lambda_{2,p}\} \leftarrow$ determine regularization parameters

  Solve cluster-specific learning problem (13) via r_p independent least-squares problems

end for

return $\{\widehat{\mathbf{C}}_p, \widehat{\mathbf{A}}_p, \widehat{\mathbf{H}}_p, \widehat{\mathbf{B}}_p\}_{p=1}^{n_p}$

3.3.2 ROM evaluation

With the localized operator inference approach, the ROM (3) takes the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\widehat{\mathbf{s}}_p(t) = \widehat{\mathbf{C}}_p + \widehat{\mathbf{A}}_p\widehat{\mathbf{s}}_p(t) + \widehat{\mathbf{H}}_p\left(\widehat{\mathbf{s}}_p(t) \otimes \widehat{\mathbf{s}}_p(t)\right) + \widehat{\mathbf{B}}_p\mathbf{u}(t); \quad \widehat{\mathbf{s}}_p(0) = \mathbf{V}_p^\top \mathbf{s}_0, \tag{15}$$

where the subscript p denotes the pth local ROM and $\widehat{\mathbf{C}}_p \in \mathbb{R}^{r_p}$, $\widehat{\mathbf{A}}_p \in \mathbb{R}^{r_p \times r_p}$, $\widehat{\mathbf{H}}_p \in \mathbb{R}^{r_p \times r_p^2}$, and $\widehat{\mathbf{B}}_p \in \mathbb{R}^{r_p \times m}$ are the corresponding local reduced model operators. The high-dimensional state approximation based on the pth POD basis is written as $\mathbf{s}(t) \approx \mathbf{V}_p\widehat{\mathbf{s}}_p(t)$. To switch from a local approximation $\widehat{\mathbf{s}}_b$ to a different approximation $\widehat{\mathbf{s}}_a$, the solution $\mathbf{V}_b\widehat{\mathbf{s}}_b(t)$ is projected onto the ath local subspace to account for the change in the POD coordinate systems:

$$\widehat{\mathbf{s}}_a = \mathbf{V}_a^\top \mathbf{V}_b \widehat{\mathbf{s}}_b \in \mathbb{R}^{r_a}. \tag{16}$$

The reduced-order matrices $\mathbf{V}_a^\top \mathbf{V}_b \in \mathbb{R}^{r_a \times r_b}$ may be pre-computed for any cluster-combination $\{a, b\}$ such that the projection (16) can be performed efficiently during the online solution of the reduced model.

3.4 Computational costs

We now present an offline-online breakdown of the computational cost of the proposed approach. During the the offline stage of the localized operator inference method, we incur additional computational and storage costs by performing a clustering analysis (step i). Generally speaking, soft clustering approaches (such as FCM) are more computationally demanding compared to hard clustering algorithms (e.g. k-means) because each data point has a weighting associated with every cluster. However, the low-dimensional representation of the snapshot data leads to a lower-dimensional clustering task and thus a more efficient clustering procedure. For the model selection, we first train a classifier (step ii). In this work we employ nearest neighbor classifiers, which are cheap to evaluate if the number of neighbors is kept low. Efficiency of the classifier is crucial since we evaluate it during the online phase of the approach. Additionally, we infer n_p local ROMs using the operator inference approach, instead of just one (step iii). However, learning a set of reduced model operators, (13), is an embarrassingly parallel task that can fully benefit from parallel architectures. An analysis of the computational cost associated with constructing a reduced model using operator inference was presented in [2]. While the proposed approach leads to an overall increase of the offline costs, the extra steps are justified by the expected online computational cost reduction.

The cost of the online phase of the localized operator inference approach is very similar to that of the original (global) operator inference method from [2]. The only additional cost in the online stage is related to evaluating a classifier (step iv). However, this cost does not depend on the number of data points used to

train the classifier nor on the number of clusters, n_p. The evaluation of the ROMs (step v) does not require any extra steps other than projections of the solution upon switching between ROMs when necessary. To summarize, the localized operator inference approach is well-suited for applications in which the online cost benefits due to the reduction of the basis dimension outweighs the increase of the extra offline costs.

4 Numerical experiments

In this section we demonstrate localized operator inference on two examples. We consider the Burgers' problem for shock propagation and the Cahn-Hilliard equation for modeling the process of phase separation. The FCM clustering algorithm and k-nearest neighbor search engine available in MATLAB are employed for clustering and classification, respectively. The minimization of objective function (8) in the clustering approach terminates when the objective function improves by less than a specified minimum threshold of 10^{-14}. Since the quality of the clustering procedure hinges on a randomized initial guess, we perform several replicates and select the clustering assignment that comes with the smallest value for the objective function (8). The fuzzy partition matrix exponent is set to $q = 1.5$. The time derivatives are approximated with a five-point stencil, which has fourth-order accuracy. The first two and last two time derivatives are computed using one-sided finite-difference approximations with the same order of accuracy. To remove redundant terms in the original Kronecker product, in practice we employ a compact Kronecker product in which the duplicate terms are removed. For instance, the standard Kronecker product of a vector $\mathbf{s} = [s_1, s_2]^\top$ with itself yields $[s_1^2, s_1 s_2, s_2 s_1, s_2^2]^\top$, whereas for the compact Kronecker product we have $[s_1^2, s_1 s_2, s_2^2]^\top$. The Kronecker product of two column vectors of length r contains only $r(r+1)/2$ unique quadratic terms.

4.1 Burgers' problem for shock propagation

We assess the method's performance on the inviscid Burgers' equation for shock propagation in a one-dimensional domain [46, 17, 18, 47, 21]. This problem is particularly challenging for projection-based ROMs. The difficulty arises from the fact that ROMs approximate the solution as a linear combination of spatially fixed reduced basis functions. However, when the underlying system dynamics exhibit motion with respect to a fixed grid, ROMs that use a global POD basis generally fail to capture the space-time evolution of the local phenomena accurately. Consider the following one-dimensional initial-boundary-value problem problem:

$$\frac{\partial}{\partial t} s(x,t) + s(x,t) \frac{\partial}{\partial x} s(x,t) = g(x); \quad x \in [0, 100], t \in [0, 50], \tag{17}$$

with $s(x,t)$ the unknown conserved quantity, and the forcing term $g(x) = 0.02e^{0.02x}$. The Dirichlet boundary condition used for the above PDE is

$$s(0,t) = s_{\mathrm{BC}}, \tag{18}$$

which is constant with time. The initial condition is given by

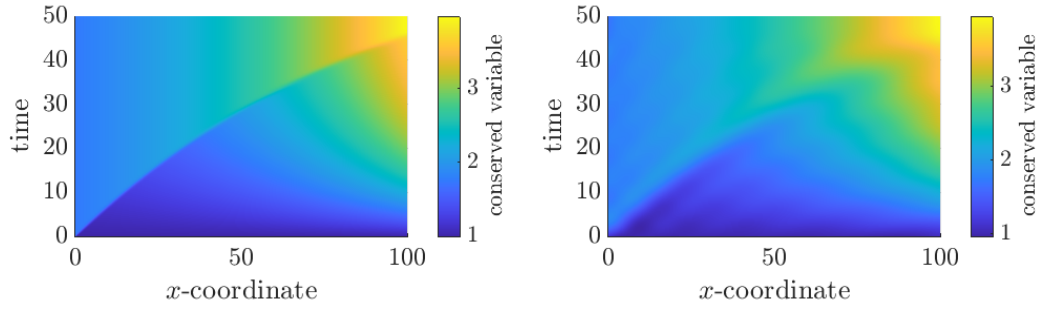
$$s(x,0) = \begin{cases} s_{\mathrm{BC}} & \text{for } x = 0 \\ 1 & \text{for } x > 0 \end{cases} \tag{19}$$

and is therefore parametrized by a single degree of freedom, s_{BC}. Its discontinuous nature in space is what gives rise to shock formation.

The problem is semi-discretized by a Godunov type finite volume method with 2^9 control volumes leading a system of nonlinear ODEs. The corresponding inferred reduced model takes the polynomial form

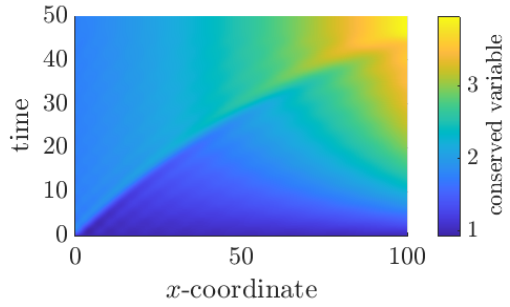
$$\frac{\mathrm{d}}{\mathrm{d}t} \widehat{\mathbf{s}}_p(t) = \widehat{\mathbf{C}}_p + \widehat{\mathbf{H}}_p \left(\widehat{\mathbf{s}}_p(t) \otimes \widehat{\mathbf{s}}_p(t) \right) + \widehat{\mathbf{B}}_p \mathbf{u}(t), \tag{20}$$

where $\widehat{\mathbf{s}}_p(t) \in \mathbb{R}^{r_p}$ is the reduced local state vector corresponding to the pth local ROM, $\widehat{\mathbf{H}}_p \in \mathbb{R}^{r_p \times r_p(r_p+1)/2}$ is the local operator corresponding to the nonlinear (quadratic) term, and $\widehat{\mathbf{C}}_p \in \mathbb{R}^r_p$ is the local vector corresponding to the constant terms. The local input matrix $\widehat{\mathbf{B}}_p$ imposes the Dirichlet boundary condition with the constant input $u(t) = s_{\mathrm{BC}}$. Note that the inviscid Burgers' equation (17) contains no linear

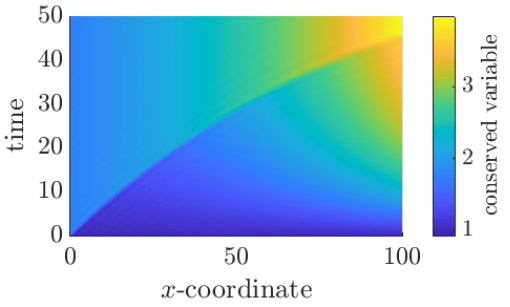


(a) FOM (reference)

(b) OpInf ROM, $r = 10$



(c) OpInf ROM, $r = 20$



(d) OpInf ROM, $r = 40$

Figure 2: Numerical simulation of a Burgers' problem at a testing point $s_{\text{BC}} = 1.75$. We compare the FOM with global operator inference ROMs constructed using different POD basis sizes.

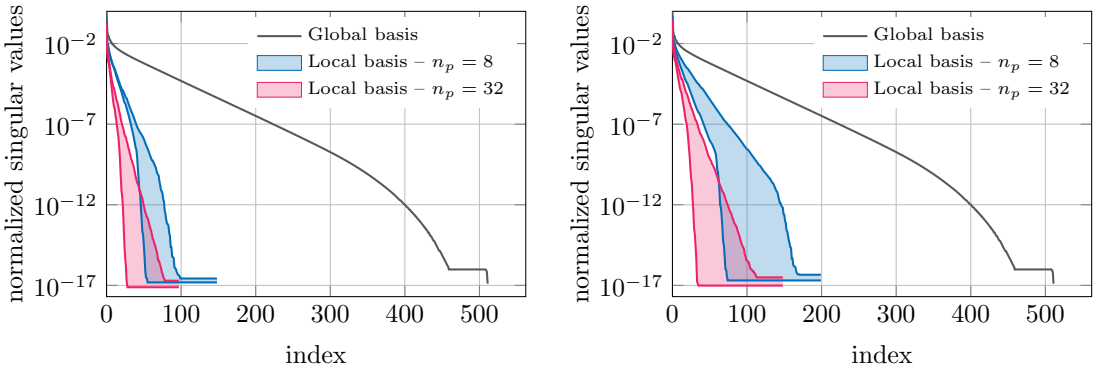
dependencies on the state, so in our postulated polynomial ROM (15), we have $\widehat{\mathbf{A}}_p = 0$. The local learning (13) problem thus takes the following form in each snapshot cluster:

$$\begin{aligned}\left(\widehat{\mathbf{C}}_p, \widehat{\mathbf{H}}_p, \widehat{\mathbf{B}}_p\right) = \underset{\widehat{\mathbf{C}}_p, \widehat{\mathbf{H}}_p, \widehat{\mathbf{B}}_p}{\arg\min} \Bigg(&\left\| \widehat{\mathbf{C}}_p \mathbf{1}_{k_p} + \widehat{\mathbf{H}}_p \left(\widehat{\mathbf{S}}_p \odot \widehat{\mathbf{S}}_p\right) + \widehat{\mathbf{B}}_p \mathbf{U}_p - \dot{\widehat{\mathbf{S}}}_p \right\|_F^2 \\ &+ \lambda_{1,p} \left(\|\widehat{\mathbf{C}}_p\|_F^2 + \|\widehat{\mathbf{B}}_p\|_F^2 \right) + \lambda_{2,p} \|\widehat{\mathbf{H}}_p\|_F^2 \Bigg); \quad p = 1, \ldots, n_p,\end{aligned} \tag{21}$$

where the reduced model operators are inferred with Algorithm 1.

To train the reduced models of the form (20) a training data set is constructed containing twenty equidistant points in the parameter domain $s_{\text{BC}} \in [1.5; 2.0]$. We consider a test set in which an additional 100 trajectories are generated with s_{BC} randomly drawn from a uniform distribution $\mathcal{U}([1.5; 2.0])$. The training trajectories are used to learn the reduced models. $k = 5,000$ time steps were taken to march forward in time. The simulation outputs the system state and time derivative every ten time steps, yielding a total of 10,000 snapshots for the learning problem. Both the FOM and the ROMs are solved with a semi-explicit Euler time-stepping scheme.

The operator inference approach from [2], described in Section 2, is applied to the solution of the aforementioned problem for reference. The method constructs a ROM in which a single POD basis represents the entire state. Figure 2 shows an example result of the shock propagation phenomenon. It depicts the evolution of the conserved quantity $s(x, t)$ in time over the one-dimensional domain at an arbitrary test point. The translating nature of the shock front proves to be a challenging problem for model reduction: when the number of POD basis functions is relatively low, the moving front is smeared out in space. It is also observed that the conserved variable displays oscillatory behavior upstream and downstream from the shock front. The small over- and undershoots in the solution of the learned model are a consequence of Gibbs phenomena arising from using a limited number of POD modes for ROM construction. As more POD modes are added, oscillations in the ROM solutions decrease consistently and the shock movement is reproduced with increasing levels of accuracy.



(a) Decay of singular values with non-overlapping snapshot clusters

(b) Decay of singular values with overlapping snapshot clusters

Figure 3: Comparison of the singular value decay of the collected snapshots for global and a set of local POD basis functions for Burgers' problem.

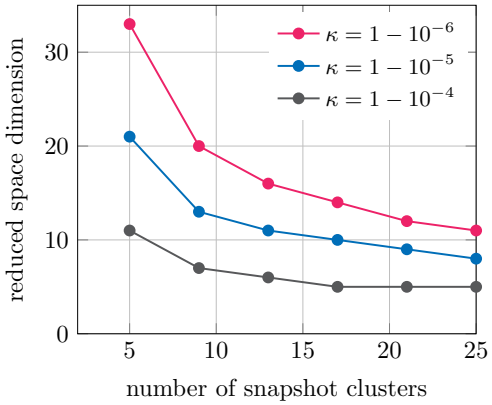


Figure 4: Analysis of the reduced space dimension as a function of the number of snapshot clusters n_p at different retained POD energy levels κ for Burgers' problem.

The proposed localized operator inference method is applied to the Burgers' problem in which cluster-specific ROMs are constructed. A dimensionality reduction is applied to the matrix of the concatenated initial conditions and trajectories $\mathbf{S}$, after which it is partitioned into n_p snapshot clusters. To determine the number of basis functions to retain in the low-dimensional representation of the snapshot matrix, $\widetilde{\mathbf{S}}$, we examine the quantity

$$\sum_{i=1}^{r_G} \sigma_i^2 \Big/ \sum_{i=1}^{n} \sigma_i^2 \,, \tag{22}$$

where σ_i are the singular values of the snapshot matrix $\mathbf{S}$. The quantity in Eq. (22) is the relative energy of the r_G retained modes of the full-state training data. Recall that this dimension reduction based on a global POD basis is used only for the purposes of snapshot clustering and training the classifier, and thus should not be viewed as a representative metric for ROM accuracy. The required numbers of modes, r_G, needed to retain 99.9%, 99.99% and 99.999% of the energy in the training data equals 6, 20 and 41, respectively. Our experiments found $r_G = 6$ to be sufficiently accurate for clustering and classification tasks in the Burgers' problem for shock propagation.

The singular values of the training data set are plotted in Figure 3. The partitioning of the snapshot data into clusters was found to significantly increase the rate at which the singular values decay. When the snapshot clusters are made to overlap with one another by sharing a number of states between neighboring clusters, the decay of the singular values is slightly slower compared to the non-overlapping cluster (Fig. 3b).

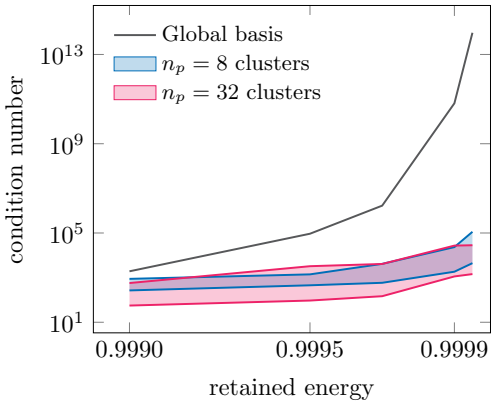


Figure 5: Condition number of the operator inference data matrix $[\mathbf{1}_{k_p}, \widehat{\mathbf{S}}_p \odot \widehat{\mathbf{S}}_p, \mathbf{U}_p]$ versus retained energy threshold, κ, for $n_p = 8$ and $n_p = 32$ snapshot clusters. The equivalent condition number for the baseline global operator inference approach [2] is included for reference. Shaded areas indicate the condition number range over the set of n_p localized ROMs.

While in principle snapshot clusters do not need overlap, this procedure avoids the formation of gaps between neighboring clusters and is reported to improve the quality of the sought-after reduced state solution upon transitioning between neighboring clusters [17].

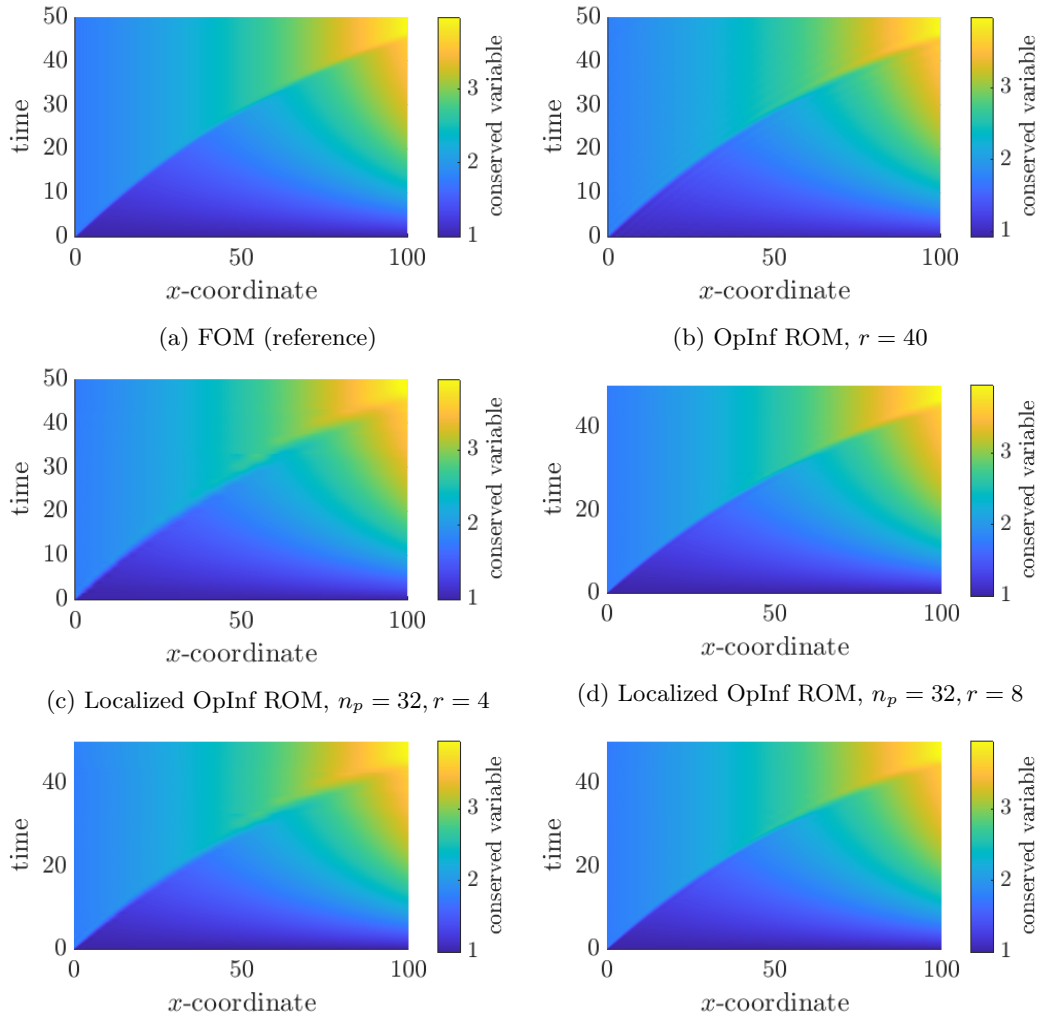
We now take a closer look at the impact of the number of adopted overlapping clusters, n_p. The number of modes to retain in each localized ROM is determined by evaluating the retained energy in each individual cluster, p, as a function of the reduced dimension in that cluster, r_p:

$$\kappa_{r_p} = \sum_{i=1}^{r_p} \sigma_{p,i}^2 \Big/ \sum_{i=1}^{k_p} \sigma_{p,i}^2 \; ; \quad p = 1, \ldots, n_p. \tag{23}$$

Each local ROM dimension, r_p, is chosen independently to satisfy $\kappa_{r_p} > \kappa$, where κ is a user-specified cumulative energy threshold. Figure 4 shows the median of the local reduced space dimension, r_p, as a function of the number of clusters at different values of κ. It is clear that, generally speaking, if κ is set to be constant across all clusters, the local bases do not contain the same number of POD basis vectors. It is also clear that even with a relatively low number of snapshot clusters, the local bases are reduced considerably in dimension compared to the global basis. Growing the number of clusters follows the law of diminishing returns. We note that the values of κ in Figure 4 are relatively high; this is required so that the ROMs sufficiently resolve the moving interface with time and is an indication of the challenge of approximating Burgers' equation (17).

The conditioning of operator inference regression problem poses numerical challenges. As the reduced dimension grows, the conditioning can deteriorate rapidly. In some cases the conditioning can be improved by adding more training data, although if the required size of the basis increases correspondingly then this can lead to diminishing returns [2, 34]. Regularization is the primary numerical strategy to address this issue, although the resulting ROM performance can be sensitive to the regularization hyperparameters. Our localized operator inference approach addresses this challenge because each localized operator inference regression problem has a smaller number of degrees of freedom than if we were to infer global ROM operators. Figure 5 compares the condition number of the localized operator inference data matrix $[\mathbf{1}_{k_p}, \widehat{\mathbf{S}}_p \odot \widehat{\mathbf{S}}_p, \mathbf{U}_p]$ (which is formed to solve (21)) at different retained POD energy thresholds. For the baseline operator inference approach (with a global basis) the condition number can get quite large ($> 10^{10}$) for this application as we add more POD modes. In contrast, for the same retained energy threshold in the local training data, the proposed approach reduces the condition number up to several orders of magnitude.

Figure 6 compares solutions computed with the baseline operator inference approach (using a global POD basis) and the localized operator inference method. In the latter ROM simulations a number of $n_p = 32$ snapshot clusters are employed. Examining the results in Figure 6, it can be seen that localized operator inference predictions capture the space-time evolution of the moving of the shock front well. Figures 6c–6d



(a) FOM (reference)

(b) OpInf ROM, $r = 40$

(c) Localized OpInf ROM, $n_p = 32, r = 4$

(d) Localized OpInf ROM, $n_p = 32, r = 8$

(e) Localized OpInf ROM, $n_p = 32$, $\kappa = 99.99\%$

(f) Localized OpInf ROM, $n_p = 32$, $\kappa = 99.999\%$

Figure 6: The numerical simulation of Burgers' problem at a testing point $s_{\text{BC}} = 1.75$. We compare the FOM with the proposed localized operator inference ROMs constructed using $n_p = 32$ snapshot clusters.

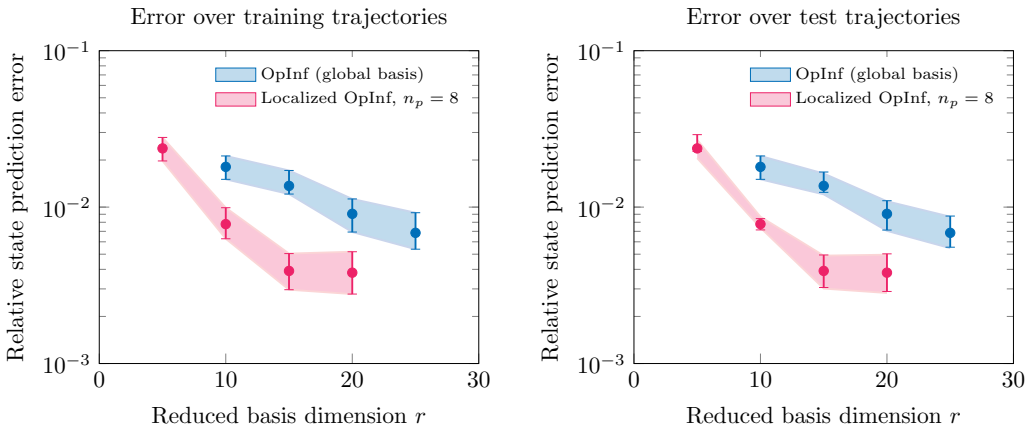


Figure 7: Operator inference model prediction error for Burgers' problem for shock propagation. Comparison of the baseline operator inference and localized operator inference methods for (left) the training trajectories and (right) a set of testing trajectories. Median and first/third quartile errors are shown.

set the dimension of every local basis at the levels of $r = 4$ and $r = 8$. It can be seen show that upon increasing the local basis size to $r = 8$, the ROM predictions for the conserved variable become virtually indistinguishable from that of the FOM, whereas upward of $r = 40$ POD modes are needed to produce the same level of accuracy with a global basis. Instead of employing a predefined number of localized basis functions, we now choose each local ROM's dimension based on the metric (23) (see also Algorithm 1). Figures 6e–6f plot the ROM solutions at $\kappa = 99.99\%$ and $\kappa = 99.999\%$. The corresponding set of ROMs has a reduced state dimension with ranges $r_p = 2$ to 6 and $r_p = 4$ to 9, respectively, for the two different levels of energy. Due to its local nature, the proposed approach also suppresses the usual undershoot-overshoot of the solution exhibited around the interface. The test and training errors are given in Figure 7. For each training and testing input, the error relative to the original high-dimensional solution, $\mathbf{S}_{\text{orig}} \in \mathbb{R}^{n\times k}$, is computed by solving the reduced model to generate the predictions in the reduced state, then reconstructing the full states of the trajectory in the original full state space to obtain $\mathbf{S}_{\text{rec}} \in \mathbb{R}^{n\times k}$. The relative error is given by the norm of the difference between the FOM and ROM solutions

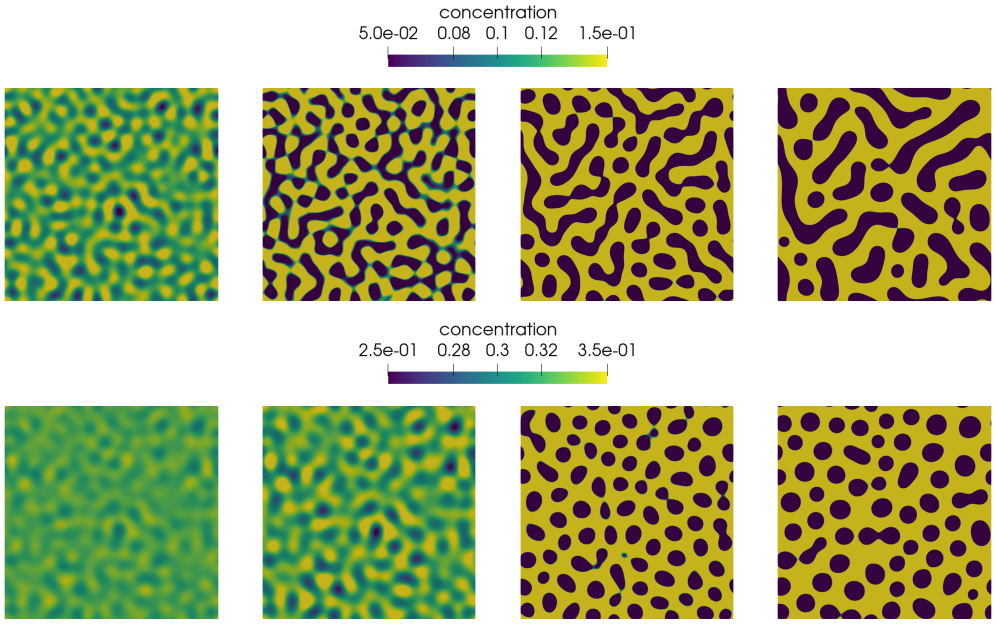
$$\frac{\|\mathbf{S}_{\text{rec}} - \mathbf{S}_{\text{orig}}\|_F}{\|\mathbf{S}_{\text{orig}}\|_F}. \tag{24}$$

For a limited computational budget, expressed through the number of POD basis functions in the reduced models, our testing indicates that the proposed method produces more accurate results compared to a baseline operator inference approach that relies on a single global POD basis, which is a key contribution of this work.

4.2 The Cahn-Hilliard phase-field model

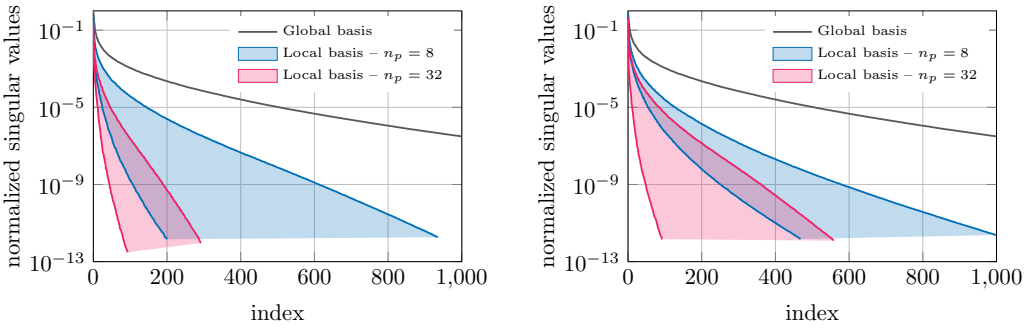
We consider the Cahn-Hilliard equation for modeling the dynamics of spinodal decomposition. For this system, localized reduced-order modeling was found to be imperative due to the prohibitive computational cost associated with traditional model order reduction techniques in capturing its rich dynamics across multiple time scales. We briefly introduce the problem and its governing equation, after which we apply the proposed localized non-intrusive reduced-order modeling framework. While the literature on reduced models for Cahn-Hilliard is sparse, some recent efforts have targeted this problem [48, 49, 50].

The Cahn-Hilliard phase-field model has been widely recognized as the generic equation modeling the leading-order dynamics of phase separation driven by chemical potential gradients. It is of vital importance in, for instance, understanding microstructure evolution during the additive manufacturing of metallic alloys, which is an important motivation for this work. The Cahn-Hilliard equation is a parabolic equation and involves first-order time derivatives, and second- and fourth-order spatial derivatives. The interested reader is referred to [51] for a recent discussion on its basic principles and practical applications.



(a) $t/T = 0.1$ (b) $t/T = 0.2$ (c) $t/T = 0.5$ (d) $t/T = 1$

Figure 8: Full-state time snapshots of the concentration field $s(\mathbf{x}, t)$ from different initial conditions: (top) $\bar{s} = 0.1$, (bottom) $\bar{s} = 0.3$.



(a) Decay of singular values with non-overlapping clusters

(b) Decay of singular values with overlapping clusters

Figure 9: Comparison of the singular value decay of the collected snapshots for global and a set of local POD basis functions for the Cahn-Hilliard system.

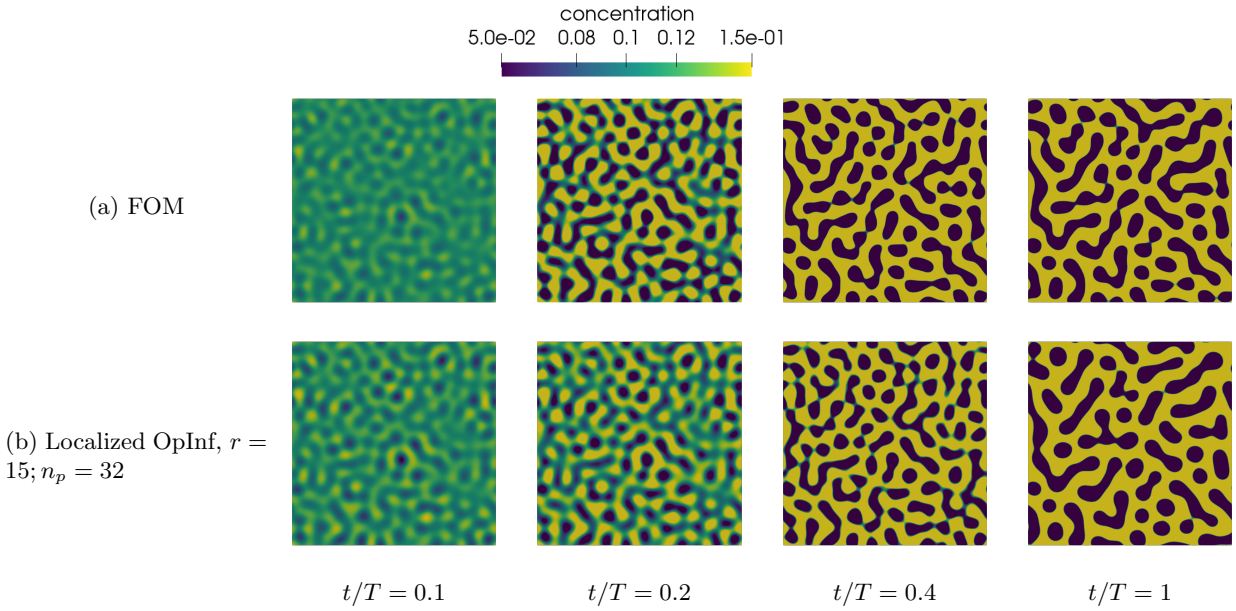


Figure 10: Time snapshots of the concentration field at a test point $\bar{s} = 0.15273$ for the FOM and a ROM constructed with $n_p = 32$ snapshot clusters.

Let $\Omega = [0, 1]^2 \subset \mathbb{R}^2$ be the spatial domain and $[0, T] \subset \mathbb{R}$ the time domain. Consider the Cahn-Hilliard equation of the form

$$\frac{\partial}{\partial t} s(\mathbf{x}, t) = M\nabla^2 \left(s^3(\mathbf{x}, t) - s(\mathbf{x}, t) - \ell\nabla^2 s(\mathbf{x}, t) \right), \quad \mathbf{x} \in \Omega; t \in [0, T) \tag{25}$$

where $s(\mathbf{x}, t)$ is the concentration (also 'phase-field' or 'order parameter') taking the values in $[-1, 1]$. It is related to the concentration of a substance concentration at the spatial coordinate $\mathbf{x} \in \Omega$ at time $t \in [0, T]$. The concentration is normalized such that the regions $s = \pm 1$ represent domains that are pure in each of the phases. The mobility M and the interface parameter ℓ, a measure proportional to the thickness of the regularized interface, are both assumed to be constant in space and time. We choose $M = 1$ and $\ell = 10^{-4}$ for the remainder of this section. Periodic boundary conditions are applied to all boundaries. The discretization is performed with the finite difference method on a uniform grid with a nodal spacing of 2^{-7} for a total of $n = 16,384$ degrees of freedom in the full model. The conducted numerical experiments focus on phase separation and coarsening in the time interval $t \in [0, 0.01]$.

The Cahn-Hilliard equation considered contains a cubic nonlinear term, i.e., $M\nabla^2 s^3$ on the righthand side of (25). Using operator inference, we thus seek to infer the local reduced model operators of the following dynamical system:

$$\frac{\mathrm{d}}{\mathrm{dt}} \widehat{\mathbf{s}}_p(t) = \widehat{\mathbf{A}}_p \widehat{\mathbf{s}}_p(t) + \widehat{\mathbf{G}}_p \left(\widehat{\mathbf{s}}_p(t) \otimes \widehat{\mathbf{s}}_p(t) \otimes \widehat{\mathbf{s}}_p(t) \right), \tag{26}$$

where $\widehat{\mathbf{s}}_p(t) \in \mathbb{R}^{r_p}$ is the local reduced state vector, $\widehat{\mathbf{A}}_p \in \mathbb{R}^{r_p \times r_p}$ the local operator corresponding to the reduced model representation of the linear terms of the Cahn-Hilliard equation, and $\widehat{\mathbf{G}}_p \in \mathbb{R}^{r_p \times r_p(r_p+1)(r_p+2)/6}$ the local operator corresponding to the reduced model representation of the cubic nonlinear term. Note that the Cahn-Hilliard equation (25) contains no quadratic dependencies on the state, nor are there constant terms. In our postulated polynomial ROM (3) this means that $\widehat{\mathbf{C}}_p = 0$ and $\widehat{\mathbf{H}}_p = 0$. There is also no external forcing term, that is $\widehat{\mathbf{B}}_p = 0$, because of periodicity in the state. Note that the presence of the cubic term leads to a modified operator inference learning problem (13). In each snapshot cluster p we solve the

following linear least-squares problem:

$$\left(\widehat{\mathbf{A}}_p, \widehat{\mathbf{G}}_p\right) = \underset{\widehat{\mathbf{A}}_p, \widehat{\mathbf{G}}_p}{\arg\min} \left(\left\| \widehat{\mathbf{A}}_p \widehat{\mathbf{S}}_p + \widehat{\mathbf{G}}_p \left(\widehat{\mathbf{S}}_p \odot \widehat{\mathbf{S}}_p \odot \widehat{\mathbf{S}}_p \right) - \dot{\widehat{\mathbf{S}}}_p \right\|_F^2 + \lambda_{1,p} \|\widehat{\mathbf{A}}_p\|_F^2 + \lambda_{2,p} \|\widehat{\mathbf{G}}_p\|_F^2 \right); \quad p = 1, \ldots, n_p. \tag{27}$$

The initial condition for (26) is given by a random perturbation about the nominal concentration, $\bar{s}$. At every grid point of the underlying discretization a noise value is added independent of one another. The initial condition is then formulated as $\mathbf{s}_0 = \bar{s} + 0.05\texttt{rand}(x_i, y_i)$, where the nominal concentration, $\bar{s}$, is drawn uniformly from the interval $[0.1, 0.4]$ and $\texttt{rand}(x_i, y_i)$ is the noise term whose values at grid point (x_i, y_i) are distributed uniformly in $[-1, 1]$. As evidenced by Figure 8, variations in the nominal concentration can give rise to a broad range of phase separation responses and morphologies in the Cahn-Hilliard system.

In what follows we adopt the strategy introduced in [49] by constructing pseudo-random initial states. By fixing the random perturbation across our training samples we generate a set of training data which is reproducible by our ROMs. A training set of 100 trajectories is constructed corresponding to a set of equidistant points in the parameter domain. By setting the time step size to $\Delta t = 10^{-6}$ and recording the data every 50 time steps, we obtain 200 state snapshots per high-fidelity simulation. This amounts to 20,000 snapshots used in the learning problem. The learned ROMs are then integrated with a fourth-order Runge-Kutta scheme (RK45). The test set for this problem consists of an additional 20 trajectories based on initial conditions drawn randomly from their distribution.

The singular values of the training data set are plotted in Figure 9. The proposed approach leverages the low-rank structure of the system: the singular values of the local POD bases decay faster than the singular values of the global basis. However, in comparison to the numerical experiments from the previous subsection, the Cahn-Hilliard system needs a greater number of snapshot clusters to fully exploit the capabilities of localized ROMs.

Table 1 tabulates the online runtime of the localized ROM for different retained energy thresholds. The measurements are wall-clock time obtained from a MATLAB R2020a implementation on a local machine with a 1.60 GHz Intel Core i5 CPU and 16GB of RAM. The runtimes are averaged over a total of ten runs. The time marching scheme and configuration are identical for the FOM and the ROM simulations. The speedup achieved by the proposed approach is about two orders of magnitude when the local reduced basis dimension is relatively small. As expected, the speedups drop significantly as the ROMs become richer with an increasing number of basis functions.

Table 1: Online runtimes for the ROMs constructed with the proposed localized operator inference framework at different values for the locally retained energy. The runtime for the FOM is 61.65 sec.

local basis size r_p	10	20	30	40
runtime (sec)	0.35	0.79	2.92	10.81
speedup ($\times$)	175	77.53	20.98	5.67

Figure 10 gives an overview of the localized operator inference ROM prediction performance for the primary quantity of interest, which is the concentration field $s(\mathbf{x}, t)$. The ROM predictions at selected times agree well with the solutions to the original high-dimensional model at an arbitrary test point. To further assess the performance of our reduced models, we calculate the autocorrelation of the spatially dependent concentration field to characterize the evolving microstructure in a statistical sense [50]. Autocorrelation functions represent a widely used mathematical framework for microstructural characterization and may be interpreted as the conditional probability that two different points within the microstructure are found to be in the same phase [52, 53]. Importantly, the invariance of autocorrelations with respect to translations and rotations of the periodic microstructure make it a good candidate to assess the accuracy of our ROMs. Figure 11 plots the differences in the autocorrelations of the concentration field at the final simulation time $t = 0.01$. As the number of modes in the local training data increases from $r_p = 6$ to $r_p = 15$, the point-wise error of the autocorrelation of the concentration function drops below 1% throughout the computational domain. Minimum and maximum of the mean spatial error in the autocorrelations over the training and

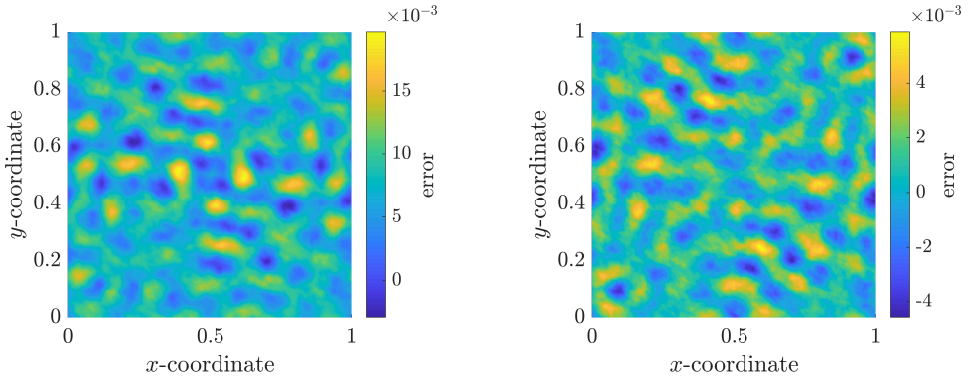


Figure 11: Point-wise error between the autocorrelations of the true (FOM) and predicted (localized ROM) concentration field at an arbitrary testing point $s = 0.15273$ (at final time $t = 0.01$) using (a) 6 local POD modes and (b) 15 local POD modes.

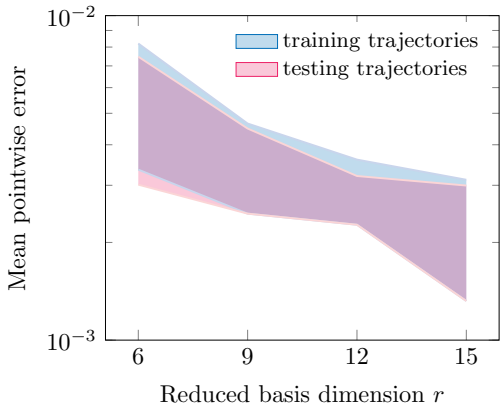


Figure 12: Localized operator inference model prediction error for the Cahn-Hilliard equation for the training and testing trajectories. The mean spatial error in the autocorrelations is shown over the trajectories.

test are shown in Figure 12. The error in the concentration decreases with the increase in the size of the POD basis, as expected. This shows that a localized ROM constructed using the proposed localized operator inference approach is applicable for approximating, in a statistical sense, the leading-order dynamics of phase separation as governed by the Cahn-Hilliard equation.

5 Conclusions and future directions

Many applications in engineering and science can benefit from the concept of localized reduced-order models. We have presented the localized operator inference approach, a new learning method for constructing local reduced-order models from simulation data. Our method differs from existing works in the literature in that the reduced models are learned in a non-intrusive fashion. This is an important consideration for enabling model-based reduction of complex dynamical systems. The proposed framework is applicable to nonlinear static, dynamic, and parametric problems. Like most reduced-order modeling methods, it consists of two stages. In the offline stage several reduced models are constructed by first clustering a set of training data into a number of snapshot clusters. The operator inference framework then finds the local reduced model operators that yield the reduced model that best matches the projected snapshot data in a minimum-residual sense. In the online phase of the approach, model recommendation is performed via a nearest neighbor classifier that was trained offline with simulation data. Numerical experiments on two different

test problems demonstrate the ability of localized operator inference to reduce the dimensions of the learned models without compromising the accuracy of the method. It is also shown that the proposed approach can decrease the condition number of the associated least-squares learning problems by several orders of magnitude.

Future efforts will focus on devising more refined regularization strategies that fully exploit the locality of the available data. Another interesting avenue is the use of higher-order approximations of the solution manifold, rather than the usual piecewise linear approximations. Representations beyond linear subspaces have recently been proposed and, in principle, are applicable to any dynamical system and fully compatible with the proposed localized non-intrusive reduced-order modeling framework.

References

[1] O. Ghattas and K. Willcox, "Learning physics-based models from data: perspectives from inverse problems and model reduction," *Acta Numerica*, vol. 30, p. 445–554, 2021.

[2] B. Peherstorfer and K. Willcox, "Data-driven operator inference for nonintrusive projection-based model reduction," *Computer Methods in Applied Mechanics and Engineering*, vol. 306, pp. 196 – 215, 2016.

[3] L. Sirovich, "Turbulence and the dynamics of coherent structures part I: Coherent structures," *Quarterly of applied mathematics*, vol. 45, no. 3, pp. 561–571, 1987.

[4] G. Berkooz, P. Holmes, and J. L. Lumley, "The proper orthogonal decomposition in the analysis of turbulent flows," *Annual review of fluid mechanics*, vol. 25, no. 1, pp. 539–575, 1993.

[5] J. L. Lumley, "The structure of inhomogeneous turbulent flows," *Atmospheric turbulence and radio wave propagation*, 1967.

[6] C. Greif and K. Urban, "Decay of the Kolmogorov N-width for wave problems," *Applied Mathematics Letters*, vol. 96, pp. 216–222, 2019.

[7] N. Cagniart, Y. Maday, and B. Stamm, *Model Order Reduction for Problems with Large Convection Effects*, pp. 131–150. Cham: Springer International Publishing, 2019.

[8] M. Nonino, F. Ballarin, G. Rozza, and Y. Maday, "Overcoming slowly decaying Kolmogorov n-width by transport maps: application to model order reduction of fluid dynamics and fluid–structure interaction problems," *arXiv preprint arXiv:1911.06598*, 2019.

[9] M. Redeker and B. Haasdonk, "A POD-EIM reduced two-scale model for crystal growth," *Advances in Computational Mathematics*, vol. 41, no. 5, pp. 987–1013, 2015.

[10] J. L. Eftang, A. T. Patera, and E. M. Rønquist, "An "*hp*" Certified Reduced Basis Method for Parametrized Elliptic Partial Differential Equations," *SIAM Journal on Scientific Computing*, vol. 32, no. 6, pp. 3170–3200, 2010.

[11] J. L. Eftang, D. J. Knezevic, and A. T. Patera, "An hp certified reduced basis method for parametrized parabolic partial differential equations," *Mathematical and Computer Modelling of Dynamical Systems*, vol. 17, no. 4, pp. 395–422, 2011.

[12] J. L. Eftang and B. Stamm, "Parameter multi-domain 'hp' empirical interpolation," *International Journal for Numerical Methods in Engineering*, vol. 90, no. 4, pp. 412–428, 2012.

[13] B. Haasdonk, M. Dihlmann, and M. Ohlberger, "A training set and multiple bases generation approach for parameterized model reduction based on adaptive grids in parameter space," *Mathematical and Computer Modelling of Dynamical Systems*, vol. 17, no. 4, pp. 423–442, 2011.

[14] B. Peherstorfer, D. Butnaru, K. Willcox, and H.-J. Bungartz, "Localized Discrete Empirical Interpolation Method," *SIAM Journal on Scientific Computing*, vol. 36, no. 1, pp. A168–A192, 2014.

[15] M. Dihlmann, M. Drohmann, and B. Haasdonk, "Model reduction of parametrized evolution problems using the reduced basis method with adaptive time-partitioning," *Proc. of ADMOS*, vol. 2011, p. 64, 2011.

[16] M.-L. Rapún and J. M. Vega, "Reduced order models based on local POD plus galerkin projection," *Journal of Computational Physics*, vol. 229, no. 8, pp. 3046–3063, 2010.

[17] D. Amsallem, M. J. Zahr, and C. Farhat, "Nonlinear model order reduction based on local reduced-order bases," *International Journal for Numerical Methods in Engineering*, vol. 92, no. 10, pp. 891–916, 2012.

[18] K. Washabaugh, D. Amsallem, M. Zahr, and C. Farhat, "Nonlinear model reduction for CFD problems using local reduced-order bases," in *42nd AIAA Fluid Dynamics Conference and Exhibit*, p. 2686, 2012.

[19] D. Amsallem, M. J. Zahr, and K. Washabaugh, "Fast local reduced basis updates for the efficient reduction of nonlinear systems with hyper-reduction," *Advances in Computational Mathematics*, vol. 41, no. 5, pp. 1187–1230, 2015.

[20] S. Grimberg, C. Farhat, R. Tezaur, and C. Bou-Mosleh, "Mesh sampling and weighting for the hyper-reduction of nonlinear Petrov–Galerkin reduced-order models with local reduced-order bases," *International Journal for Numerical Methods in Engineering*, vol. 122, no. 7, pp. 1846–1874, 2021.

[21] D. Amsallem and B. Haasdonk, "PEBL-ROM: Projection-error based local reduced-order models," *Advanced Modeling and Simulation in Engineering Sciences*, vol. 3, no. 1, pp. 1–25, 2016.

[22] E. Kaiser, B. R. Noack, L. Cordier, A. Spohn, M. Segond, M. Abel, G. Daviller, J. Östh, S. Krajnović, R. K. Niven, and et al., "Cluster-based reduced-order modelling of a mixing layer," *Journal of Fluid Mechanics*, vol. 754, p. 365–414, 2014.

[23] T. Daniel, F. Casenave, N. Akkari, and D. Ryckelynck, "Model order reduction assisted by deep neural networks (ROM-net)," *Advanced Modeling and Simulation in Engineering Sciences*, vol. 7, no. 1, pp. 1–27, 2020.

[24] T. Daniel, A. Ketata, F. Casenave, and D. Ryckelynck, "Physics-informed cluster analysis and a priori efficiency criterion for the construction of local reduced-order bases," *arXiv preprint arXiv:2103.13683*, 2021.

[25] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 3932–3937, 2016.

[26] H. Schaeffer, "Learning partial differential equations via data discovery and sparse optimization," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 473, no. 2197, p. 20160446, 2017.

[27] B. Galletti, C. H. Bruneau, L. Zannetti, and A. Iollo, "Low-order modelling of laminar flow regimes past a confined square cylinder," *Journal of Fluid Mechanics*, vol. 503, p. 161–170, 2004.

[28] B. Galletti, A. Bottaro, C.-H. Bruneau, and A. Iollo, "Accurate model reduction of transient and forced wakes," *European Journal of Mechanics-B/Fluids*, vol. 26, no. 3, pp. 354–366, 2007.

[29] L. Cordier, B. A. El Majd, and J. Favier, "Calibration of POD reduced-order models using Tikhonov regularization," *International Journal for Numerical Methods in Fluids*, vol. 63, no. 2, pp. 269–296, 2010.

[30] T. G. Kolda and B. W. Bader, "Tensor Decompositions and Applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.

[31] C. Gu, "QLMOR: A Projection-Based Nonlinear Model Order Reduction Approach Using Quadratic-Linear Representation of Nonlinear Systems," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 30, no. 9, pp. 1307–1320, 2011.

[32] B. Kramer and K. E. Willcox, "Nonlinear Model Order Reduction via Lifting Transformations and Proper Orthogonal Decomposition," *AIAA Journal*, vol. 57, no. 6, pp. 2297–2307, 2019.

[33] E. Qian, B. Kramer, B. Peherstorfer, and K. Willcox, "Lift & Learn: Physics-informed machine learning for large-scale nonlinear dynamical systems," *Physica D: Nonlinear Phenomena*, vol. 406, p. 132401, 2020.

[34] R. Swischuk, B. Kramer, C. Huang, and K. Willcox, "Learning physics-based reduced-order models for a single-injector combustion process," *AIAA Journal*, vol. 58, no. 6, pp. 2658–2672, 2020.

[35] P. Khodabakhshi and K. Willcox, "Non-intrusive data-driven model reduction for differential algebraic equations derived from lifting transformations," *Oden Institute Report*, pp. 21–08, 2021.

[36] G. Golub and C. Van Loan, *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, 2013.

[37] S. A. McQuarrie, C. Huang, and K. E. Willcox, "Data-driven reduced-order models via regularised Operator Inference for a single-injector combustion process," *Journal of the Royal Society of New Zealand*, vol. 51, no. 2, pp. 194–211, 2021.

[38] N. Sawant, B. Kramer, and B. Peherstorfer, "Physics-informed regularization and structure preservation for learning stable reduced models from data with operator inference," *arXiv preprint arXiv:2107.02597*, 2021.

[39] I. Assent, "Clustering high dimensional data," *WIREs Data Mining and Knowledge Discovery*, vol. 2, no. 4, pp. 340–350, 2012.

[40] J. C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters," *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973.

[41] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.

[42] S. Ghosh and S. K. Dubey, "Comparative analysis of k-means and fuzzy c-means algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 4, 2013.

[43] M. G. Kapteyn and K. E. Willcox, "From physics-based models to predictive digital twins via interpretable machine learning," *arXiv preprint arXiv:2004.11356*, 2020.

[44] A. Gordon, *Classification, 2nd Edition*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, 1999.

[45] C. M. Bishop, *Pattern recognition and machine learning*. Information science and statistics, New York, NY: Springer, 2006. Softcover published in 2016.

[46] M. Rewieński and J. White, "Model order reduction for nonlinear dynamical systems based on trajectory piecewise-linear approximations," *Linear Algebra and its Applications*, vol. 415, no. 2, pp. 426 – 454, 2006. Special Issue on Order Reduction of Large-Scale Systems.

[47] K. Carlberg, "Adaptive h-refinement for reduced-order models," *International Journal for Numerical Methods in Engineering*, vol. 102, no. 5, pp. 1192–1210, 2015.

[48] C. Gräßle and M. Hinze, "POD reduced-order modeling for evolution equations utilizing arbitrary finite element discretizations," *Advances in Computational Mathematics*, vol. 44, no. 6, pp. 1941–1978, 2018.

[49] E. N. Karatzas and G. Rozza, "A reduced order model for a stable embedded boundary parametrized Cahn-Hilliard phase-field system based on cut finite elements," *arXiv preprint arXiv:2009.01596*, 2021.

[50] D. M. de Oca Zapiain, J. A. Stewart, and R. Dingreville, "Accelerating phase-field-based microstructure evolution predictions via surrogate models trained by machine learning methods," *npj Computational Materials*, vol. 7, no. 1, pp. 1–11, 2021.

[51] J. Kim, S. Lee, Y. Choi, S.-M. Lee, and D. Jeong, "Basic Principles and Practical Applications of the Cahn-Hilliard Equation," *Mathematical Problems in Engineering*, vol. 2016, 2016.

[52] S. Torquato and G. Stell, "Microstructure of two-phase random media. I. the n-point probability functions," *The Journal of Chemical Physics*, vol. 77, no. 4, pp. 2071–2077, 1982.

[53] J. G. Berryman, "Measurement of spatial correlation functions using image processing techniques," *Journal of Applied Physics*, vol. 57, no. 7, pp. 2374–2384, 1985.