Learning latent representations in high-dimensional state spaces using polynomial manifold constructions

Rudy Geelen, Laura Balzano and Karen Willcox

Abstract— We present a novel framework for learning cost-efficient latent representations in problems with high-dimensional state spaces through nonlinear dimension reduction. By enriching linear state approximations with low-order polynomial terms we account for key nonlinear interactions existing in the data thereby reducing the problem's intrinsic dimensionality. Two methods are introduced for learning the representation of such low-dimensional, polynomial manifolds for embedding the data. The manifold parametrization coefficients can be obtained by regression via either a proper orthogonal decomposition or an alternating minimization based approach. Our numerical results focus on the one-dimensional Korteweg-de Vries equation where accounting for nonlinear correlations in the data was found to lower the representation error by up to two orders of magnitude compared to linear dimension reduction techniques.

I. INTRODUCTION

Dimension reduction rests on the fundamental assumption that samples from high-dimensional state spaces can be represented with a smaller number of variables (often called "latent" variables) without a significant loss of information. An arsenal of mathematical procedures has been proposed in the literature to find a compact representation of the data. The identification of such intrinsic, low-dimensional structure in problems with high-dimensional state spaces sits at the heart of this paper. In particular, we explore the use of state approximations that embed low-order polynomial terms. These terms are introduced to account for nonlinear correlations in the data thereby enabling a greater reduction in the dimensionality for the problem at hand. Our proposed approach remains interpretable through its formulation in terms of a modal basis expansion. As the least-squares method underlies the construction of these approximations, the framework can also be deployed effectively for large-scale, real-world datasets.

We start by summarizing the main linear technique for dimensionality reduction, namely, the principal component analysis (PCA) [1]–[3]. The PCA performs a linear mapping of the data to a lower-dimensional space by maximizing its variance in the low-dimensional representation [3]. The

breadth of applications for PCA, and its numerous adaptations, is largely due to its simplicity and ease of use. It should be noted, however, that PCA is sometimes also referred to as proper orthogonal decomposition (POD) [4]–[6]. PCA and POD are closely related and in some cases algorithmically equivalent, but have arisen from different communities. When it comes to dimension reduction in problems with high-dimensional state spaces associated with simulation data, POD is the method of choice. Both techniques have at their core the singular value decomposition (SVD) to identify low-rank structure. The methodology in this paper is presented from the perspective of POD, but everything we propose is equally applicable to PCA.

POD analysis dictates that the coefficients of the basis functions are linearly uncorrelated. Consequently, POD provides an optimal embedding of the original, high-dimensional data only into a linear subspace. It is unable to deal with nonlinear correlations existing in the data. The ability to define an adequate nonlinear embedding in high-dimensional state spaces is thus of great importance in dimension reduction problems. Many alternatives to POD have been proposed for identifying and unraveling such nonlinear correlations. These methods include, for instance, kernel PCA [7], Isomap [8], locally linear embedding [9], multidimensional scaling (MDS) [10], different variants of autoencoders, and other architectures that rely on neural networks as recently reviewed in, e.g., [11]. Despite their successful deployment in many applications where linear dimensionality reduction techniques were found to fall short, machine learning methods are often criticized for their limited interpretability and susceptibility to overfitting. This can hinder, for instance, the embedding of physical constraints, such as boundary conditions, and the preservation of other data features and properties [12].

In this paper we introduce a new procedure for capturing key nonlinear interactions of the data by taking into account the existence of a low-dimensional, nonlinear manifold. The proposed strategy remains interpretable in that the nonlinear state approximations are given in terms of a modal basis expansion, as is the case with POD. The presented approach is also readily applicable in the dimension reduction of large-scale problems as its underlying optimization process is fully data-driven and can be carried out using standard least-squares solvers. The paper is organized as follows. First, we briefly review some known results for the POD as it applies to data compression problems. This is followed by a description of two instances of a nonlinear dimension reduction technique based on polynomial manifold construc-

Rudy Geelen is with the Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX 78712 USA `rudy.geelen@austin.utexas.edu`

Laura Balzano is with the Department of Electrical Engineering and Computer Science, University of Michigan Ann Arbor, Ann Arbor, MI 48109 USA `girasole@umich.edu`

Karen Willcox is with the Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX 78712 USA `kwillcox@oden.utexas.edu`

tions. The effectiveness of the proposed framework is then demonstrated in a nonlinear problem setting.

II. PROBLEM FORMULATION

The POD provides a systematic manner for the identification of a set orthonormal basis functions that can be used for approximating solutions in problems with high-dimensional state spaces. The POD basis is computed as the set of left singular vectors of the data matrix $\mathbf{S} \in \mathbb{R}^{n \times k}$ given by

$$\mathbf{S} = \begin{pmatrix} | & | & & | \\ \mathbf{s}(t_1) & \mathbf{s}(t_2) & \dots & \mathbf{s}(t_k) \\ | & | & & | \end{pmatrix} \quad (1)$$

whose columns consist of a collection of high-dimensional system states $\mathbf{s}(t) \in \mathbb{R}^n$, where $t \in \mathcal{T}$ is some parameter on which the state depends. The POD then seeks to produce rank-r approximations of the state variable expressed through a linear combination of POD modes as follows:

$$\mathbf{s}(t) \approx \mathbf{\Phi}\widehat{\mathbf{s}}(t), \quad (2)$$

where $\mathbf{\Phi} \in \mathbb{R}^{n \times r}$ is a basis matrix (the POD basis) containing as columns the left singular vectors of $\mathbf{S}$ corresponding to the r largest singular values. The column space of $\mathbf{\Phi}$ then defines an r-dimensional subspace of the full state space $\mathbb{R}^n$ with, generally speaking, $r \ll n$. The vector $\widehat{\mathbf{s}}(t) \in \mathbb{R}^r$ contains the reduced state (also POD) coordinates.

Provided that the total number of training samples is smaller than the dimension of the high-dimensional system states, that is $k < n$, the singular values of $\mathbf{S}$ are denoted $\sigma_1 \geq \dots \sigma_k \geq 0$. The POD basis minimizes the least-squares error between the original data and their representation in the reduced space (for a fixed basis size). This error is equal to the sum of the squares of the singular values corresponding to those left singular vectors not included in the basis

$$\sum_{i=r+1}^{k} \sigma_i^2, \quad (3)$$

where σ_i is the ith singular value of $\mathbf{S}$.

From a statistics viewpoint, the orthogonal projection underlying the POD provides linearly uncorrelated features. Despite its optimality properties, it cannot unravel nonlinear correlations in the data. Accounting for such nonlinear correlations is a crucial factor in further reducing the dimensionality of the problem. In what follows, we seek to represent the high-dimensional data on a low-dimensional submanifold $\mathcal{M} \subset \mathbb{R}^n$ of the original state space defined as $\mathcal{M} = \{\mathbf{s}(t) \,|\, t \in \mathcal{T}\}$, which is constructed to approximate well the elements of the solution manifold for the problem of interest.

In the field of machine learning, the numerous tools aimed at identifying dominant nonlinear correlations are often referred to collectively as manifold learning or representation learning [11]. We here focus on nonlinear transformations that can be expressed through a basis expansion. Of particular interest are nonlinear state approximations of polynomial form:

$$\mathbf{s}(t) \approx \underbrace{\mathbf{V}\widehat{\mathbf{s}}(t)}_{\text{linear}} + \underbrace{\mathbf{Z}\mathbf{g}(\widehat{\mathbf{s}}(t))}_{\text{nonlinear}}, \quad (4)$$

where

$$\mathbf{g}(\widehat{\mathbf{s}}(t)) = \begin{pmatrix} \widehat{\mathbf{s}}^2(t) \\ \widehat{\mathbf{s}}^3(t) \\ \vdots \\ \widehat{\mathbf{s}}^p(t) \end{pmatrix}, \quad (5)$$

in which $\mathbf{V} = [\mathbf{v}_1 \,|\, \dots \,|\, \mathbf{v}_r] \in \mathbb{R}^{n \times r}$ is the linear basis matrix that has $\mathbf{v}_j \in \mathbb{R}^n$, $j = 1, \dots, r$, as its jth column. The column space of $\mathbf{V}$ spans an r-dimensional linear subspace of the full state space $\mathbb{R}^n$. The matrix $\mathbf{Z} \in \mathbb{R}^{n \times (p-1)r}$ is a matrix whose columns span a different subspace from the one spanned by the columns of $\mathbf{V}$. The vector $\widehat{\mathbf{s}} \in \mathbb{R}^r$ is the reduced state vector, representing the coefficients of the basis expansion. The vector $\mathbf{g}(\widehat{\mathbf{s}}(t)) \in \mathbb{R}^{(p-1)r}$ denotes a set of polynomial terms up to degree $p \geq 2$ associated with the nonlinear aspect of solution manifold $\mathcal{M}$. Throughout this paper, the exponent used in (5) raises each element of a given vector to the corresponding power. For example, if $\widehat{\mathbf{s}} = [\widehat{s}_1, \widehat{s}_2]^\top$, then $\widehat{\mathbf{s}}^p = [\widehat{s}_1^p, \widehat{s}_2^p]^\top$. The use of polynomial constructions is not new in representation learning problems. However, the precise manner in which the polynomial approximations are built and deployed leads to vastly different approaches [13]–[17].

It is noted that one is permitted to omit the linear contribution $\mathbf{V}\widehat{\mathbf{s}}(t)$ from approximation (4). The assumption underlying this choice is that the correlations between the latent space coefficient are expected to be exclusively nonlinear. The philosophy of our approach is that the structure of the low-dimensional submanifold of the state space is unknown and must thus be approximated using regression techniques.

If matrix $\mathbf{V}$ in (4) is chosen to be the POD basis matrix, $\mathbf{Z}$ can be inferred from the high-dimensional dataset by means of linear least-squares regression [15]. A different approach is to consider a parametrization of the form

$$\mathbf{Z} = \overline{\mathbf{V}}\mathbf{\Xi} \in \mathbb{R}^{n \times (p-1)r}, \quad (6)$$

where the matrix $\overline{\mathbf{V}} = [\overline{\mathbf{v}}_1 \,|\, \dots \,|\, \overline{\mathbf{v}}_q] \in \mathbb{R}^{n \times q}$ represents a basis matrix, and $\mathbf{\Xi} \in \mathbb{R}^{q \times (p-1)r}$ is the corresponding coefficient matrix that weighs each of the polynomial terms contained in $\mathbf{g}(\widehat{\mathbf{s}}(t))$.

For optimization purposes, see Section III, we choose the columns of $\mathbf{V}$ and $\overline{\mathbf{V}}$ to form orthonormal sets of dimension r and q, respectively. Additionally, the columns of $\overline{\mathbf{V}}$ are chosen to be orthogonal to those of $\mathbf{V}$. Equivalently we can write

$$[\mathbf{V}, \overline{\mathbf{V}}]^\top [\mathbf{V}, \overline{\mathbf{V}}] = \mathbf{I}_{r+q}, \quad (7)$$

where $\mathbf{I}_a$ denotes an identity matrix of dimension $a \times a$. It is noted that constraint set (7) forms a smooth submanifold of $\mathbb{R}^{n \times (r+q)}$ known as the Stiefel manifold.

To construct nonlinear approximations of the form (4) in an optimal fashion, the representation learning problem is

posed in terms of the constrained optimization problem

$$\min_{\mathbf{V},\overline{\mathbf{V}},\boldsymbol{\Xi},\widehat{\mathbf{S}}} J(\mathbf{V},\overline{\mathbf{V}},\boldsymbol{\Xi},\widehat{\mathbf{S}}) \tag{8}$$
$$\text{such that } [\mathbf{V},\overline{\mathbf{V}}]^{\top}[\mathbf{V},\overline{\mathbf{V}}] = \mathbf{I}_{r+q},$$

with objective function

$$J = \frac{1}{2}\sum_{j=1}^{k} \left\| \mathbf{s}_j - [\mathbf{V},\overline{\mathbf{V}}] \begin{bmatrix} \widehat{\mathbf{s}}_j \\ \boldsymbol{\Xi}\mathbf{g}(\widehat{\mathbf{s}}_j) \end{bmatrix} \right\|_2^2, \tag{9}$$

where $\widehat{\mathbf{S}} := (\widehat{\mathbf{s}}_1, \widehat{\mathbf{s}}_2, \ldots, \widehat{\mathbf{s}}_k) \in \mathbb{R}^{r \times k}$ is the reduced-state representation of the given system states $\mathbf{s}_j$ for $j = 1,\ldots,k$. Solving (8) means that (4) holds, in the least-squares sense, at the optimum of the objective of (8).

The use of Frobenius norm regularization was advocated in [15] to avoid the overfitting of the nonlinear approximations to the training data. Here we choose to penalise only the entries of the coefficient matrix $\boldsymbol{\Xi}$, leading to the modified optimization problem

$$\min_{\mathbf{V},\overline{\mathbf{V}},\boldsymbol{\Xi},\widehat{\mathbf{S}}} \left(J(\mathbf{V},\overline{\mathbf{V}},\boldsymbol{\Xi},\widehat{\mathbf{S}}) + \frac{\gamma}{2}\|\boldsymbol{\Xi}\|_F^2 \right) \tag{10}$$
$$\text{such that } [\mathbf{V},\overline{\mathbf{V}}]^{\top}[\mathbf{V},\overline{\mathbf{V}}] = \mathbf{I}_{r+q},$$

which is the problem of interest in this paper. Particular choices for each of the terms $\{\mathbf{V},\overline{\mathbf{V}},\boldsymbol{\Xi}\}$ and the corresponding latent space representation $\widehat{\mathbf{s}}(t)$, and how these may be inferred from the data, will be presented next.

The proposed dimensionality reduction method is unique in its ability to map directly from the latent space to the original, high-dimensional state space—via (4)—without invoking additional approximations or the need to evaluate neural network architectures. Despite the wealth of literature on manifold learning methods, they do (generally speaking) not provide such a mapping in explicit form. This inhibits their application in large classes of problems such as various model reduction and classification tasks.

III. LEARNING NONLINEAR REPRESENTATIONS

A. Approach 1 – POD-based modeling

Our first option views the learning of nonlinear state approximations through the lens of the POD. In the following we demonstrate in step-by-step fashion how conventional POD approximations can be enriched to produce approximations of the form (4).

Step 1: Compute basis: In this proposed approach we start by populating the basis matrix $\mathbf{V}$ with the left singular vectors of $\mathbf{S}$ corresponding to its r largest singular values. We then choose the basis vectors corresponding to the next consecutive q largest singular values as the columns of $\overline{\mathbf{V}}$. A similar idea was also proposed recently in [18]. Figure 1 illustrates the left singular vectors used in conventional POD and the proposed POD-based polynomial manifold approach. Note that this choice for the pair of basis matrices $\{\mathbf{V},\overline{\mathbf{V}}\}$ automatically satisfies constraint (7) through the orthogonality properties of the SVD.
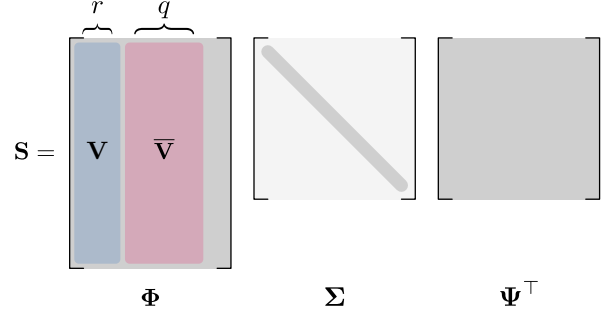

Fig. 1: The left singular vectors used in the state approximations for proposed POD-based polynomial manifold approach following a singular value decomposition of the data matrix $\mathbf{S} = \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Psi}^{\top}$.

Step 2: Low-dimensional representation: Finding the latent space representation for this approach simply amounts to computing the representation of each data sample j in the POD coordinates as $\widehat{\mathbf{s}}_j = \mathbf{V}^{\top}\mathbf{s}_j$ for $j = 1,\ldots,k$.

Step 3: Linear least-squares regression: Given the matrices $\{\mathbf{V},\overline{\mathbf{V}}\}$ and the latent space representation $\widehat{\mathbf{S}}$, this leaves the coefficient matrix $\boldsymbol{\Xi}$ to be inferred from the training data. It is then trivial to show that (10) reduces to an unconstrained linear optimization problem:

$$\min_{\boldsymbol{\Xi}} \left(J(\mathbf{V},\overline{\mathbf{V}},\boldsymbol{\Xi},\widehat{\mathbf{S}}) + \frac{\gamma}{2}\|\boldsymbol{\Xi}\|_F^2 \right), \tag{11}$$

whose exact solution can be computed using the normal equations as

$$\boldsymbol{\Xi} = \overline{\mathbf{V}}^{\top}(\mathbf{I}_n - \mathbf{V}\mathbf{V}^{\top})\mathbf{S}\mathbf{W}^{\top}(\mathbf{W}\mathbf{W}^{\top} + \gamma\mathbf{I}_{(p-1)r})^{-1}, \tag{12}$$

with $\boldsymbol{\Xi} \in \mathbb{R}^{q \times (p-1)r}$, where we introduced the data matrix $\mathbf{W}$ as

$$\mathbf{W} := \begin{bmatrix} | & | & & | \\ \mathbf{g}(\widehat{\mathbf{s}}_1) & \mathbf{g}(\widehat{\mathbf{s}}_2) & \ldots & \mathbf{g}(\widehat{\mathbf{s}}_k) \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{(p-1)r \times k}. \tag{13}$$

The vector accounting for the nonlinear interactions between the reduced-state coefficients of the jth data sample, $\mathbf{g}(\widehat{\mathbf{s}}_j)$, contains low-order polynomial nonlinear terms (up to degree p) and is given in (5).

Contributions & comparison with previous work: The proposed POD-based polynomial manifold formulation described above produces only an approximate minimizer to optimization problem (8), as explained in [15, Remark 1]. Nonetheless, the formulation is particularly appealing from a computational perspective: it permits us to embed readily available POD modes explicitly into state approximation (4) without modifying the dimensionality of the reduced problem, an important contribution of this work.

Despite also having its roots in polynomial regression, the POD-based manifold approach presented in [15] is fundamentally different from the one described here in a number of key aspects. A parametrization of the form (6) permits us to explicitly leverage POD modes when approximating system states, rather than needing to compute the columns of $\mathbf{Z}$ from the data. As a result, the number of coefficients to be

inferred in the proposed approach no longer scales with the full-state dimension, n, thereby reducing the computational and storage burden in constructing such approximations. Furthermore, the nonlinear part of the approximation is expressed through low-order polynomial terms instead of Kronecker quadratic product form. Accounting for higher-order nonlinear correlations existing between the various POD mode amplitudes will promote an improved compression of the data.

B. Approach 2 – The alternating minimization approach

In this section we take a different approach from above in that we propose an iterative algorithm for solving problems of the form (10) using alternating minimization. Classical alternating minimization (or projection) algorithms have been successfully applied in solving optimization problems over two or more variables or equivalently of finding a point in the intersection of two sets. Although one sacrifices a certain degree of interpretability by solving the representation learning problem through optimization procedures, each step of the proposed method is intuitive yet computationally tractable and thus scalable. Our approach is detailed in individual subsections below in step-by-step fashion. We also provide an adequate choice of the initialization as well as a relevant stopping criterion in the alternating minimization process.

Step 1: The Procrustes problem: We start with initial guesses for the representation of the training data in the low-dimensional coordinate system at each sample $\widehat{\mathbf{s}}_j$, $j = 1, \ldots, k$, and the coefficient matrix $\boldsymbol{\Xi}$. The matrix of the concatenated bases $\boldsymbol{\Omega} := [\mathbf{V}, \overline{\mathbf{V}}]$ can then be inferred from the available data. Because the regularizer in (10) depends neither on $\mathbf{V}$ nor on $\overline{\mathbf{V}}$, an equivalent optimization problem may be considered without regularization. In the Frobenius norm (10) then simplifies to

$$\boldsymbol{\Omega} = \underset{\boldsymbol{\Omega} \in \mathbb{R}^{n \times (r+q)}}{\arg \min} \frac{1}{2} \left\| \mathbf{S} - \boldsymbol{\Omega} \begin{bmatrix} \widehat{\mathbf{S}} \\ \boldsymbol{\Xi} \mathbf{W} \end{bmatrix} \right\|_F^2 \qquad (14)$$
$$\text{such that } \boldsymbol{\Omega}^\top \boldsymbol{\Omega} = \mathbf{I}_{r+q}.$$

Matrix approximation problem (14) is better known in the statistical literature as the orthogonal Procrustes problem [19]. It aims to learn the best possible unitary (orthogonal) transformation, that is a rotation or reflection, that relates two given matrices. The solution to (14) was derived by Schönemann [20] as

$$\boldsymbol{\Omega} = \mathbf{U}_P \mathbf{V}_P^\top, \qquad (15)$$

where $\mathbf{U}_P \boldsymbol{\Sigma}_P \mathbf{V}_P^\top = \mathbf{S}[\widehat{\mathbf{S}}^\top, (\boldsymbol{\Xi} \mathbf{W})^\top]$ is an SVD, and is unique if $\mathbf{S}[\widehat{\mathbf{S}}^\top, (\boldsymbol{\Xi} \mathbf{W})^\top]$ is full rank.

Step 2: Linear least-squares regression: Given an estimate for the basis matrices, $\{\mathbf{V}, \overline{\mathbf{V}}\}$, and the same low-dimensional coordinate system used in Step 1, we now solve for the coefficient matrix $\boldsymbol{\Xi}$. In practice this amounts to solving (11), with the main difference being one of interpretation: in Approach 1 the basis matrices $\{\mathbf{V}, \overline{\mathbf{V}}\}$ are the left singular vectors of the high-dimensional training data whereas in the alternating minimization approach this is no

longer the case as one has to solve an orthogonal Procrustes problem to find a set of orthonormal basis matrices.

Step 3: Unconstrained nonlinear least-squares: In the final step of the minimization process we compute the representation of the high-dimensional data on $\mathcal{M}$. For this we have new estimates for $\mathbf{V}, \overline{\mathbf{V}}$ and $\boldsymbol{\Xi}$ at hand, meaning that (10) simplifies to

$$\widehat{\mathbf{S}} = \underset{\widehat{\mathbf{S}} \in \mathbb{R}^{r \times k}}{\arg \min} \frac{1}{2} \sum_{j=1}^{k} \left\| \mathbf{s}_j - [\mathbf{V}, \overline{\mathbf{V}}] \begin{bmatrix} \widehat{\mathbf{s}}_j \\ \boldsymbol{\Xi} \mathbf{g}(\widehat{\mathbf{s}}_j) \end{bmatrix} \right\|_2^2, \qquad (16)$$

which is an unconstrained optimization problem that can thus be treated efficiently using Levenberg-Marquardt and trust-region-reflective methods available in many off-the-shelf solvers.

Initial guess and termination criterion: The iterative process must be initiated with guesses for the representation of the data in the low-dimensional coordinate system, $\widehat{\mathbf{s}}_j$ for $j = 1, \ldots, k$, as well as the coefficient matrix $\boldsymbol{\Xi}$. These are readily available in our POD-based polynomial manifold approach (Approach 1). In the proposed alternating minimization scheme we solve the problems in Steps 1 to 3 in a sequential fashion, constituting one iteration in the optimization procedure. Iterations are terminated when a global stopping criterion is satisfied. We choose to monitor the retained information content, see [15] for a motivation, accounted for by approximation (4) across the iterations, denoted by the ℓ superscript, as

$$e^\ell := \frac{\|\mathbf{V}^\ell \widehat{\mathbf{S}}^\ell + \overline{\mathbf{V}}^\ell \boldsymbol{\Xi}^\ell \mathbf{W}^\ell\|_F^2}{\|\mathbf{S}\|_F^2}. \qquad (17)$$

We then define convergence if $|e^{\ell+1} - e^\ell| \leq \texttt{TOL}$ with $\texttt{TOL}$ a predefined tolerance.

IV. NUMERICAL RESULTS

A. Error metric & good practices

In the numerical experiments conducted in this work we represent a given dataset $\mathbf{S}(\boldsymbol{\mu}_i)$ in the computed reduced-state coordinate system, after which we attempt to reconstruct that same dataset in the original, high-dimensional state space as $\widetilde{\mathbf{S}}(\boldsymbol{\mu}_i)$. The different parameter instances are given by $\boldsymbol{\mu}_i$ with $i = 1, \ldots, m$. Importantly, the data used for testing are unseen in learning process. This enables us to quantify the representation error of nonlinear state approximation (4) as follows:

$$\frac{1}{m} \sum_{i=1}^{m} \frac{\|\mathbf{S}(\boldsymbol{\mu}_i) - \widetilde{\mathbf{S}}(\boldsymbol{\mu}_i)\|_F^2}{\|\mathbf{S}(\boldsymbol{\mu}_i)\|_F^2} \qquad (18)$$

The data matrices under consideration are centered by their column-averaged mean value. The regularization parameter γ in (10) is found through a grid search and confirmed to provide an adequate amount of regularization irrespective of the polynomial degree of the state approximation or the number of basis vectors $r + q$. There usually is a trade-off between choosing r, the dimension of the linear subspace described by basis $\mathbf{V}$, and q, the number of basis vectors in $\overline{\mathbf{V}}$. In what follows, r is user-specified as it corresponds to the
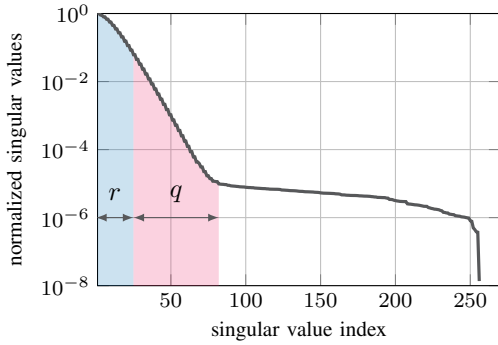
Fig. 2: Normalized singular values of the mean-subtracted data matrix for the Korteweg-de Vries problem. The blue and red areas denote the singular values whose corresponding left singular vectors are in $\mathbf{V}$ and $\overline{\mathbf{V}}$, respectively.

(a) Reference solution   (b) POD reconstruction with $r = 6$

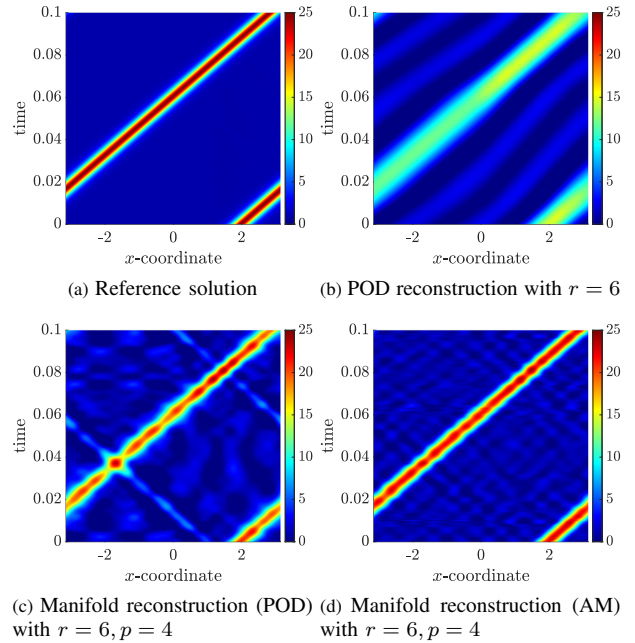(c) Manifold reconstruction (POD) with $r = 6, p = 4$ (d) Manifold reconstruction (AM) with $r = 6, p = 4$

Fig. 3: Reconstruction of the high-dimensional system-states using the different state approximation at $\mu_j = 1.9298$ with reduced basis dimension $r = 6$. The manifold models have polynomial embeddings of degree $p = 4$.

desired dimension of the latent space. Dimension q is then chosen in accord with the targeted level of representation accuracy which is informed by the singular value decay of the shifted data matrices. The tolerance for the alternating minimization algorithm, see Section III-B, is chosen to be `TOL` $= 10^{-3}$.

B. Korteweg-de Vries equation

Our numerical experiments are concerned with traveling wave physics. Consider a single soliton propagating at constant speed over a domain with periodic boundary conditions. The evolution of the wave field is given by the one-dimensional Korteweg-de Vries equation

$$\partial_t s = -\alpha s \partial_x s - \beta \partial_x^3 s \tag{19}$$

in the space-time domain $[-\pi, \pi] \times [0, 0.1]$ with initial condition $s_0(x) = 1 + 24 \operatorname{sech}^2\left(\sqrt{8}(x - \mu)\right)$ and with ∂_x and ∂_t denoting partial derivatives with respect to x and t, respectively [21]. An equidistant grid was used consisting of 256 grid points. The state data are saved every 0.0002 time units. We choose the model constants to be $\alpha = 8$ and $\beta = 1$ and generate five individual datasets at $\mu = [0, 0.5, 1, 1.5, 2]$, yielding a total of 2,500 samples to be used for learning the state representations. Each data sample appears in the form of an n-dimensional column vector and is stored column-wise in the data matrix $\mathbf{S}$, as given in (1).

As evidenced by the decay of the singular values of the dataset, shown in Fig. 2, 82 POD basis vectors are necessary for approximating the training data with a projection error below 10^{-5} in the Euclidean norm. The decay is also found to reach a plateau around this value, at which point increasing the reduced basis dimension no longer meaningfully improves the quality of the state approximation. In the numerical experiments that follow we therefore maintain a constant $r + q = 82$. The regularization parameter was chosen to be $\gamma = 500$.

An additional ten datasets are generated, based on the initial conditions, by uniformly drawing μ in the range $[0, 2]$. In this experiment we are looking to better understand the ability of the POD, Manifold-POD (Section III-A), and Manifold-AM (Section III-B) formulations in building an

r-dimensional reduced-space representation and, from that, reconstruct a set of high-dimensional system-states with a minimal loss of information. Fig. 3 gives an illustration of the performance of the different reduction methods at a reduced dimension of $r = 6$. The numerical experiment considers a polynomial embedding of degree $p = 4$ for the nonlinear state approximations. The POD clearly struggles with capturing the space-time evolution of the Korteweg-de Vries equation. As it represents the optimal solution that can be approximated using a linear subspace spanned by only r basis vectors, the reconstruction fails to capture the soliton's amplitude and shows Gibbs oscillations in its vicinity. The reconstructions from the polynomial manifold formulations on the other hand, shown in Figs. 3c and 3d, are clearly better suited for this by accounting for correlations of polynomial nature between the latent space coefficients. This ability is beneficial in reducing the problem's dimensionality.

The test errors in function of the reduced basis dimension are shown in Fig. 4. As expected, POD is the least able in accurately reconstructing the test data. When accounting for an additional q POD modes through the addition of low-order polynomial terms to the state approximation, the representation error decreases. However, the proposed alternating minimization approach, in which we rely on the solution to a Procrustes-type problem for inferring a set of orthonormal basis vectors, is considerably more accurate than both aforementioned alternatives. Improvements up to two orders of magnitude in accuracy were obtained compared to POD. Both nonlinear manifold approaches were found to benefit from increasing the polynomial degree in the nonlinear part of the approximation (4).

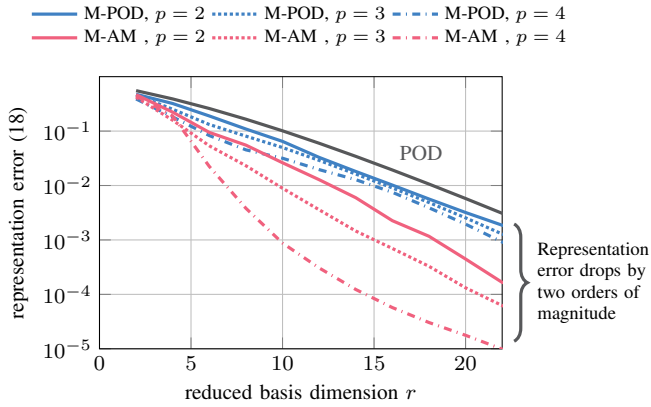The number of iterations at which (17) falls below the

Fig. 4: Plot of the representation error, (18), in function of the reduced basis dimension, r, for the standard POD approach (black curve), POD-based polynomial manifold approach (blue curves) and alternating minimization based polynomial manifold approach (red curves). For the manifold formulations we vary the degree p of the polynomial embedding.

TABLE I: The number of alternating minimization cycles at which (17) falls below the specified threshold at different reduced basis dimensions, r, and values for the polynomial embedding degree, p.

Reduced basis dim. r		2	4	6	8	10	12	14
	$p = 2$	4	8	11	20	9	10	5
# iterations	$p = 3$	6	13	18	15	6	5	4
	$p = 4$	19	29	9	6	4	3	3

specified threshold for the alternating minimization procedure is listed in Table I. While there is a quite a bit of variation at different values of the reduced basis dimension, r, and the degree of the polynomial embeddings, p, the number of iterations remains small for all numerical experiments reported here. This shows that learning nonlinear state approximations of the form (4) through minimization techniques is computationally tractable.

V. CONCLUDING REMARKS

We present a general framework for building nonlinear state approximations that is particularly applicable for dimension reduction in problems with high-dimensional state spaces. By embedding low-order polynomial terms in the modal basis expansion we account for nonlinear correlations in the data through interactions of the reduced space coordinates. The proposed framework prompts two different representation learning approaches. The POD-based polynomial manifold approach is intuitive due its connection to conventional POD, yet enables a reduction of the dimensionality for the problem at hand. If one is willing to depart from POD, and its interpretable nature, the proposed alternating minimization based variant offers an even better compression performance by finding basis matrices with orthonormal columns through a Procrustes-type problem. Importantly, both learning approaches can effectively be deployed at scale using only least-squares techniques.

VI. ACKNOWLEDGEMENTS

REFERENCES

[1] K. Pearson, "On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

[2] H. Hotelling, "Analysis of a complex of statistical variables with principal components," *Journal of Educational Psychology*, vol. 24, pp. 417–441, 498–520, 1933.

[3] I. Jolliffe, "Principal component analysis," in *Encyclopedia of Statistics in Behavioral Science*, John Wiley & Sons, Ltd., 2005.

[4] J. L. Lumley, "The structures of inhomogeneous turbulent flow," *Atmospheric Turbulence and Radio Wave Propagation*, pp. 166–178, 1967.

[5] L. Sirovich, "Turbulence and the dynamics of coherent structures. I. Coherent structures," *Quarterly of Applied Mathematics*, vol. 45, no. 3, pp. 561–571, 1987.

[6] P. Holmes, J. L. Lumley, G. Berkooz, and C. W. Rowley, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge, UK: Cambridge University Press, 1996.

[7] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *Artificial Neural NetworksICANN'97: 7th International Conference Lausanne, Switzerland, October 8–10, 1997 Proceeedings*, pp. 583–588, Springer, 2005.

[8] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[9] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[10] I. Borg and P. J. Groenen, *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media, 2005.

[11] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[12] R. Swischuk, L. Mainini, B. Peherstorfer, and K. Willcox, "Projection-based model reduction: Formulations for physics-based machine learning," *Computers & Fluids*, vol. 179, pp. 704–717, 2019.

[13] H. Qiao, P. Zhang, D. Wang, and B. Zhang, "An explicit nonlinear mapping for manifold learning," *IEEE Transactions on Cybernetics*, vol. 43, no. 1, pp. 51–63, 2013.

[14] J. Barnett and C. Farhat, "Quadratic approximation manifold for mitigating the kolmogorov barrier in nonlinear projection-based model order reduction," *Journal of Computational Physics*, vol. 464, p. 111348, 2022.

[15] R. Geelen, S. Wright, and K. Willcox, "Operator inference for non-intrusive model reduction with quadratic manifolds," *Computer Methods in Applied Mechanics and Engineering*, vol. 403, p. 115717, 2023.

[16] J. L. Barnett, C. Farhat, and Y. Maday, "Mitigating the kolmogorov barrier for the reduction of aerodynamic models using neural-network-augmented reduced-order models," in *AIAA SCITECH 2023 Forum*, p. 0535, 2023.

[17] B. Kramer, S. Gugercin, and J. Borggaard, "Nonlinear balanced truncation: Part 2 – model reduction on manifolds," 2023.

[18] A. Cohen, C. Farhat, A. Somacal, and Y. Maday, "Nonlinear compressive reduced basis approximation for PDE's." working paper or preprint, Mar. 2023.

[19] J. C. Gower and G. B. Dijksterhuis, *Procrustes problems*, vol. 30. OUP Oxford, 2004.

[20] P. H. Schönemann, "A generalized solution of the orthogonal procrustes problem," *Psychometrika*, vol. 31, no. 1, pp. 1–10, 1966.

[21] A. Mendible, S. L. Brunton, A. Y. Aravkin, W. Lowrie, and J. N. Kutz, "Dimensionality reduction and reduced-order modeling for traveling wave physics," *Theoretical and Computational Fluid Dynamics*, vol. 34, pp. 385–400, 2020.