2024 IEEE 63rd Conference on Decision and Control (CDC)
December 16-19, 2024. MiCo, Milan, Italy

Manifold Clustering Based Nonlinear Model Reduction with Application to Nonlinear Convection

T. Wu, D. Wilson, S.M. Djouadi

Abstract— **This paper proposes a new cluster method combined with Dynamic Mode Decomposition with Control (DMDc), and the Proper Orthogonal Decomposition (POD) to construct more accurate reduced order models. DMDc and POD are popular data-driven techniques that extract low-order models from high-dimensional complex dynamic systems. However, these methods are inherently linear, i.e., the data is assumed to belong to linear manifolds. However, this may lead to inaccuracies in the reduced models commensurate with the presence of nonlinearities. To capture the nonlinear behavior, manifold clustering is introduced to group the snapshots obtained by experiments or numerical simulation into several sub-regions based on the underlying non-linear structure. Manifold clustering is a powerful approach for exploratory data analysis, allowing the discovery of patterns and structures that are not apparent in raw high-dimensional data. It does not require knowing the number of clusters and the intrinsic manifold dimensions in advance. Manifold clustering is combined with DMDc and POD to construct the local reduced-order models. Time clustering is applied to the snapshots generated by a nonlinear convective flow governed by the 2D Burgers' equations with boundary actuation. The manifold cluster reduced order model outperforms standard and other cluster-based (K-means) reduced order models.**

I. INTRODUCTION

The Proper Orthogonal Decomposition (POD) [1] and Dynamic Mode Decomposition with control (DMDc) [2] are sophisticated analytical methods employed to extract dynamic features and construct reduced-order models of complex systems. These methodologies are particularly significant in fields where the systems are governed by high-dimensional, nonlinear dynamical equations, such as in fluid dynamics, atmospheric science, and more recently, in the burgeoning area of data-driven science and machine learning. The combined use of POD and DMDc [3] [4] [5] [6] in fluid dynamics serves a dual purpose. While POD offers a lens through which the most energetic and influential flow features are viewed, DMDc expands the view to encompass the dynamics of how the flow evolves under the influence of external control. Together, these methods empower researchers and engineers to not only observe and describe fluid behaviors but also to exert precise influence over these behaviors, paving the way for advanced flow control techniques that can be applied to modern aerodynamics, climate modeling, and energy systems.

However, popular reduced order models such as DMD and POD [7] are inherently linear, that is, the snapshots are assumed to belong to a linear vector space. In [8] [9] [10] [11] [12], the authors propose to quantify the global nonlinear manifold geodesic by partitioning the data manifold into several regions and then approximating using local Euclidean distances. Snapshot data were grouped into several sub-regions using the K-means algorithm, and reduced-order models based on POD and DMDc were constructed in each sub-region. The results show a significant improvement over conventional POD and DMDc.

The general notion of clustering is used in various domains such as data compressing, image processing, etc. In [13], [14], [7] [15], Centroid Voronoi Tessellations (CVT) is a clustering method based on k-means algorithm. CVT is used to construct a reduced order model by choosing the cluster centroids as the reduced basis. In [16], the authors group snapshots into several clusters in the state space, then cluster centroids partition the state space in complementary non-overlapping regions, based on these regions, the state transition matrix is constructed using a Markov process and finally applied to a mixing layer problem. In [17], differential geometry properties of Riemannian manifolds are used to produce suitable reduced-order bases for nonlinear dynamical systems.

Here, other than the K-means cluster method, we investigate manifold clustering that works by combining the concepts of manifold learning and clustering to group data points that lie on the same manifold in a high-dimensional space. The development of manifold clustering originated from the manifold hypothesis, which suggests that while datasets might appear high-dimensional, they often rest on underlying low-dimensional manifolds [18]. Early studies in the realm of manifold learning, particularly with techniques such as Isomap (Isometric Mapping) [19] and Locally Linear Embedding (LLE) [20], offered essential methods to approximate the intrinsic geometry of these manifolds. These initial efforts laid the crucial groundwork for the advancement of manifold clustering techniques. When the data is modeled as a union of several manifolds, manifold clustering is needed in addition to manifold learning [21]. In [22], [23], [24], different types of manifold clustering have been proposed to solve the problem when the manifolds intersect each other [18].

This paper is a continuation of the work undertaken in [8] [9] [11]. Herein, the manifold clustering method is used to group the system snapshots obtained through experimental data or numerical simulation into several sub-regions where the underlying manifold structure is similar in each region to construct a reduced-order model for each sub-

Department of Electrical Engineering and Computer Science twu19@vols.utk.edu, dwilso81@utk.edu, mdjouadi@utk.edu

region. Moreover, for comparison purposes, the K-means algorithm [25] is used to group the snapshots that have similar behavior into clusters. In particular, time clustering is applied to the snapshots generated of a nonlinear convective flow governed by the 2D Burgers' equation with boundary actuation resulting in a cluster-reduced order method that yields better results than the K-means cluster DMDc and POD-reduced order models. The main discovery is that the manifold clustered reduced order model outperforms the K-means clustered reduced order model. The 2D Burgers' equation is used as a surrogate to the Navier-Stokes equation since it has the same nonlinearity as the latter [26] [27]. It is a fundamental partial differential equation occurring in several fields including fluid mechanics, acoustics, gas dynamics, and traffic flow [28].

This paper is organized as follows. Section II provides a brief background about DMDc and POD, and two types of clustering methods, the K-means and Manifold clustering algorithms, followed by the proposed clustered reduced order modeling method. Section III presents a numerical study that illustrates the performance of the manifold cluster reduced order model with comparison to the K-means cluster reduced order models. Concluding remarks are provided in Section IV.

II. Background: DMDc, POD, and Clustering

A. DMDc

DMD aims to extract information about a dynamical system from an ensemble of experimental or numerical snapshot data. Dynamic Mode Decomposition with Control (DMDc) extends the capabilities of DMD by incorporating control inputs or external forcing into the analysis [29]. This enhancement is particularly valuable in systems where control strategies significantly influence the system's behavior. By integrating control inputs into the DMD framework, DMDc enables the identification of control-affine dynamic modes, prediction of system responses under different control scenarios, and design of optimal control strategies [29]. Letting $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^l$ be a measurement/snapshot made at time k, DMDc assumes that (x_k, x_{k+1}, u_k) can be approximated by a linear operator A and B as [29]:

$$x_{k+1} = Ax_k + Bu_k \tag{1}$$

where $B \in \mathbb{R}^{n\times l}$, Where $u_k \in \mathbb{R}^l$ and $B \in \mathbb{R}^{n\times l}$. To handle the incorporation of control input, consider a new matrix of control input snapshots defined as [29]:

$$\Gamma = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \dots & u_{m-1} \\ | & | & & | \end{bmatrix} \tag{2}$$

In conjunction with the previously defined matrices X and $X^{'}$, the dynamical system can be represented according to [29]

$$X' \approx AX + B\Gamma \tag{3}$$

represented in matrix form:

$$X = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \dots & x_{m-1} \\ | & | & & | \end{bmatrix} \tag{4}$$

$$X' = \begin{bmatrix} | & | & & | \\ x_2 & x_3 & \dots & x_m \\ | & | & & | \end{bmatrix} \tag{5}$$

Where m is the total number of the snapshots and X' is the time shift snapshots matrix of X, The goal of DMDc is to find the best linear operators A and B to approximate the dynamical system with actuation. In this case, to construct the relationship between unknown A, B with known data X and Γ, (3) is rewritten as [29]:

$$X' \approx \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} X \\ \Gamma \end{bmatrix} = C\Omega \tag{6}$$

where $C = [A\ B]$, $\Omega = [X\ \Gamma]^T$. With a similar form as DMD in [29], DMDc is defined as:

$$C = X'\Omega^{\dagger} \tag{7}$$

where C can be found by minimizing the Frobenius norm $\|C - X'\Omega^{\dagger}\|_F$. As in DMD, SVD is used on the augmented input data matrix Ω, that is, $\Omega = U\Sigma V^* \approx \tilde{U}\tilde{\Sigma}\tilde{V}^*$, with the truncation value r. Then, the approximated linear operators A and B can be found as:

$$\begin{aligned} C &\approx X'\tilde{V}\tilde{\Sigma}^{-1}\tilde{U}^* \\ [A\ \ B] &\approx [X'\tilde{V}\tilde{\Sigma}^{-1}\tilde{U}_1^*\ \ X'\tilde{V}\tilde{\Sigma}^{-1}\tilde{U}_2^*] \end{aligned} \tag{8}$$

where $C \in \mathbb{R}^{n\times(n+l)}, \tilde{U}_1 \in \mathbb{R}^{n\times r}, \tilde{U}_2 \in \mathbb{R}^{l\times r}, \tilde{U} = [\tilde{U}_1\ \tilde{U}_2]$. To find a reduced order representation of dynamic systems, a second SVD on the output matrix X' is used [29], where $X' \approx \hat{U}\hat{\Sigma}\hat{V}$ with the truncation value p, here $\hat{U} \in \mathbb{R}^{n\times p}, \hat{\Sigma} \in \mathbb{R}^{p\times p}, \hat{V}^* \in \mathbb{R}^{p\times m-1}$. Then the reduced order system matrices for A and B can be realized with the linear transformation $x = \hat{U}\tilde{x}$:

$$\begin{aligned} \tilde{A} &= \hat{U}^*A\hat{U} = \hat{U}^*X'\tilde{V}\tilde{\Sigma}^{-1}\tilde{U}_1^*\hat{U} \\ \tilde{B} &= \hat{U}^*A\hat{U} = \hat{U}^*X'\tilde{V}\tilde{\Sigma}^{-1}\tilde{U}_2^* \end{aligned} \tag{9}$$

The reduced-order dynamical system with the given control inputs can be constructed as follows:

$$\tilde{x}_{k+1} = \tilde{A}\tilde{x}_k + \tilde{B}\tilde{u}_k \tag{10}$$

After presenting the background for standard DMDc, POD is introduced next.

B. POD

POD is a widely used numerical method that reduces the computational burden in complex high-dimensional systems. The snapshot method is used to construct the POD modes of the ensemble data $x_i, i = 1, \cdots, N$ generated by the open-loop system. The $N \times N$ correlation matrix L is defined as [1] [30]:

$$L_{i,j} =< x_i, x_j > \tag{11}$$

where:

$$< x_i, x_j >= \int_\Omega x_i x_j^* dx \tag{12}$$

belongs to $L^2(\Omega)$ standard inner product and x_j^* is the complex conjugate of x_j, where Ω is the spatial domain. Next, compute the eigenvalues of L, and sort the first M eigenvalues in descending order such that [1]:

$$(\frac{\sum_{i=1}^{M} \lambda_i}{\sum_{i=1}^{N} \lambda_i}) \geq 1 - \delta \tag{13}$$

where the ratio is the percentage of energy contained in the ensemble snapshots that is preserved in the POD basis. By choosing a desired relatively small δ, the smallest M is determined and each eigenvector $\{\alpha_i\}_{i=1}^{M}$ is normalized as :

$$\|\alpha_i\|^2 = 1/\lambda_i \tag{14}$$

The orthonormal POD basis $\{\phi_i\}_{i=1}^{M}$ can be constructed as:

$$\phi_i(x) = \sum_{j=1}^{N} \alpha_{i,j} x_j \tag{15}$$

where $\alpha_{i,j}$ is the jth component of α_i. The approximated solution X is a linear combination of the POD basis described as:

$$X \approx \sum_{i=1}^{M} \alpha_i \phi_i \tag{16}$$

where $\phi_i(x)$ is the i-th POD basis and $\alpha_i(t)$ is its corresponding temporal coefficient. The POD reduced order mode can be represented as [1]:

$$\min_{\phi} \|X - \sum_{i=1}^{M} \alpha_i \phi_i\|_F^2 \qquad s.t. < \phi_i, \phi_j >= \delta_{ij} \tag{17}$$

C. Clustered Reduced Order Basis

This work is a continuation of our previous work [11] where we used K-means clustering in conjunction with DMDc. It turns out that the particular clustering employed affects the performance of the reduced order model. The K-means clustering method groups the snapshots by calculating the minimal distance between each data point and the kth cluster centroid according to the Euclidean distance [25]:

$$d(x_i, x_j) = \sqrt{(x_i - x_j)^T (x_i - x_j)} \tag{18}$$

where d is the Euclidean distance between two distinct snapshots x_i and x_j.

Let c_i be the argument of the minimum distance between x_i and x_{c_j}, i.e.,

$$c_i = \arg \min_{j=1,\ldots,K} d(x_i, x_{c_j}) \tag{19}$$

The new centroids are:

$$x_{c_j} = \frac{\sum_{i=1}^{N} 1_{c_i=j} x_i}{\sum_{i=1}^{N} 1_{c_i=j}} \tag{20}$$

where $j = 1, \ldots, K$, K is the number of clusters, and $1_{c_j=j} = 1$ if $c_i = j$ and $1_{c_j=j} = 0$ otherwise.

If $\tilde{S}$ is the set of all the snapshots, and if $\tilde{S}_i$ represents the i^{th} cluster with center x_{c_i}, then $\tilde{S} = \cup_{i=1}^{k} \tilde{S}_i$, where the union is a union of disjoint sets. It is readily seen that the K-means algorithm is based on the notion that a cluster is centered around a single point when measuring similarity. However, snapshots generated by simulation or experiments from processes governed by nonlinear PDEs reside on low-dimensional manifolds. It is more advantageous to consider clusters as groups of snapshots around compact manifolds leading to manifold clustering [21]. Here, we adopt the manifold clustering method proposed in [18] where the data is assumed unorganized, i.e., it is not known which snapshots belong to which manifolds and the latter may intersect. In addition, the method identifies the number of manifolds and their intrinsic dimensions and partitions the snapshots into the manifolds they belong to. To deal with the presence of single and intersecting manifolds, unsymmetrical normalized spectral clustering [31] is first applied to partition the snapshots coarsely into different connected subsets according to the following steps [18]:

- Construct a similarity graph $G = (V, E)$ with vertex set V and edge set E: draw an edge between snapshot X_i and X_j if X_i is among the L nearest neighbors of X_j, and vice versa. The nearest neighbor is determined if the Euclidean distance $d(X_i, X_j)$ is less than a threshold.
- Construct the weight matrix $W = (w_{ij})$ as $w_{ij} = 1$ if snapshots X_i and X_j are connected, and $w_{ij} = 0$ if not.
- Spectral decomposition: Define the diagonal matrix F with $F_{ii} = \sum_j w_{ij}$, and the matrix $E = F - W$. Compute the first r eigenvectors u_1, u_2, $\cdots$, u_r corresponding to the r smallest eigenvalues by solving the generalized eigenvalue problem. The number r is determined by using the so-called eigenmap heuristic [32]:

$$\text{If } \ |\lambda_j - \lambda_{j-1}| \leq 10^{-6} < |\lambda_{j+1} - \lambda_j| \ \text{ then } \ r = j \tag{21}$$

  where 10^{-6} is used to replace zero to void numerical issues. Let the matrix $U := [u_1, u_2, \cdots, u_r]$.
- Cluster by the K-means algorithm: Group the vectors corresponding to the rows of the matrix U using K-means.

After the different groups have been identified, we need to determine whether their structure is intersecting or single. In [18], it is suggested to resort to the intrinsic dimension id, that is, if the snapshots belong to a single manifold then the intrinsic dimension of each point on this manifold should be the same, otherwise, they are different. The points in the intersecting areas have, in general, higher dimensions than the other areas. Thus, the snapshots with the highest dimension d_{max} are grouped in the intersecting manifolds [18]. The latter can be accurately identified using the ϵ-neighbors rule.

The intersecting area due to different manifolds passing across each other should be revealed [18]. Here the K-plane clustering algorithm [33] is employed to reveal the different manifolds in each intersecting cluster. Given the number of

clusters k and their dimensions $d_1, d_2, \cdots, d_k$, the steps are as follows [33] [18]:

- Initialization: Randomly assign each snapshot to a cluster to obtain an initial partition. Then, alternate between the following steps until convergence.
- Cluster update: Find a center μ_i and a set of bases $\Phi_i = [\phi_{id_1}, \phi_{id_2}, \cdots, \phi_{id_k}]$ for the i-th cluster such that the reconstruction error is minimum.
- Cluster assignment: For each snapshot X_m in the intersecting clusters, determine the space j such that:

$$(X_m - \mu_j)^T (I - \Phi_j \Phi_j^T)(X_m - \mu_j) = \min_{i=1,2,\cdots,k} (X_m - \mu_i)^T (I - \Phi_i \Phi_i^T)(X_m - \mu_i) \quad (22)$$

The intrinsic dimension id is estimated locally by using the local covariance matrix of the snapshots. More specifically, for each snapshot X_i, find its nearest M nearest snapshot neighbors $X_i^1, X_i^2, \cdots, X_i^M$, and compute the local covariance matrix Cov_i as [34]:

$$Cov_i = \frac{1}{M} \sum_{j=1}^{M} (X_i^j - m_i)(X_i^j - m_i)^T \quad (23)$$

$$m_i = \frac{1}{M} \sum_{j=1}^{M} X_i^j \text{ is the mean vector} \quad (24)$$

The intrinsic dimension is then determined by sorting the eigenvalues of Cov_i in descending order $\lambda_1^i \geq \lambda_2^i \geq \lambda_2^i \cdots$, and if [18]:

$$\frac{\lambda_j^i}{\lambda_1^i} < 0.05 \leq \frac{\lambda_{j-1}^i}{\lambda_1^i} \quad (25)$$

then, $id = j - 1$. In the next section, the clusters $\tilde{S}_i$ are used to derive cluster DMDc and POD. The manifold clustering algorithm can be summarized as follows:

Algorithm 1: Manifold clustering

Input: Snapshots dataset $X = x_1, \ldots, x_n$

1 Create a graph $G = (V, E)$ with a node for each snapshots, and an edge between pairs of neighboring snapshots and set the edge weight to that distance;

2 Create $k \times n$ weight matrix $W = [w_{ij}]$, take spectral decomposition on W obtaining eigenmap

3 For each connected subset, compute the intrinsic dimension id, select the connected subset as a cluster if $id's$ are equal. if $id's$ are different, construct a new graph G_{new} and repeat step 2

4 Repeat until Step 3 converges

Output: $s_1, s_2, \ldots, s_k$: the results of clustering

D. Cluster DMDc and POD

Each cluster of snapshots $\tilde{S}_i$ is used to build the matrix of snapshots X_i at time k and X_i' at time $k+1$. In this case, since for any $m \times n$ matrix $M = (m_{ij})$, $\|M\|_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} |m_{ij}|^2$. The following Lemma, proved in [11], shows that clustering DMD/DMDc provides better results than standard DMD/DMDc in general:

Lemma: [11]

$$\begin{aligned} \min_A \|X' - AX\|_F^2 &= \min_A \sum_{i=1}^{K} \|X_i' - AX_i\|_F^2 \\ &\geq \sum_{i=1}^{K} \min_A \|X_i' - AX_i\|_F^2 \\ &= \sum_{i=1}^{K} \|X_i' - A_i X_i\|_F^2 \end{aligned} \quad (26)$$

where the sum is taken over K clusters. However, the clustering technique used does affect the performance of the reduced-order model as shown here. A similar result can be shown to hold for POD as well.

III. Numerical Results

A. A Prototype Nonlinear Convective Flow

Let $\Omega_1 \subseteq \mathbb{R}^2$ be the rectangle given by $(a, b] \times (c, d)$ [35] [27]. Let $\Omega_2 \subseteq \Omega_1$ be the rectangle given by $[a_1, a_2] \times [b_1, b_2]$ where $a < a_1 < a_2 < b$ and $c < b_1 < b_2 < d$. The problem domain, Ω, is given by $\Omega = \Omega_1 \backslash \Omega_2$. In this configuration, Ω_2 is the obstacle. Dirichlet boundary controls are located on the obstacle bottom and top, denoted by Γ_B and Γ_T, respectively. The dynamics are described by the 2D Burgers' equation [35] [27]:

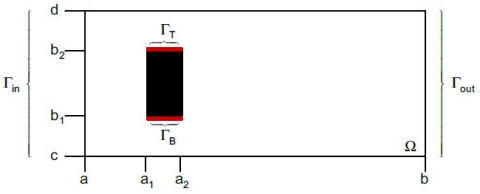

Fig. 1. Problem Geometry

$$\frac{\partial}{\partial t} w(t, x, y) + \nabla \cdot F(w) = \frac{1}{Re} \Delta w(t, x, y) \quad (27)$$

for $t > 0$ and $(x, y) \in \Omega$. In (27) $F(w)$ has the form

$$F(w) = \left[C_1 \frac{w^2(t, x, y)}{2} \quad C_2 \frac{w^2(t, x, y)}{2} \right]^T \quad (28)$$

where C_1, C_2 are non-negative constants. This system has a convective nonlinear term similar to the Navier-Stokes equations governing turbulent fluid flows. The quantity Re is the Reynolds number counterpart. Here, the boundary control is assumed to be separable for simplicity [35] [27]. Under this assumption, at the bottom and top of the obstacle, boundary conditions are specified as [27], [35]:

$$\begin{aligned} w(t, \Gamma_B) &= u_B(t) \Psi_B(x) \\ w(t, \Gamma_T) &= u_T(t) \Psi_T(x) \end{aligned} \quad (29)$$

where $u_B(t)$ and $u_T(t)$ are the boundary controls at the obstacle bottom and top, respectively. The profile functions $\Psi_B(x)$ and $\Psi_T(x)$ describe the spatial influence of the controls on the boundary. A parabolic inflow condition is specified of the form [27], [35]:

$$w\left(t, \Gamma_{in}\right) = f(y) \tag{30}$$

At the outflow, a Neumann condition is specified according to [27], [35]

$$\frac{\partial}{\partial x} w\left(t, \Gamma_{out}\right) = 0 \tag{31}$$

The boundary values along Γ_U are supposed to be zero for all times, that is,

$$w(t, \Gamma_U) = 0 \tag{32}$$

The initial condition is given as

$$w(0, x, y) = w_0(x, y) \in L^2(\Omega). \tag{33}$$

B. Numerical setting

To generate the snapshots, a finite difference approximation [35] is used to obtain the solution snapshots for the convective flow from left to right with parabolic inflow condition $f(y)$, and where $C_1 = 1$ and $C_2 = 0$. The rectangle Ω_1, Ω_2 are set to be $(0, 0.99] \times (0, 0.48]$ and $[0.15, 0.24] \times [0.15, 0.33]$. The parameter Re is specified as 300, 500, 800. Finite difference spatial discretization is performed as in [35]. A grid with spatial size $h = 0.015$ is built to generate the snapshots using inputs of the following forms:

$$u_B(t) = m_1 \sin\left(\frac{3}{16}\pi t^{f_1}\right), \; u_T(t) = m_2 \sin\left(\frac{3}{8}\pi t^{f_2}\right) \tag{34}$$

where the steady state solution arising from the inflow is taken as the initial condition in the simulation. In Equation (34), snapshots are taken with $\Delta t = 0.02$ increments starting from $t = 0$ and ending at $T = 40$. The resulting ensemble consists of roughly 2000 time snapshots each with dimension 2156.

C. Reduced Order Modeling

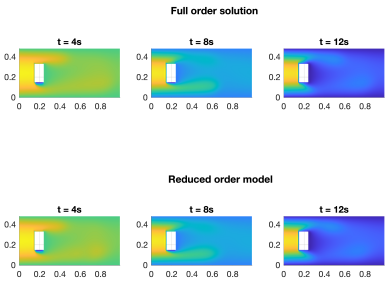

Fig. 2. Under case 3, the full order solution and the manifold-clustered DMDc reduced order model

Here, two types of cluster-reduced order models are constructed, respectively. Different Re values with different boundary inputs u_B, u_T are tested.
1) $Re = 300, u_B = \sin(3/16 \times \pi t^{1.4}), u_T = 2 \times \sin(3/8 \times \pi t^{1.4})$,
2) $Re = 500, u_B = \sin(3/16 \times \pi t^{1.4}), u_T = 2 \times \sin(3/8 \times \pi t^{1.4})$,
3) $Re = 800, u_B = \sin(3/16 \times \pi t^{1.4}), u_T = 2 \times \sin(3/8 \times \pi t^{1.4})$
4)$Re = 800, u_B = \sin(3/16 \times \pi t^{1.6}), u_T = 2 \times \sin(3/8 \times \pi t^{1.2})$
5) $Re = 800, u_B = 2\times\sin(3/16\times\pi t^{1.6}), u_T = 3\times\sin(3/8\times \pi t^{1.2})$. Here we compare the accuracy of the reduced order

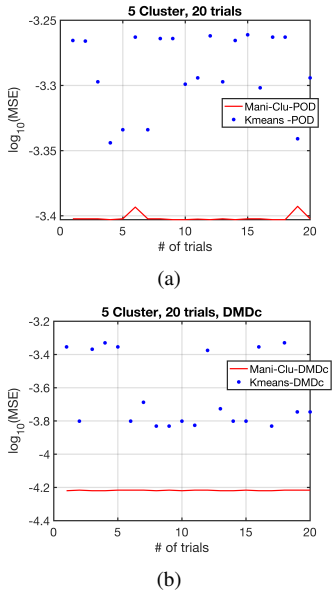

Fig. 3. Under case 2 and 4, the K-means and Manifold cluster-DMDc/POD reduced order models are constructed. The MSE between the reduced-order and full-order models is displayed after 20 runs.

model between manifold clustering and K-means clustering combined with DMDc and POD. $k = 5$ is chosen to generate clusters for each method. As mentioned in the previous section, K-means clustering randomly chooses the centroids at the beginning of the clustering process, which leads to different constructed reduced-order models. For each case, 20 runs are set for the performance comparison. For instance, as seen in Fig. 2, the manifold-clustered DMDc reduced order model is shown to be efficient in representing the full-order solution well. To evaluate the performance, the mean square error (MSE) metric is used. In Fig. 3, the K-means and manifold clustering methods are applied with POD to obtain the reduced-order models. The proposed manifold method outperforms the K-means cluster method. We take the experiment for 5 cases with different cluster methods combined with 2 reduced order techniques and take the average MSE over 20 trials for each case. More details can be found in Table I, under the condition given in the 5 cases with $k = 5$ clusters, the clustered-DMDc reduced order model has better performance than the clustered-POD reduced order model. Here, our purpose is to investigate how the manifold cluster reduced-order model performs from case 1 to case 5.

The non-linearity of the system increases as Re increases and more complex boundary inputs result. The manifold clustered reduced models are all better than the K-means clustered reduced order models.

$log_{10}(MSE)$	K-means		Manifold-cluster	
	POD	DMDc	POD	DMDc
Case 1	-3.20	-3.8	-3.33	-4.14
Case 2	-3.24	-3.74	-3.39	-4.21
Case 3	-3.38	-3.63	-3.49	-4.12
Case 4	-3.29	-3.56	-3.40	-4.08
Case 5	-2.71	-2.92	-2.94	-3.28

TABLE I

UNDER THE 5 CASES, THE PERFORMANCE OF THE K-MEANS AND MANIFOLD CLUSTERING REDUCED ORDER MODELS CAN BE VIEWED IN THE TABLE. FOR EACH CASE, WE RUN THE CODE 20 TIMES. FOR INSTANCE, FOR CASE 4, THE PERFORMANCE OF THE CLUSTERED-POD REDUCED ORDER MODEL CAN BE VIEWED IN FIG. 3

IV. CONCLUSIONS

A new model reduction method based on manifold clustering combined with POD and DMDc is proposed. Its performance is compared to the K-means cluster-DMDc and cluster-POD using a prototype nonlinear convective flow governed by the 2D Burgers' equation for different cases. The proposed manifold cluster method reduction outperforms both K-means cluster-DMDc and cluster-POD. Future work includes using the new reduced-order models to carry out flow feedback control for the full-order model.

ACKNOWLEDGMENT

This paper was supported in part by the National Science Foundation under grant NSF-CMMI-2024111.

REFERENCES

[1] P. Holmes, J. L. Lumley, G. Berkooz, and C. W. Rowley, *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge university press, 2012.

[2] J. H. Tu, *Dynamic mode decomposition: Theory and applications*. PhD thesis, Princeton University, 2013.

[3] C. W. Rowley, I. Mezić, S. Bagheri, P. Schlatter, and D. S. Henningson, "Spectral analysis of nonlinear flows," *Journal of fluid mechanics*, vol. 641, pp. 115–127, 2009.

[4] A. Towne, O. T. Schmidt, and T. Colonius, "Spectral proper orthogonal decomposition and its relationship to dynamic mode decomposition and resolvent analysis," *Journal of Fluid Mechanics*, vol. 847, pp. 821–867, 2018.

[5] I. Mezić, "Analysis of fluid flows via spectral properties of the koopman operator," *Annual Review of Fluid Mechanics*, vol. 45, pp. 357–378, 2013.

[6] K. K. Chen, J. H. Tu, and C. W. Rowley, "Variants of dynamic mode decomposition: boundary condition, koopman, and fourier analyses," *Journal of nonlinear science*, vol. 22, pp. 887–915, 2012.

[7] J. Burkardt, M. Gunzburger, and H.-C. Lee, "Pod and cvt-based reduced-order modeling of navier–stokes flows," *Computer methods in applied mechanics and engineering*, vol. 196, no. 1-3, pp. 337–355, 2006.

[8] S. Sahyoun and S. Djouadi, "Local proper orthogonal decomposition based on space vectors clustering," in *3rd IEEE International Conference on Systems and Control*, pp. 665–670, 2013.

[9] S. Sahyoun and S. M. Djouadi, "Time, space, and space-time hybrid clustering pod with application to the burgers' equation," in *53rd IEEE Conference on Decision and Control*, pp. 2088–2093, IEEE, 2014.

[10] S. Sahyoun, "Control oriented nonlinear model reduction for distributed parameter systems," 2017.

[11] T. Wu, D. Wilson, and S. M. Djouadi, "Adaptive cluster-dynamic mode decomposition with application to the burgers' equation," in *American Control Conference*, IEEE, 2024.

[12] A. Narasingam and J. S.-I. Kwon, "Development of local dynamic mode decomposition with control: Application to model predictive control of hydraulic fracturing," *Computers & Chemical Engineering*, vol. 106, pp. 501–511, 2017.

[13] J. Burkardt, M. Gunzburger, and H.-C. Lee, "Centroidal voronoi tessellation-based reduced-order modeling of complex systems," *SIAM Journal on Scientific Computing*, vol. 28, no. 2, pp. 459–484, 2006.

[14] Q. Du, V. Faber, and M. Gunzburger, "Centroidal voronoi tessellations: Applications and algorithms," *SIAM review*, vol. 41, no. 4, pp. 637–676, 1999.

[15] B. Telsang and S. M. Djouadi, "Computation of centroidal voronoi tessellations in high dimensional spaces," *IEEE Control Systems Letters*, vol. 6, pp. 3313–3318, 2022.

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

[17] D. Amsallem, *Interpolation on manifolds of CFD-based fluid and finite element-based structural reduced-order models for on-line aeroelastic predictions*. Stanford University, 2010.

[18] Y. Wang, Y. Jiang, Y. Wu, and Z.-H. Zhou, "Multi-manifold clustering," in *PRICAI 2010: Trends in Artificial Intelligence: 11th Pacific Rim International Conference on Artificial Intelligence, Daegu, Korea, August 30–September 2, 2010. Proceedings 11*, pp. 280–291, Springer, 2010.

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

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

[21] R. Souvenir and R. Pless, "Manifold clustering," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 1, pp. 648–653, IEEE, 2005.

[22] A. Babaeian, A. Bayestehtashk, and M. Bandarabadi, "Multiple manifold clustering using curvature constrained path," *PloS one*, vol. 10, no. 9, p. e0137986, 2015.

[23] A. Little, M. Maggioni, and J. M. Murphy, "Path-based spectral clustering: Guarantees, robustness to outliers, and fast algorithms," *Journal of machine learning research*, vol. 21, no. 6, pp. 1–66, 2020.

[24] Z. Li, Y. Chen, Y. LeCun, and F. T. Sommer, "Neural manifold clustering and embedding," *arXiv preprint arXiv:2201.10000*, 2022.

[25] G. Gan, C. Ma, and J. Wu, "Data clustering, theory, algorithms, and applications," 2007.

[26] R. C. Camphouse and J. Myatt, "Reduced order modelling and boundary feedback control of nonlinear convection," in *AIAA Guidance, Navigation, and Control Conference and Exhibit*, p. 5844, 2005.

[27] R. C. Camphouse, "Boundary feedback control using proper orthogonal decomposition models," *Journal of guidance, control, and dynamics*, vol. 28, no. 5, pp. 931–938, 2005.

[28] L. Yang and X. Pu, "Derivation of the burgers' equation from the gas dynamics," *Communications in Mathematical Sciences*, vol. 14, no. 3, pp. 671–682, 2016.

[29] J. L. Proctor, S. L. Brunton, and J. N. Kutz, "Dynamic mode decomposition with control," *SIAM Journal on Applied Dynamical Systems*, vol. 15, no. 1, pp. 142–161, 2016.

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

[31] J. B. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 22, pp. 888–905, 2000.

[32] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 4, no. 17, pp. 395–416, 2007.

[33] P. S. Bradley and O. L. Mangasarian, "K-plane clustering," *Journal of Global Optimization*, vol. 1, no. 16, pp. 23–32, 2000.

[34] K. Fukunaga and D. R. Olsen, "Algorithm for finding intrinsic dimensionality of data," *IEEE Transactions on Computers*, vol. C-20, no. 2, pp. 176–183, 1971.

[35] R. C. Camphouse and J. Myatt, "Feedback control for a two-dimensional burgers' equation system model," in *2nd AIAA flow control conference*, p. 2411, 2004.