ME493 - Methods of Data-Driven Control

Benjamin Aziel

2024-04-27

Table of contents

Preface **3**

1 The Four Fundamental Subspaces **4**

2 Singular Value Decomposition **6**
2.1 Overview 6
2.2 POD Analysis of Low Reynolds Number Pitching Airfoil DNS 6

3 Data-Driven Dynamical Systems **12**
3.1 Overview: Dynamic Mode Decomposition 13
3.2 DMD Analysis of Low Reynolds Number Pitching Airfoil DNS 13
3.3 Overview: Sparse Identification of Nonlinear Dynamics 15
3.4 SINDy Implementation on Temporal Amplitudes 15

4 Model Reduction and Projection **16**
4.1 Overview 16

5 System Identification Techniques **17**
5.1 Overview: Eigensystem Realization Algorithm 17
5.2 Overview: DMD with Control 17

References **18**

Preface

Documentation of the independent study ME493: Methods of Data-Driven Control, taken during Spring 2024.

Instructor: Dirk M. Luchtenburg **Co-conspirators:** Sohaib Bhatti, Khushant Khurana, Eunkyu (Q) Kim, Sonam Okuda **Email:** dirk.luchtenburg@cooper.edu

Description: This independent study provides an introduction to state-of-the-art methods employed in the field of data-driven engineering. Data-driven methods such as the proper orthogonal decomposition and the dynamic mode decomposition are used in a variety of fields including fluid dynamics, climate analysis, mixing problems, the study of infectious diseases, etc. Data-driven methods rely heavily on concepts from linear algebra, calculus, probability, and statistics. We will review these essential elements and develop a hands-on understanding of a selection of data-driven methods. The study culminates with an application of some method(s) to a practical problem which is to be carefully detailed in a technical report.

Referenced texts include the following:

- Strang (2019)
- Brunton and Kutz (2022)
- Boyd and Vandenberghe (2018)
- Géron (2022)
- Sutton and Barto (2018)

1 The Four Fundamental Subspaces

In the Strang-ian view of linear algebra, an m by n matrix $\mathbf{A}$ is associated with four fundamental subspaces - two of $\mathbb{R}^m$ and two of $\mathbb{R}^n$. It's easiest to illustrate this using an example matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 \end{bmatrix}$$

where:

$$\mathbf{r}_1 = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \quad \mathbf{r}_2 = \begin{bmatrix} 4 & 5 & 6 \end{bmatrix}$$

$$\mathbf{c}_1 = \begin{bmatrix} 1 & 4 \end{bmatrix}^T \quad \mathbf{c}_2 = \begin{bmatrix} 2 & 5 \end{bmatrix}^T \quad \mathbf{c}_3 = \begin{bmatrix} 3 & 6 \end{bmatrix}^T$$

Column Space

The column space (or range) $R(\mathbf{A})$ contains all linear combinations of the column vectors of $\mathbf{A}$.

It only takes two linearly independent vectors to span $\mathbb{R}^2$, and we have three! Our column space is $\mathbb{R}^2$.

Row Space

The row space $R(\mathbf{A}^T)$ contains all linear combinations of the column vectors of $\mathbf{A}^T$ (or equivalently, the row vectors of A).

Transposed row vectors $\mathbf{r}_1^T$ and $\mathbf{r}_2^T$ span the following plane:

$$\left\{ \begin{bmatrix} x_1 + 4x_2 \\ 2x_1 + 5x_2 \\ 3x_1 + 6x_2 \end{bmatrix} : \mathbf{x} \in \mathbb{R}^2 \right\}$$

Null Space

The null space $N(\mathbf{A})$ contains all solutions $\mathbf{u}$ to $\mathbf{Au} = \mathbf{0}$.

Solving this equation yields the nontrivial solution:

$$\mathbf{u} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

Our null space is the span of $\mathbf{u}$, or the following line:

$$\left\{ \begin{bmatrix} x \\ -2x \\ x \end{bmatrix} : x \in \mathbb{R} \right\}$$

Left Null Space

The left null space $N(\mathbf{A}^T)$ contains all solutions $\mathbf{v}$ to $\mathbf{A}^T\mathbf{v} = \mathbf{0}$.

This equation has no nontrivial solutions, so the left null space is the zero subspace.

The "big picture of linear algebra," as Gil Strang puts it, is that for an m by n matrix $\mathbf{A}$:

- The column space/range $R(\mathbf{A})$ is perpendicular to the left null space $N(\mathbf{A}^T)$ in $\mathbb{R}^m$
- The row space $R(\mathbf{A}^T)$ is perpendicular to the null space $N(\mathbf{A})$ in $\mathbb{R}^n$

Rank

The rank r of a matrix $\mathbf{A}$ is the number of independent rows/columns, i.e., the row space and column space/range have the same dimension r.

- The dimension of the null space $N(\mathbf{A})$ is $n - r$ and the dimension of the left null space $N(\mathbf{A}^T)$ is $m - r$.

2 Singular Value Decomposition

2.1 Overview

The singular value decomposition (SVD) is a fundamental matrix factorization with numerous applications in data analysis and scientific computing. Mathematically, the SVD of an $m \times n$ matrix $\mathbf{X}$ is a factorization of the form:

$$\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^*$$

where $\mathbf{U}$ is an $m \times m$ orthogonal matrix, Σ is an $m \times n$ diagonal matix with non-negative real numbers on the diagonal, and $\mathbf{V}$ is an $n \times n$ orthogonal matrix.

The SVD is particularly useful for analyzing large, high-dimensional datasets that can be well-approximated by matrices of much lower rank. By extracting the dominant patterms in the data, the SVD enables efficient dimensionality reduction, noise removal, and data compression. It is the foundation of techniques like principal component analysis (PCA) and is widely applied in fields such as signal processing, machine learning, and image analysis.

Proper orthogonal decomposition (POD) modes are a set of orthogonal basis functions that optimally represent a given dataset in a least-squares sense. They are obtained by performing an SVD on a data matrix. POD modes form an orthonormal basis, meaning the modes are mutually orthogonal and have unit norm.

The modes are ranked by their energy content, with the first mode capturing the most energy and subsequent modes capturing progressively less energy.

2.2 POD Analysis of Low Reynolds Number Pitching Airfoil DNS

The dataset analyzed contains direct numerical simulations (DNS) of two-dimensional stationary and pitching flat-plate airfoils at a Reynolds number of 100. The dataset includes time-resolved snapshots of the velocity field, lift and drag coefficients, and airfoil kinematics spanning 40-100 convective time units. The cases consist of a stationary airfoil and eight different pitching frequencies. This dataset is part of a database intended to aid in the conception, training, demonstration, evaluation, and comparison of reduced-complexity models for fluid mechanics, created by Aaron Towne and Scott Dawson.

The dataset also includes a MATLAB function that provides a simple implementation of Dynamic Mode Decomposition (DMD), a data-driven method we'll get to later.

To analyze the DNS data using POD, I first extracted the velocity components from the provided snapshots. I then computed the mean-corrected snapshots by subtracting the mean of each snapshot and arranged them into a matrix $\mathbf{X}$.

Next, I performed an economy-sized SVD on $\mathbf{X}$. The squared singular values are plotted here, as a function of the mode index.

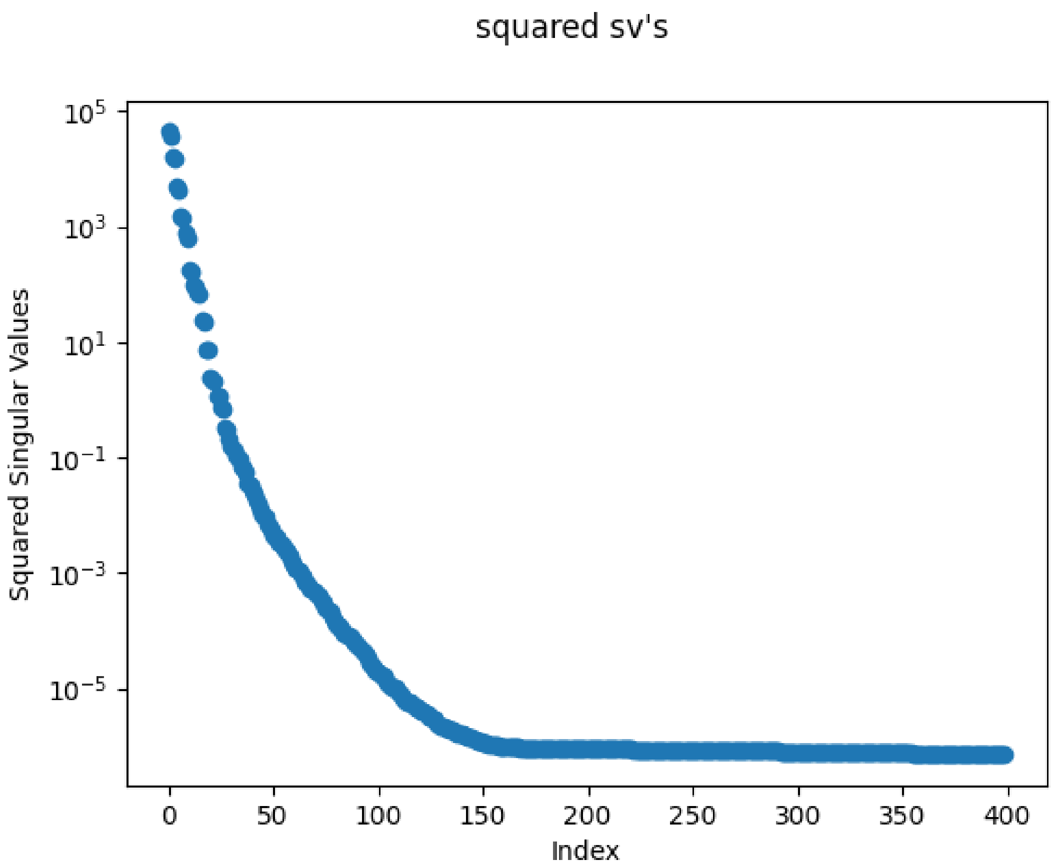


Figure 2.1: squared_sv

This plot reveals how many individual POD basis vectors there are. The rapid decay of the singular values indicates that the flow is well-approximated by a low-rank subspace, with the first 16 modes capturing the majority of the energy.

Visualizing the first six POD modes for both u_x and u_y revealed the spatial structure of the dominant flow patterns. The oscillatory modes have a characteristic wavelength that can be estimated from the spatial distribution of the mode amplitudes.

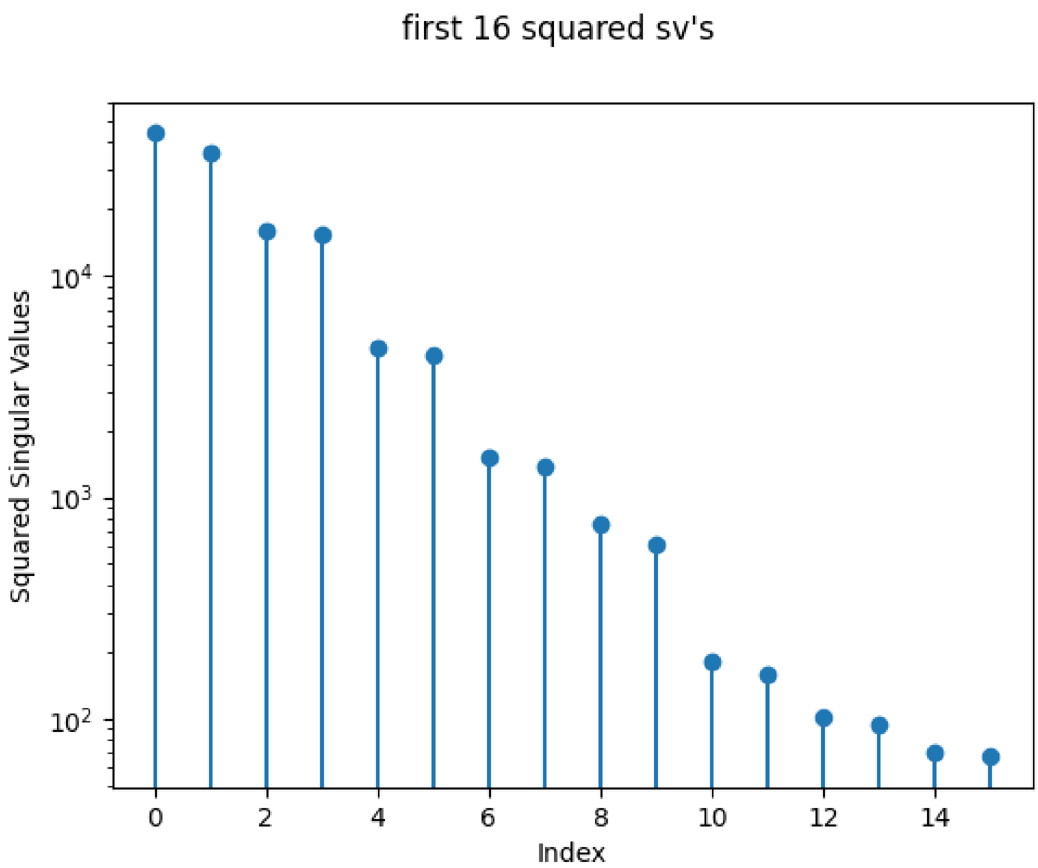


Figure 2.2: squared_sv_truncated

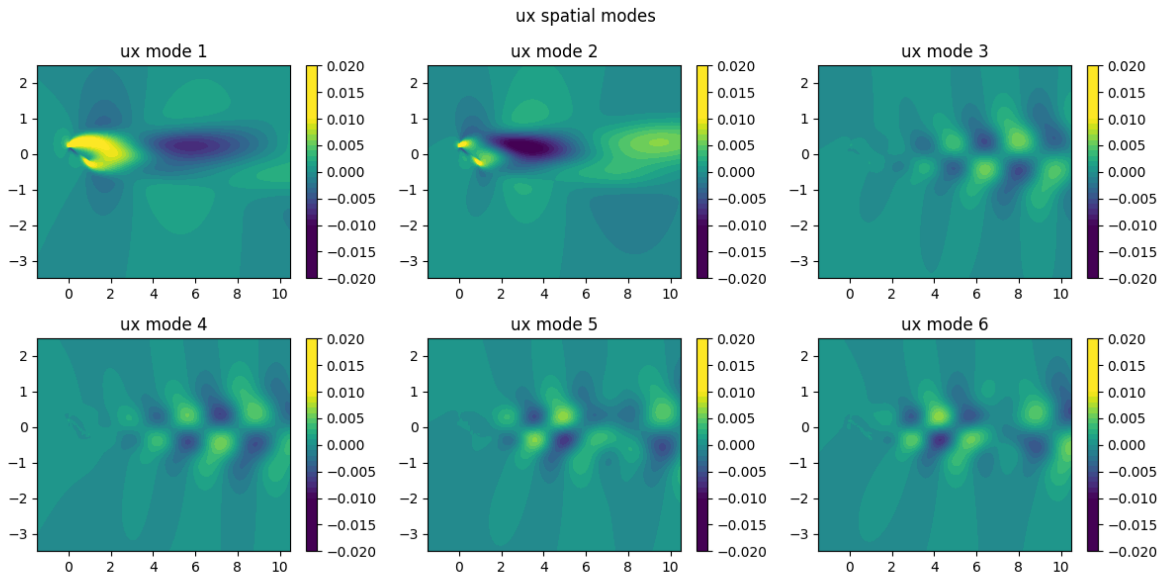


Figure 2.3: ux_spatial_modes

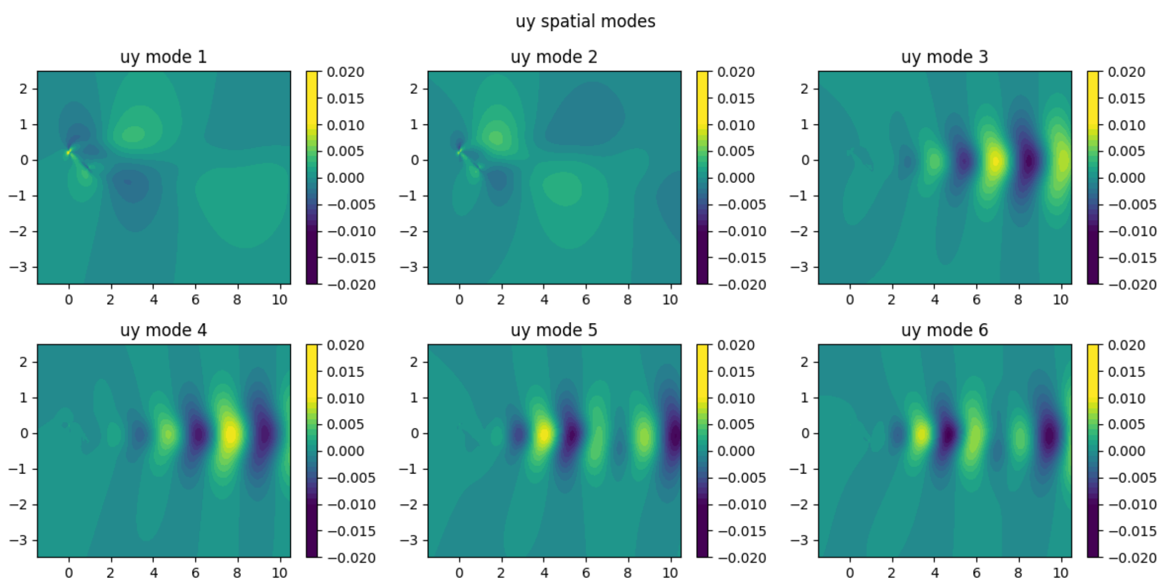


Figure 2.4: uy_spatial_modes

Here are the temporal amplitudes associated with those spatial modes:

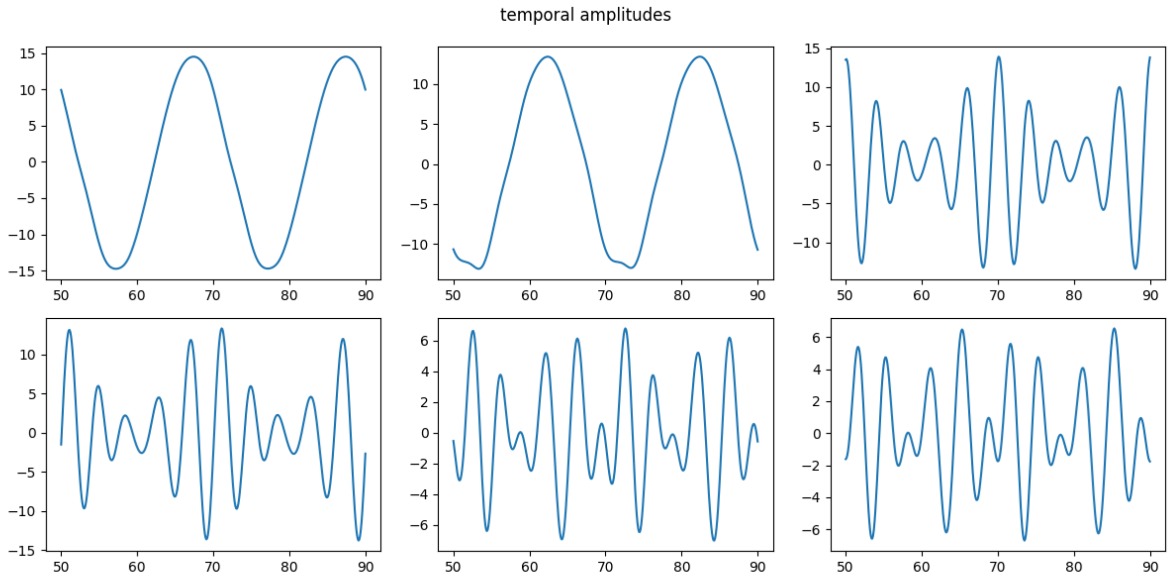


Figure 2.5: temporal_amplitudes

Finally, I reconstructed the snapshots using a rank-4 approximation, which captures the most energetic flow structures. Comparing the reconstructed snapshots with the original data showed that the low-rank approximation successfully recovers the essential flow physics, such as the presence of coherent structures and the overall flow patterns.

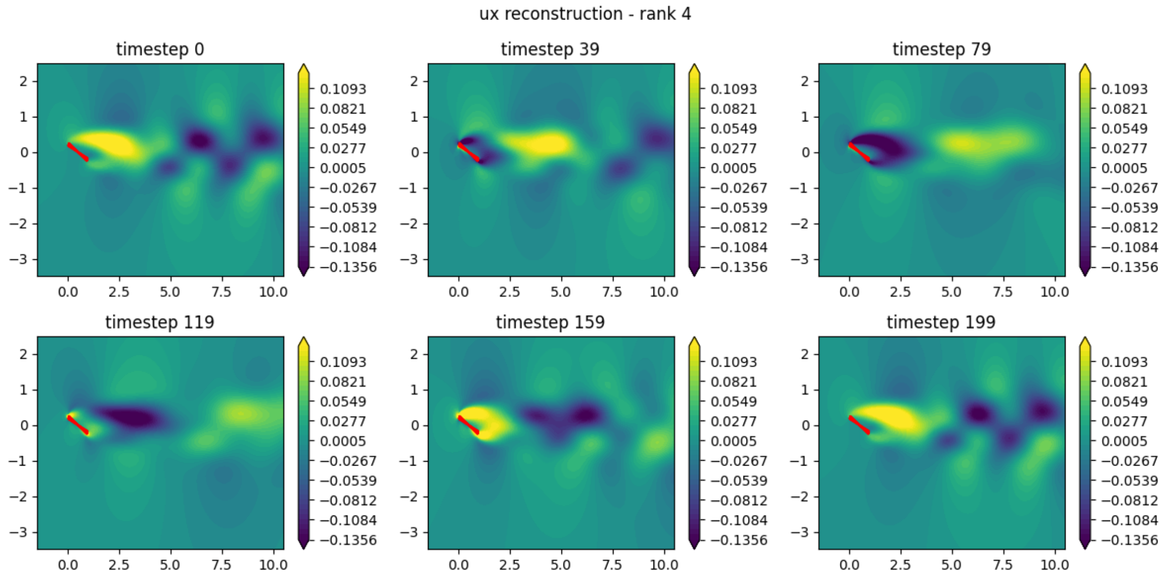


Figure 2.6: ux_reconstruction

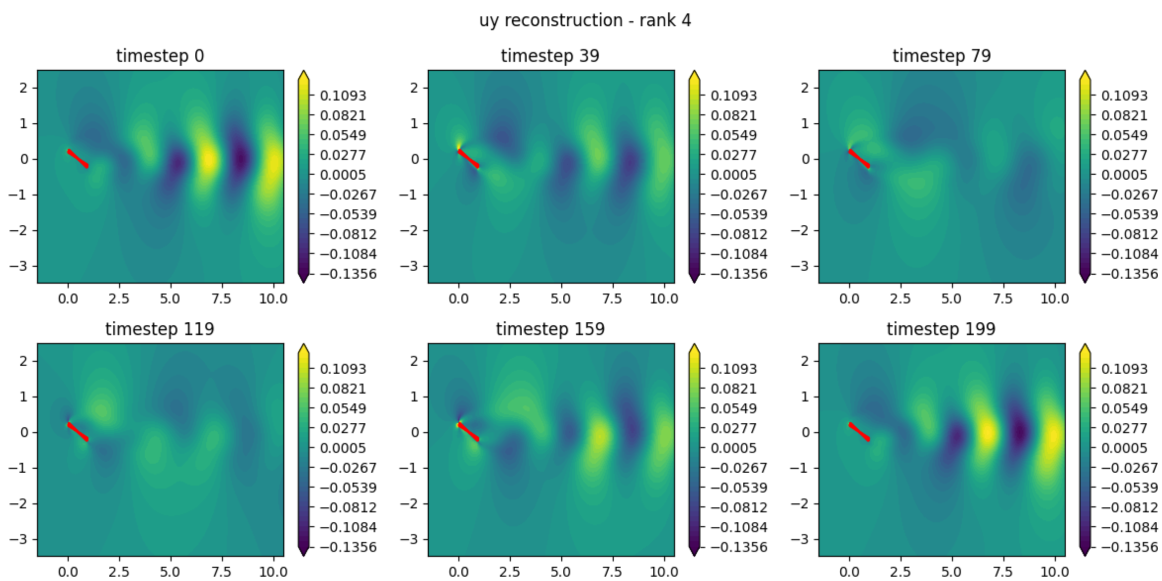
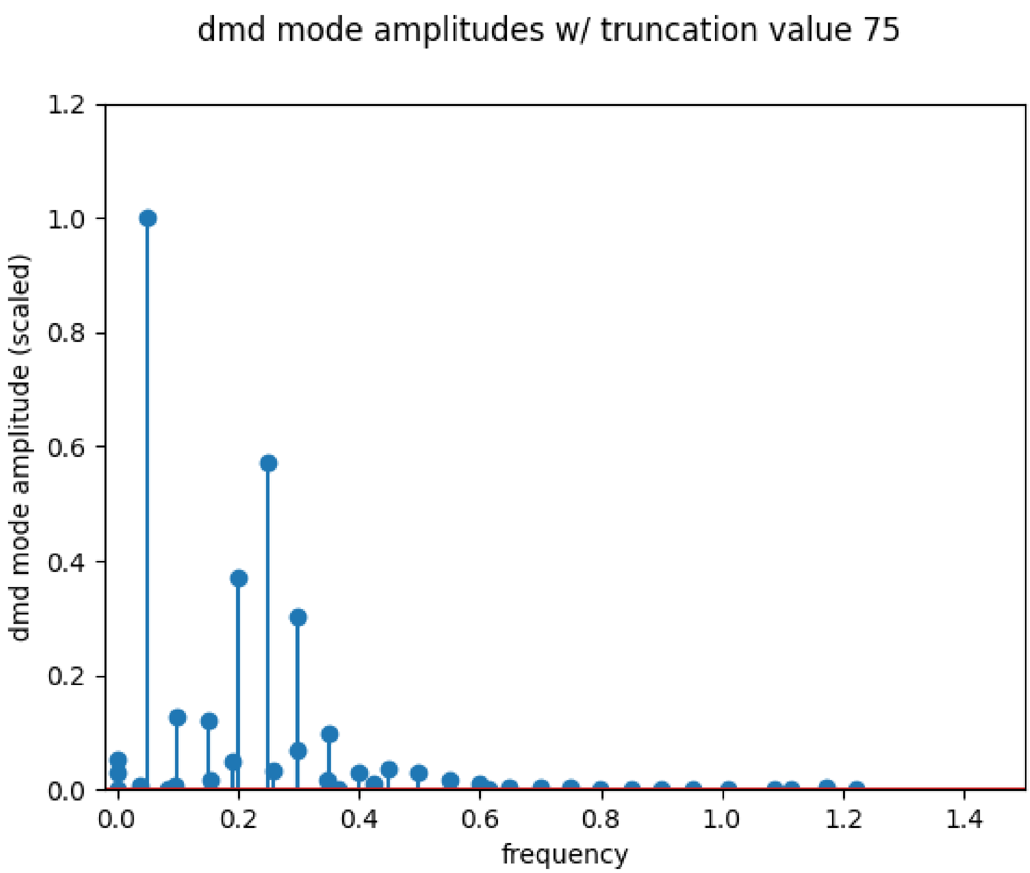


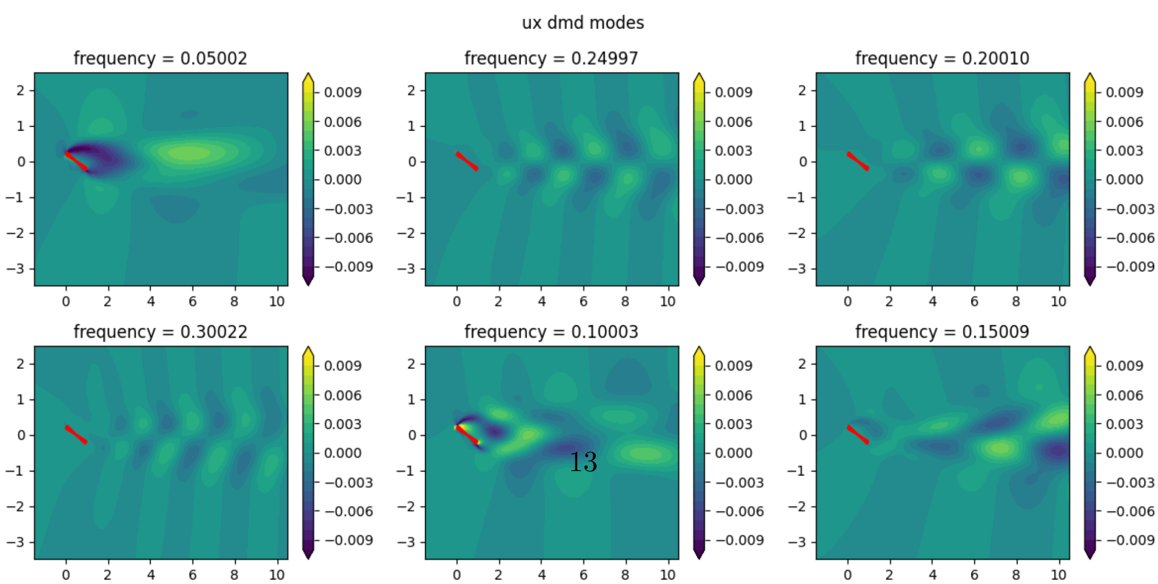
Figure 2.7: uy_reconstruction

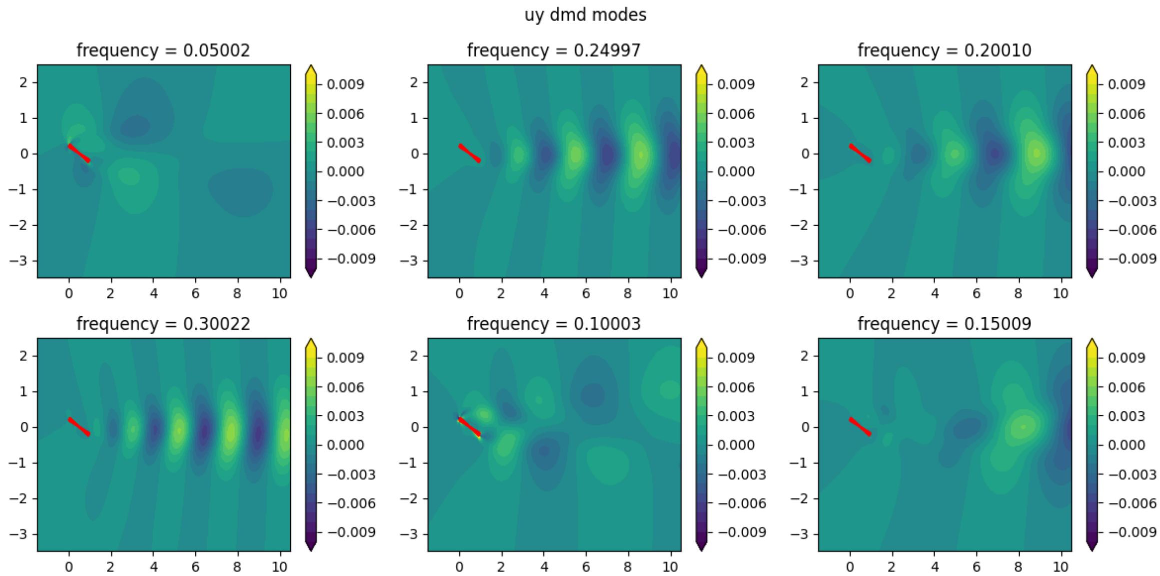
3 Data-Driven Dynamical Systems

3.1 Overview: Dynamic Mode Decomposition

3.2 DMD Analysis of Low Reynolds Number Pitching Airfoil DNS

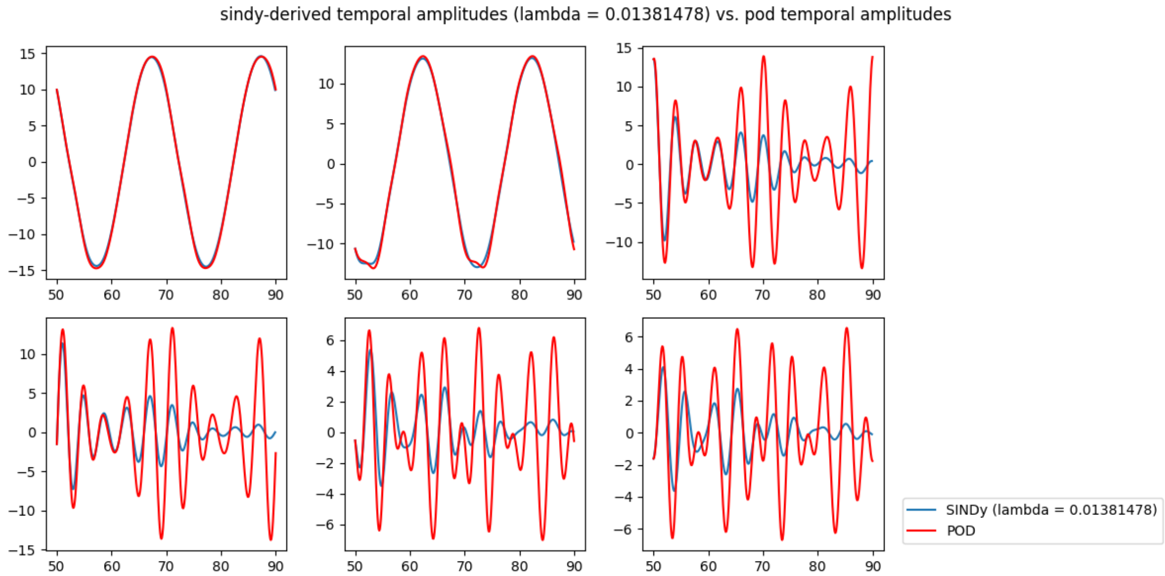


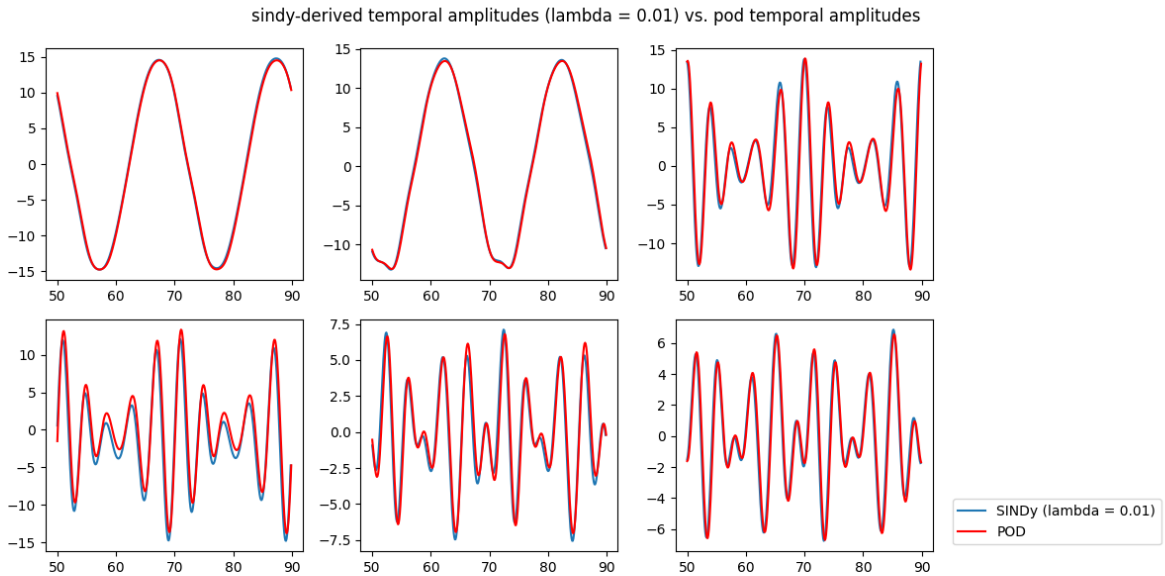



uy dmd modes
frequency = 0.05002
frequency = 0.24997
frequency = 0.20010
frequency = 0.30022
frequency = 0.10003
frequency = 0.15009

3.3 Overview: Sparse Identification of Nonlinear Dynamics

3.4 SINDy Implementation on Temporal Amplitudes





4 Model Reduction and Projection

4.1 Overview

5 System Identification Techniques

5.1 Overview: Eigensystem Realization Algorithm

5.2 Overview: DMD with Control

References

Boyd, Stephen, and Lieven Vandenberghe. 2018. *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares*. Cambridge University Press.

Brunton, Steven L, and J Nathan Kutz. 2022. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Vol. 2. Cambridge University Press.

Géron, Aurélien. 2022. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc.

Strang, Gilbert. 2019. *Linear Algebra and Learning from Data*. SIAM.

Sutton, Richard S, and Andrew G Barto. 2018. *Reinforcement Learning: An Introduction*. MIT Press.