MEL2040 Project
Data-Driven Analysis of Fluid Flows

February 12, 2024

Report Weightage: 20% **Report Due Date:** 12th April 2024, 23:59:59

Introduction

This project aims to give you an exposure to the intersection of two wonderful domains - machine learning and fluid mechanics. Machine learning is a field that is gaining a lot of momentum. Its applicability is increasing in various domains including fluid mechanics. Since fluid mechanics is a domain that generates a lot of data, numerous data-driven methods make use of this data to optimize systems and improve their performance. There is immense potential for machine learning-based fluid applications in the domains of aerospace, gas turbines, automotive, and heavy industries and several researchers across the world are working in this cutting-edge area. Numerous machine-learning methods exist in the domain of fluid mechanics. However, we shall look at one of the most simple, elegant and widely used methods - the proper orthogonal decomposition (POD).

I hope you enjoy this project and learn several tools and concepts in both fluid mechanics and machine learning. The skills you garner from this project will give you the confidence to apply machine learning to core mechanical engineering problems and also take you a long way toward other aspirations that you may have. Please don't hesitate to ask any questions during this project.

General Guidelines

- The project has to be conducted in **groups of 2 only**. No individual submissions. Please inform us of your group by **16th Feb** in the attached Google sheet. After this date, we shall assign groups randomly.
- The **project report** (and codes) carries an overall weightage of **20%** in the FM course. An additional **8%** of the overall weightage is for a **presentation** and **viva**. The presentation and viva will be held after the submission date.
- Please use **Python** for coding. Please submit the report as **ID1_ID2.pdf** and zip all your codes into one folder.

- **3 Bonus Marks** will be awarded for writing the report in LaTeX.[1]
- Some of the sections have a **page limit**. Please adhere to it.
- For all tasks in this project, you are encouraged to provide any **additional**[2] analysis, observations, and performance metrics that might help you and us evaluate the usefulness and novelty of your workflow.
- You already have the required skill set (through previously completed courses) to start working on this project immediately.
- We have asked ChatGPT and other AI tools to solve this project. You will be happy to know - it solved it absolutely incorrectly, even after multiple prompts! (This includes Sections 1 and 2 of this project). All the best! :)
- We are happy to answer your questions. Please don't reach out to us at the last minute, though!

1 Review of Machine Learning in Fluids (15 marks)

While trying to develop a new product or trying to undertake research in a domain, one of the first steps is to conduct a literature review or survey of existing products, ideas, and applications. This gives the researcher an overview of what potential they have to undertake new research or bring in innovation.

Your first task is to write a literature review of some of the advances of machine learning in the domain of fluid mechanics. You should base your literature review on what you understand from research papers. You may refer to Google Scholar for research papers and/or books. Make sure you understand what you write - as you will be asked questions on it! :)

Your literature review need not necessarily be on POD. It can be either on a topic of your choice or a mixture of topics[3]. Please keep the response of this section under 2 pages.

2 Any New Ideas (10 marks)

Propose 4 new[4] ideas of how machine learning can be applied in the domain of fluid mechanics. Ideally, these ideas should be based on existing literature. But we are happy if you think of something novel and applicable. Please keep the response to this section under 2 pages. Please don't trust the response of ChatGPT and other AI tools to the response they provide for these sections (we know its level of understanding regarding the topic).

[1] By the way, this document is LaTeXgenerated.
[2] Additional, here, refers to anything beyond what is asked to be done in the specific question.
[3] While evaluating this section we acknowledge the fact that this is only your first course on fluid mechanics
[4] New means - something which does not exist!

3 Proper Orthogonal Decomposition (30 marks)

A decomposition is a type of mathematical operation that is performed to break down a temporally evolving dynamical system into its individual components. The physical interpretation of these components can vary depending on the type of decomposition method used. One of the most famous methods used is the proper orthogonal decomposition (POD). Fun fact - this method was developed specifically for the analysis of fluid flows - turbulent flows[5]. The chaos of turbulent flows hides a lot of meaningful information that is required to uncover its physics. POD in a very broad sense, decomposes any dynamical system into its spatial and temporal components. These decomposed components are often referred to as the modes of the system. Having said that, since you have been introduced to principal component analysis (PCA) in your previous courses, we shall draw a similarity between PCA and POD for better understanding.

Any vector space can be represented by its basis vectors. For example, $\hat{i} = [1, 0, 0]^T$, $\hat{j} = [0, 1, 0]^T$ and $\hat{k} = [0, 0, 1]^T$ are the basis vectors of a Cartesian coordinate system or a Cartesian vector space. Any point/vector within this vector space can be represented as a combination of these basis vectors. Even though this is the commonly used basis system, there can be other basis vectors that are better suited to certain applications, especially in higher dimensional systems. PCA, as you may already know. is used for dimensional reduction. Here, the data is decomposed into a set of dimensions (smaller than the previous ones). This essentially projects the current n-dimensional vector space into another k-dimensional ($k < n$) vector space. Here, the k basis vectors for the k-dimensional space are formulated such that they capture the maximum possible variance from the original data. These k basis vectors are orthogonal to each other (i.e. $KK^T = K^T K = I$). Variance is the description of the spread of the data with respect to its mean, while covariance shows the relationship between two different variables.

$$\text{var}(x) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \tag{1}$$

$$\text{cov}(x, y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1} \tag{2}$$

$$\text{cov}(x, x) = \text{var}(x) \tag{3}$$

For n features, say $x_1, x_2 ... x_n$,

$$\text{cov}(x_1, x_2, ... x_n) = \begin{bmatrix} \text{var}(x_1) & \dots & \text{cov}(x_n, x_1) \\ \vdots & \ddots & \vdots \\ \text{cov}(x_1, x_n) & \dots & \text{var}(x_n) \end{bmatrix}$$

[5]Berkooz, G., Holmes, P. and Lumley, J.L., 1993. The proper orthogonal decomposition in the analysis of turbulent flows. Annual review of fluid mechanics, 25(1), pp.539-575.

One can thus perform an eigenvalue decomposition on the covariance matrix and then select the first k eigenvectors and their corresponding eigenvalues to reconstruct a k dimensional reduced space. This is what you have come to know as PCA. Looking closely, each eigenvector is a particular POD mode. The modes are arranged in their decreasing order of energy content. What this means is that POD decomposes a dynamical system in the decreasing order of the energy contained in each mode.

POD can also be performed by using the singular value decomposition (SVD). If you perform a bit of matrix algebra magic, you will be able to see how the covariance-PCA method is related to the SVD method. There are some specific advantages to using the SVD over the covariance matrix method (Hint: remember, POD decomposes systems into their spatial and *temporal* modes) and we strongly encourage you to look at this method and its physical interpretation. Additionally one may also use the Reduced SVD method for enhanced computational efficiency **(Hint: recommended)**.

The above was a brief introduction to POD, setting it in context to what you already know. You will have to refer to additional material - videos, books, and research papers to understand the POD formulation more carefully.

3.1 Image Generation (3 marks)

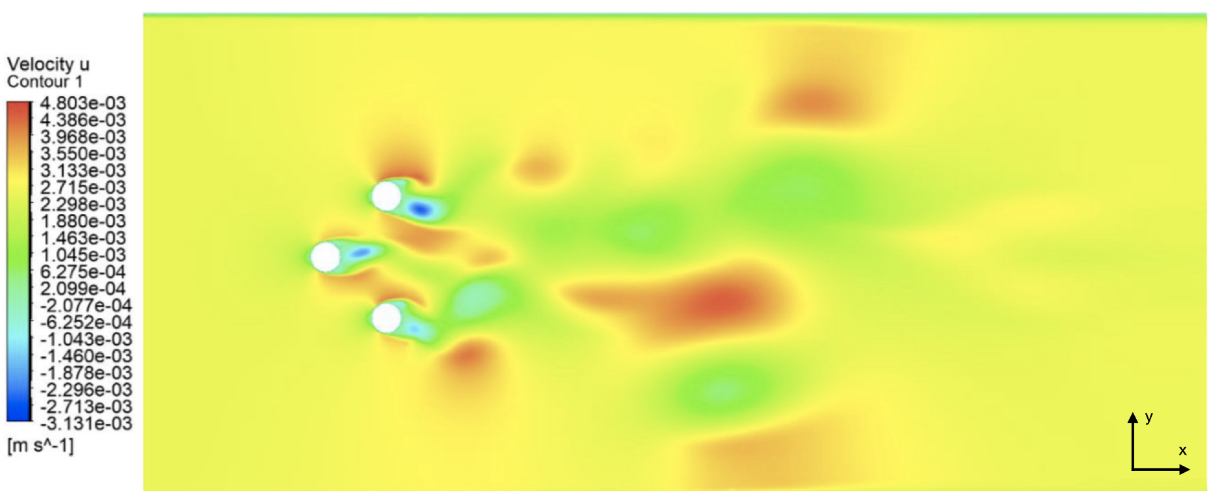

Figure 1: An instantaneous snapshot of flow around 3 cylinders.

You have been provided with a video of a flow over three cylinders present in a staggered configuration. The link to the flow is as follows: Video. A snapshot from the flow evolving in time has also been shown in Fig. 1. Additionally, information about the physical properties of the flow and simulation have been listed in Table 1 for your reference. The flow, as can be seen in the video, is evolving in time. To perform POD, you will have to first convert the video into a stack of images. This stack of images/frames shall be used for further analysis. You will get marks for the code that you provide for this task.

Table 1: Flow Parameters

Parameters	Value / Description
Fluid	Air
Fluid Viscosity	$1.7894 \times 10^{-5}\ Kg\,m^{-1}s^{-1}$
Fluid Density	$1.225\ Kg\,m^{-3}$
Inlet Velocity	$0.0025\ ms^{-1}$
Inlet Gauge Pressure	$0\ Pa$

3.2 Execute POD (12 marks)

Having obtained relevant frames, perform POD on the obtained images which are evolving in time. You can use any method to do so. However, it must not be performed using any POD finding functions executed and present in open-source specialized libraries. In the answer to this question, tell us about the mathematical formulation used. Include the code here.

3.3 Analyse POD Modes (15 marks)

Using the POD operation, present the description/plot of how the energy is distributed among the modes. Report the modes corresponding to the top 10 energy states of this dynamical system. What is the significance of these modes? Explain your answer and findings based on research articles that use POD.

4 Noise! (25 marks)

The data that you are currently analysing has been mined using computational fluid dynamics (CFD). Though there are several experimental techniques which involve analysing physical flow such as particle image velocimetery (PIV), Schlieren imaging etc. These images may contain a lot of discrepancies in the form of digital noise, sensory errors and imperfections in equipment. So, it is essential to analyse how this noise could affect our modal studies.

4.1 Adding Noise (10 marks)

We will do this artificially in this project. There are different types of artificial noise that one can add to an image. You have to explore these. Add different types of noise separately to the original frames that you have extracted. Additionally, the noise you add should have appropriate magnitude w.r.t the original image. These shall be 20%, 40%, 60% and 80% of the maximum magnitude of the individual frames.

4.2 Effect on POD Modes (15 marks)

Perform POD on these noisy images. Explain how the noise affects different modes.

- How does the energy transcend through modes?
- Can you give any statistical evidence of the energy changes in the modes?

- Do different noise types have different type of response in terms of how the energy of the modes changes?
- How does the magnitude of noise affect the modal energy?
- What sort of noise is best suited to be used for POD for artificial analysis?

Provide your comprehensive analysis. You may add any other information that you observe.

5 Super-Resolving (20 marks)

Any exceptional finds in this section will be given upto ***5*** *bonus marks (supernumerary).*

Now that you have created noisy/defective images and analysed how these defects affect the modes, you will now have to remove the noise that you have added (Don't just subtract the noise that you had added earlier!). To do this, you will have to use data-driven methods. These may be derived from POD based algorithms, ML algorithms, Deep Learning algorithms, etc. Use your ingenuity! Provide a comparative analysis of the true images and the denoised/super-resolved images. You might want to explore existing literature for such methods.

6 Reference Material

This is some basic reference material to get you started. You are expected to explore literature, books, and other videos for further reference.

- A playlist of videos related to relevant topics is as follows: Link - Playlist for SVD, PCA, POD
- Search for research papers on Google Scholar.