

MPhil DIS Report 24

CRSiD: tmb76

University of Cambridge

May 25, 2024

Contents

1	Executive Summary	2
2	Introduction	3
3	Background	4
4	Methodology	5
5	Conducted research	6
5.1	Portability of the code	6
5.2	Reproducibility of the results	6
5.3	Exploration of other algorithms	6
5.3.1	Spectral clustering	6
5.3.2	K-Means	6
5.3.3	Weighted K-Means	6
5.4	Stability Assessment	6
5.4.1	Under different number of clusters set	6
5.4.2	Under different training set size	6
6	Elasto-inertial turbulence	7
6.1	Background	7
6.2	Methodology	7
6.3	Results	7
6.4	Discussion	7
7	Data analysis pipeline	8

Chapter 1

Executive Summary

In many areas of physics, the equations describing the observed are complex and include a large number of terms. For general use and numerical simulations, it is often necessary to simplify these equations. This report investigates the work carried out in the Callaham paper, which explores the use of unsupervised learning to identify dominant balance regimes from simulated data of the terms in the equation.

Chapter 2

Introduction

For many problems in engineering of physical sciences, equations involve a large number of terms and are complex differential equations. Simulating them is therefore computationally expensive. However, it is often the case that these systems have multiple asymptotic local behaviours where the system is dominated only by a few terms. In these cases, one can simplify the equations to a balance between these dominant terms. This method is known as dominant balance or scale analysis, and has been a powerful tool in physics. Though it is extremely useful, it also requires a lot of expertise and has been done by hand in the past. In this report, a novel approach, developed by Callaham et al. (2021) [1] is discussed, verified and used on new data. The first part of the report will be a discussion of the background of the paper and of the method used. Then, the research and results of the Callaham et al. (2021) [1] paper will be shortly investigated, evaluating the ease of reproducibility of their work. Third, focussing on one of the examples used in the paper, use of other algorithms than the one used in the paper will be carried out, to test the robustness of the method. Finally, the method will be used on a new dataset, from simulation of elasto-inertial turbulence, a property of polymer laden flow.

Chapter 3

Background

Motivation for the work. Why is it important to be able to identify the dominant balance regimes in equations? What are the challenges but also what's the same in every dominant balance regardless of physics type?

Past literature if any. Has there been models with similar objectives but less flexibility (like only for one of the examples)

Chapter 4

Methodology

Get simulated data of the terms in the equation of physical variables from which terms in the equation can be derived

- Group the data into feature space, with each term as a feature

- Cluster the data using GMM

- SPCA to identify which terms are active in each cluster

- Group together clusters that have the same active terms

Chapter 5

Conducted research

Callaham has used the algorithm in a few different cases to validate how it functions.

5.1 Portability of the code

5.2 Reproducibility of the results

5.3 Exploration of other algorithms

5.3.1 Spectral clustering

5.3.2 K-Means

5.3.3 Weighted K-Means

5.4 Stability Assessment

5.4.1 Under different number of clusters set

5.4.2 Under different training set size

Chapter 6

Elasto-inertial turbulence

6.1 Background

6.2 Methodology

6.3 Results

6.4 Discussion

Chapter 7

Data analysis pipeline

Bibliography

- [1] J.L. Callaham, J.V. Koch, and B.W. et al. Brunton. Learning dominant physical processes with data-driven balance models. *Nature Communications*, 12:1016, 2021.