Data-driven discovery of physical laws

Nicolas Boullé[†] Alex Townsend

[†]Mathematical Institute, University of Oxford,
Woodstock Road, Oxford OX2 6GG

UNIVERSITY OF OXFORD

InFoMM Industrially Focused Mathematical Modelling

1. Introduction

Can one discover physical laws from experiments? From sparse and noisy laboratory data measurements, one hopes to eventually **learn physical laws of nature** or conservation laws that elude scientists in biology, medicine, and physics.

Challenge: The existing artificial intelligence algorithms are complex and it is difficult to extract **physical understanding** from these black-box machine learning techniques.

Aim: Understand when and why it is possible to discover physical laws and design a **human-understandable** technique for learning mathematical models from data.

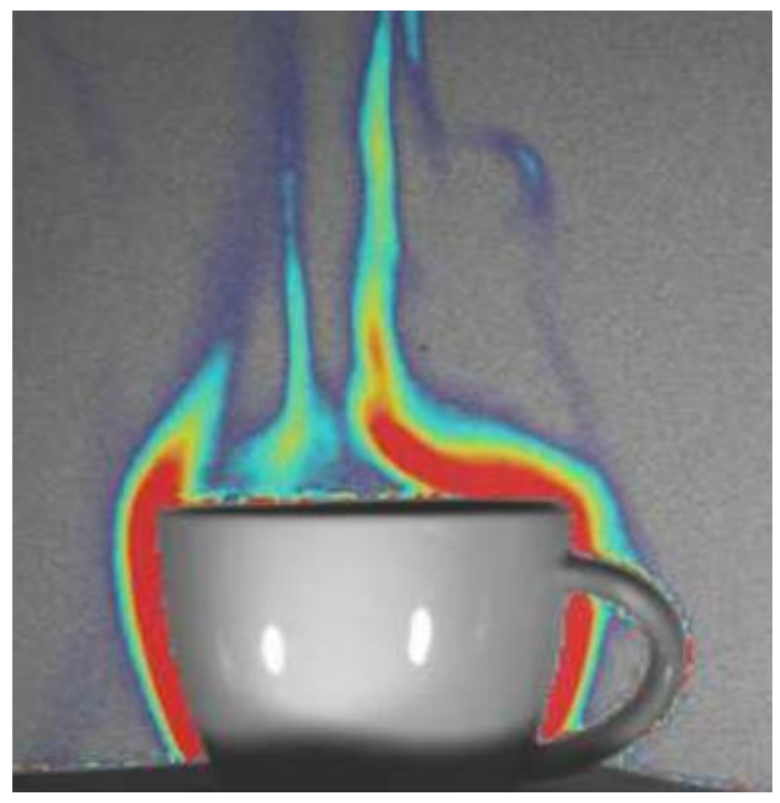

Fig. 1: Inferring flow over a coffee cup [1].

2. Theoretical Advances

Novelty: We derive the **first theoretical** scheme to learn a class of mathematical models from data. The theory characterizes the amount of training data needed.

Methodology: The proof is based on **random** excitations of the system and combines techniques of different areas of mathematics: numerical analysis, probability, analysis.

Outcome: Theoretical insights to design novel and practical algorithms for learning physical laws from data with **performance guarantees**.

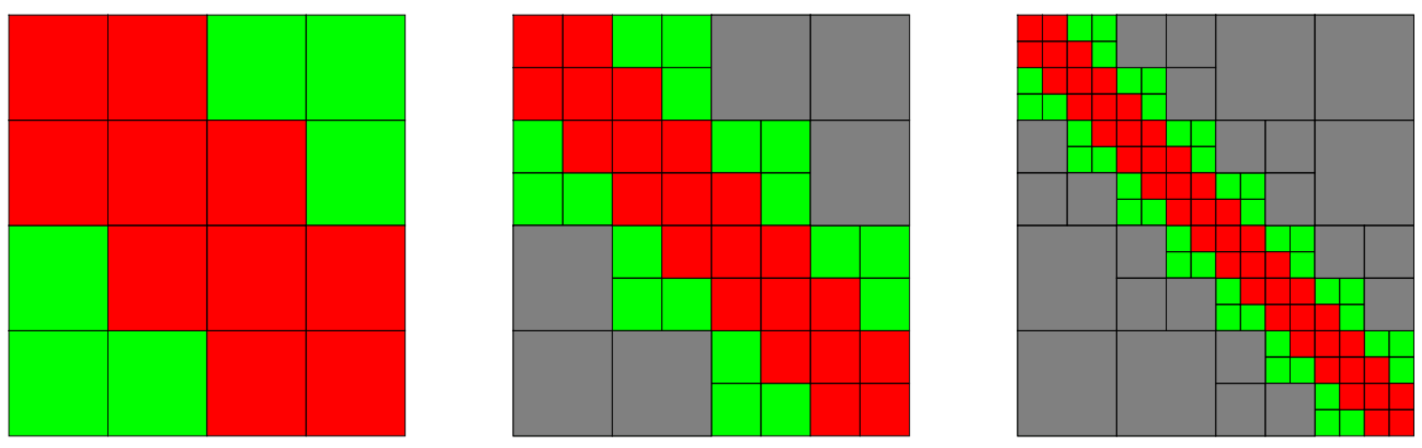

Fig. 2: Our proof exploits the geometrical structure of the problem [2].

3. Human-Understandable Technique

Novel approach: A data-driven technique that infuses an **interpretation** in the model by learning well-understood mathematical objects that imply underlying physical laws.

Ideas: Random excitations of the unknown system, a theoretically powerful machine learning algorithm [3], and the numerical analysis of the learned mathematical model.

Human-understandable **features** of the system can be discovered: conservation laws, symmetries…

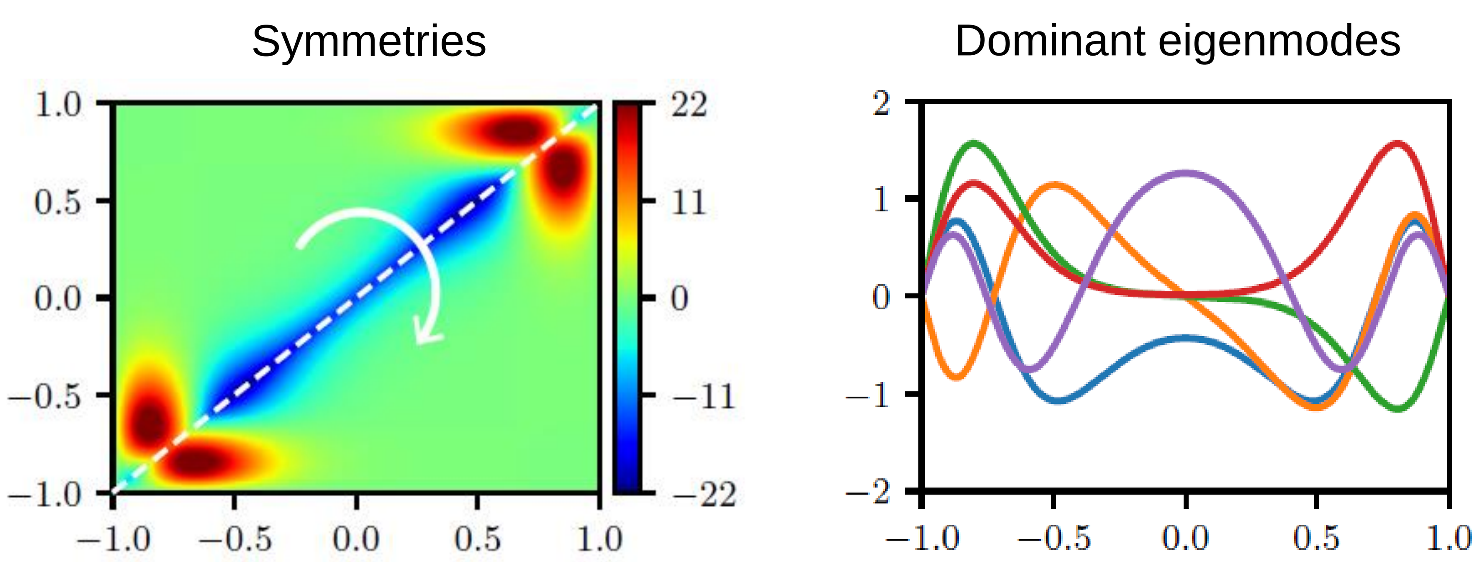

Fig. 3: Our algorithm enables the extraction of physical features [4].

4. Conclusions and Impact

Conclusions:
- A theoretical "**learning rate**", characterizing the amount of data needed to learn a mathematical model from an unknown physical system.
- A rigorous data-driven method exploiting prior mathematical knowledge to output a **human-understandable model** allowing detection and interpretation of qualitative features

Broader impact: This work aims to enable the application of data-driven methods in fields where **risk assessment** and **understanding** of the model are crucial to make decisions such as medicine, climate science, and finance.

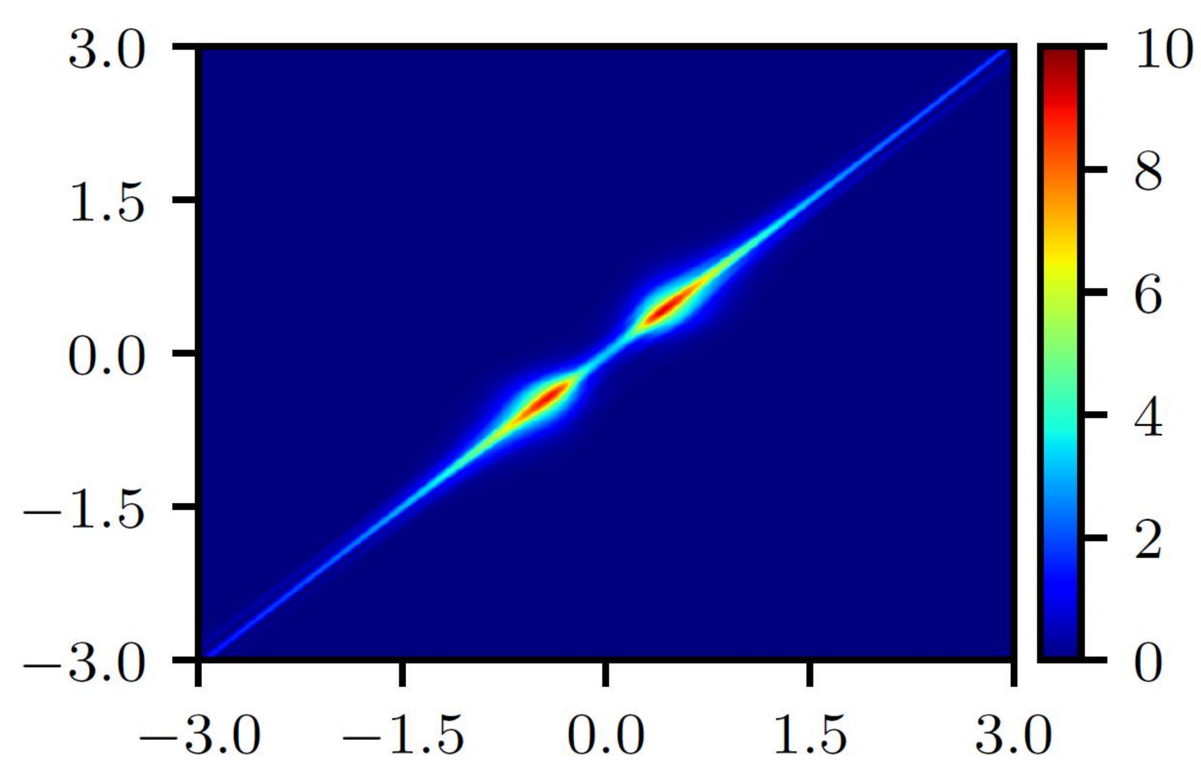

Fig. 4: Model discovered from a quantum mechanics system [4].

References: [1] Cai et al. 2021, [2] Boullé and Townsend 2022, [3] Boullé et al. 2020, [4] Boullé et al. 2021.

Acknowledgements: This poster is based on work supported by the EPSRC Centre For Doctoral Training in Industrially Focused Mathematical Modelling (EP/L015803/1) in collaboration with Simula Research Laboratory.