Project 02: Revealing Hidden Order

Leonel and Isaac

Introduction

Active matter systems, such as bacterial colonies or rod-like particle fluids, constitute fundamental models in non-equilibrium physics. These systems are characterized by complex local interactions that generate rich spatiotemporal dynamics, collective phenomena, and phase transitions (like the isotropic-to-nematic transition). Simulating these systems yields high-dimensional data (scalar, vector, and tensor fields on $D \ll 256^2$), which complicates the direct identification of the dominant modes of the dynamics and the understanding of transition mechanisms controlled by parameters like alignment and dipole strength.

Our Solution: This project proposes applying Unsupervised Deep Learning techniques (Autoencoders and Variational Autoencoders - VAEs) to *The Well Dataset* to decompose the observed dynamics into low-dimensional latent structures. This approach seeks to filter high-frequency noise and capture the *essence* of the collective dynamics, providing a compact and physically interpretable representation of the system's states.

Integration into Joint Paper 1: The main contribution of this project is the generation of an optimal, reduced set of state variables (the latent space z). This set is critical for feeding Project 01 (Kinetic Theory Closure), enabling the construction of predictive models that are significantly more efficient, stable, and computationally viable than those based on the full phase space.

Objectives

- Implement and optimize Autoencoder (AE) and Variational Autoencoder (VAE) architectures to compress the simulation data (256 x 256 grid) into a latent representation of dimension $D \ll 256^2$.
- Validate phase separation. Demonstrate that the resulting latent space significantly clusters simulations corresponding to distinct physical phases (e.g., isotropic, nematic) and parameter regimes (alignment and dipole-strength sweeps).
- Interpret the axes of the latent space. Establish an explicit correlation between key latent variables and macroscopic physical observables of interest (e.g., alignment order, kinetic energy, vorticity) to explain the learned features.

Preprocessing & Sampling

Using The Well Dataset, which provides 24 repetitions of 81-frame simulations of active rod-like particles on a 256×256 grid across alignment and dipole-strength sweeps, this project focuses on learning low-dimensional latent representations that capture dynamical regimes and transitions. The workflow is structured into three main components:

1. **Preprocessing and Sampling**
   - Decide on data representation (individual frames, short sequences, or local patches).
   - Normalize or standardize spatial fields (e.g., velocity, polarization).
   - Create a prototype subset (≈500–1000 frames) to prototype model architectures and visualizations.
2. **Exploratory Visualization**
   - Apply PCA, t-SNE, or UMAP on flattened frames or extracted descriptors (alignment order, vorticity, kinetic energy).
   - Inspect the degree of phase separation and trajectory continuity in reduced spaces.
   - Identify potential clustering structures indicative of hidden regimes or transitions.
3. **Latent Model Design**
   - Design Autoencoder (AE) and Variational Autoencoder (VAE) architectures (encoder–decoder structure, latent dimension, loss functions).
   - Define evaluation metrics:
     - ✓ Clustering separation between regimes.
     - ✓ Reconstruction error to assess fidelity.
     - ✓ Latent trajectory smoothness across control parameters.
   - Explore correlations between latent variables and physical observables to interpret learned features.

Hypothesis

Central Hypothesis: A Variational Autoencoder (VAE) can learn a very low-dimensional latent representation that not only minimizes reconstruction error but also intrinsically organizes the active matter dynamics from *The Well Dataset* along interpretable physical axes. Specifically, we hypothesize that collective phase transitions (e.g., isotropic to nematic) will manifest as smooth, continuous transitions in the latent space, while distinct control regimes (alignment and dipole strength) will form separate, well-defined clusters.

5. Integration into Joint Manuscript (Paper 1)

The integration of Project 02 is vital for the validity and scope of the *Joint Paper 1*:

- **Reduced State Variables:** The primary output of Project 02 (the latent space z) provides the optimal reduced set of state variables that Project 01 requires for building an accurate kinetic theory closure. This translates the high dimensionality of the simulation into a manageable basis for modeling.

- **Structural Support and Decoding:** Latent space visualizations demonstrating separation by phase and regime will serve as the empirical evidence illustrating how Machine Learning can "decode" the collective dynamics, thereby demonstrating a structural understanding of the transitions.
- **Physical Interpretability of the Model:** By correlating latent dimensions with physical observables (O3), we ensure that the closure model developed in Project 01 is based on physically relevant variables, increasing the scientific impact and interpretability of the joint work.