Analysis of latent space in Deep generative models and its applications to polymer dynamics

Scientific context

Studying soft matter systems still represents a challenge due to the complexity they present, in terms of the structure of the considered macromolecules, such as polymers and proteins, but also of the large number of objects in interaction. In view of current computing resources, the characteristic timescales for the study of the desired macroscopic properties remain inaccessible by the analytical and numerical resolution of the dynamics of such systems.

A classic solution to get around this problem is to reduce the number of degrees of freedom of the system by the c*oarse-graining* (CG) approach. Its goal is to decouple the *fast* degrees of freedom, which, once integrated, take the form of noise-dissipation forces, from the *slow* degrees of freedom, which are the only ones kept for the study of the dynamics. Defining the right sets of degrees of freedom and constructing an optimal representation for the CG phase space is an active ongoing research field [1].

A promising way to optimize this *coarse-graining* procedure is based on the application of new concepts developed in artificial intelligence [2]. The use of so-called generative models should indeed make it possible to reproduce the dynamics of a physical system on the characteristic times of interest but also to identify the relevant variables via the latent space. Analysis of the latter and its inversion in the original space can thus lead to a better understanding of the physical properties of the considered systems.

PhD project

First, the candidate will focus on the realization of a deep generative machine versatile and powerful enough to deal with the systems of interest such as solvated or unsolvated polymers, proteins (insulin), or highly heterogeneous systems, whether either by their composition or their geometry. A first approach will be to adapt recent methods developed for the generation of protein configurations at thermodynamic equilibrium [3].

The learning will require the production of molecular trajectories produced through *classical* simulations (eg LAMMPS), ideally in connection with other research work underway at SIMATLAB.

The second half of the thesis will be devoted to the numerical and theoretical analysis of the latent space. This investigation will require mathematical and statistical developments to exhibit the link with the real phase space of the physical systems studied.

Environment

This PhD will take place within SIMATLAB, a laboratory dedicated to research in the simulation of materials at the molecular scale. The SIMATLAB relies on the cooperation between:

- Different laboratories of the Université Clermont-Auvergne (UCA):
- The Institut de Chimie de Clermont-Ferrand is a general chemistry laboratory (UMR CNRS). The Thermodynamics and Molecular Interactions team works on the thermodynamics of fluid mixtures

and on the multi-scale study of interfaces, in particular with polymers. In particular, the team develops methodologies for multi-scale molecular simulation and modeling.

- LIMOS is a CNRS UMR mainly attached to INS2I and whose scientific positioning is centered around computer science, system modeling and optimization. One of the research axes of this laboratory concerns the processing of data, from the sensor to the analysis. In this context, the laboratory is developing activities in artificial intelligence, from a methodological and application point of view.
- The Laboratoire de Mathématiques Blaise Pascal (UMR CNRS 6620, attached to INSMI) is the main public research center in mathematics in the Auvergne region. Its fields of competence cover a wide spectrum, articulated around 4 teams, including the one of Probability, Analysis and Statistics which will take part in the supervision of this thesis.

- And of the MICHELIN company:
  - The molecular modeling team, whose objective is to enable numerical simulation to support industry in the design of innovative materials.

Supervisors

The supervision of this PhD project will be shared by:

- Alain DEQUIDT (ICCF)
- Manon MICHEL (LMBP)
- Vincent BARRA (LIMOS)
- Nicolas MARTZEL (MICHELIN)

The PhD candidate will work at SIMATLAB (ICCF), close to the other laboratories on the UCA campus in Cézeaux. The funding by UCA follows an academic doctoral contract.

PhD candidate profile

The ideal candidate comes from a M2 in mathematics or theoretical physics, with a good background in computational physics and/or machine-learning methods.

Bibliography

[1] *Grain Shape Dynamics for Molecular Simulations at the Mesoscale*, Nicolas Martzel, Alain Dequidt, Julien Devémy, Ronald Blaak, Sébastien Garruchet, et al. *Advanced Theory and Simulations*, Wiley, 2020, pp.2000124

[2] *Apprentissage artificiel - 4e édition : De Bayes et Hume au Deep Learning. Concepts et algorithmes*, Barra, V. and Cornuejols, A. and Miclet, L., 2021, Eds Eyrolles

[3] *Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning*, Frank Noé, Simon Olsson, Jonas Köhler, Hao Wu, Science 06 Sep 2019: Vol. 365, Issue 6457