Due date: Friday, December 16, 2022, 16:00
Submission contain a pdf with answers, and the notebook used to generate them

The biological functionality of proteins is tightly linked with their 3D structure. The stability of such configurations (native states) heavily depends on the temperature. For instance, egg white (albumin protein) transitions from transparent to white under heating because the 3D structure of the albumin proteins is denaturated and not soluble anymore. The simplest model to study such phenomena is the Lennard-Jones (LJ) polymer chain, i.e. a chain of identical beads linked together with harmonic springs which, in addition, interact with each other through a LJ potential. The interaction potential V of the LJ polymer chain is then given by

$$V(\{\mathbf{r}_i\}_{i=1,2,\dots,N}) = \sum_i k\left[r_{ii+1} - r_0\right]^2 + \sum_i \sum_{i<j} \epsilon \left[\left(\frac{r_0}{r_{ij}}\right)^{12} - \left(\frac{r_0}{r_{ij}}\right)^{6} \right] \quad , \tag{1}$$

where $\{\mathbf{r}_i\}_{i=1,2,\dots,N}$ is the set of particle positions, $r_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|$ the distance between two particles, r_0 the equilibrium distance, and $k = 16.67\epsilon/r_0^2$ the force constance, defined such that the stiffness of the harmonic term and of the LJ term match.

In this assignment, you will analyse molecular dynamics trajectories of a polymer chain of length $N = 100$ performed at various normalized temperatures: `T=[0.5, 1.0, 1.5, 1.6, 1.7, 1.8, 2.0, 2.1, 2.2, 2.3, 2.4, 2.6, 2.7, 3.5, 4.0]`; using a handcrafted feature, namely the radius of gyration, and learnt features from TICA/PCA.

1 Analysis of Lennard-Jones polymer trajectories

The trajectories are provided with the assignment and they are stored as a dictionary of numpy arrays containing the positions of the system over its evolution at a given temperature (the keys are the various temperatures).

1.1 Individual trajectories with TICA (12 points)

1.1.1 Radius of gyration

The squared radius of gyration of a polymer chain of N particles is given by

$$Rg^2 = \frac{1}{(N+1)^2} \sum_i^N \sum_{j<i} \|\mathbf{r}_i - \mathbf{r}_j\|^2. \tag{2}$$

Implement the functions `compute_distances` and `compute_gyration_radius` following the guidelines from the comments in the notebook.

1.1.2 Trajectory analysis

We will focus the analysis on three trajectories at `T=[0.5, 2.3, 4.0]` for simplicity. Visualize these trajectories using the function `display_trajectory`. Comment on the states that you observe qualitatively. Which temperature corresponds to a 'globular' state, an 'elongated' state, and a mixture of the two? What is the effect of thermal fluctuations on the favored state?

Compute the radius of gyration (Rg) and the 1st TICA projection ($TIC1$) for each configuration of the trajectories. Use the set of all distances as base features of each configuration for the TICA analysis. Why use distances rather than the set of positions?

Use the provided function `display_trajectory_and_property` to compare Rg and $TIC1$ for each trajectory. Are these quantities correlated with your previous qualitative observation? How well do they distinguish the two observed states? Motivate your answers with appropriate figures.

1.2 Estimate the critical temperature with PCA (8 points)

The critical temperature can be estimated from the change in a certain collective variable as a function of temperature. For the LJ polymer, the radius of gyration is an intuitive choice to determine the transition from the 'globular' to the 'elongated state'. Dimensionality reduction, in this case PCA, provides a tool to determine a low-dimensional representation without a priori knowledge of a suitable collective variable.

Compute the statistical average of the radius of gyration associated with each trajectory at the 15 different temperatures. To produce a feature that similarly differentiates the two states with PCA using the set of all distances as features requires two steps. First, determine the PCA projection using the trajectories at `T=[0.5, 4.0]` only. Indeed, these two trajectories are clearly in just one of the two states. Then, use this PCA projection to transform all the trajectories retaining the 1st PCA component. Show the average radius of gyration and average 1st PCA component as a function of temperature and identify the transition temperature.
The statistical average of an observable A is given by:

$$\langle A \rangle = \frac{1}{N} \sum_{n=1}^{N} A(t_n), \tag{3}$$

assuming that the time series $\{t_n\}_{n=1,2,\ldots,N}$ properly samples the probability distribution associated with the system.