Department of Physics
Freie Universität Berlin

Problem set 4: Computational Molecular Physics

Published: 2 June 2016 *Please send the solution by email to kkarathanou@zedat.fu-berlin.de by 7 June 2016, 12pm*

1. Cluster Analysis (25 points)

The file datapoints.txt attached to this set contains atomic cartesian coordinates xyz.

Use the k-means algorithm as implemented in matlab, or for python see

`https://datasciencelab.wordpress.com/2013/12/12/clustering-with-k-means-in-python/`

(or choose your favorite programming language and implement it yourself to

$$min \sum_{i=1}^{k} \sum_{x_n} ||x_n - c_k||^2$$

where x_n are the data points and c_k are the cluster centres.).

a) Visualise the data points and run k-means on the provided dataset using the euclidean distance and number of clusters k=11. Plot the clusters and the cluster centroids using different colors.

2. Lag-times (25 points)

The attached txt files NN distances, CC distances, psiA dihedrals, phiL dihedrals contain trajectories of different internal coordinates from a simulation of Alanine-Leucine. The first column is the time step, the second is atom distances or dihedral angles respectively.

a) Create 2D histograms, one for NN distances and psiA dihedrals, and one for CC distances and phiL dihedrals.

b) Make a plot of phiA dihedral angles in y-axis versus the corresponding psiL dihedral angles in x-axis.

c) Choose 2000 consequetive data points from the two dihedral files (at same time steps) and run the kmeans algorithm with k=5 clusters using the euclidean distance.

d) Find the count matrix C and transition probability matrix for transitions between the different clusters.

e) Choose 3 different lag-times and compute the eigenvalues of each new transition matrix.

f) Make a plot of lag-times in x-axis versus the corresponding implied time scale in y-axis (using all five eigenvalues). Are the implied time scales converged? Discuss possible reasons.