Project 07

Cluster analysis

Keywords: Markov chain Monte Carlo, cluster algorithms
Deadline: Please, hand in your report by **Tuesday, 29 July 2014**

1 Introduction

For complex molecules, which visit many different conformations, the long-lived conformations are of particular interest. These long-lived conformations correspond to minima in the free energy surface and hence maxima in the equilibrium probability density. Cluster algorithm exploit this property by trying to identify groups of conformations in a trajectory which form a particularly dense and closed group in the conformational space. These clusters are then typically interpreted as the long-lived conformational states of the molecules. In this project, you will use a Markov chain Monte Carlo algorithm to sample two-dimensional potential energy surfaces. You will then test the ability of different cluster algorithms to identify the minima in these potential energy surfaces.

2 Outline of the project

1. Familiarize yourself with cluster analysis by reading Ref. 2 and 3.
2. Extend the Markov chain Monte Carlo (MCMC) sampling algorithm from exercise 7, such that it can sample a two-dimensional potential.
3. Use this MCMC algorithm to sample suitable two-dimensional potentials. Estimate the distance matrix.
4. Implement the Jarvis-Patrick algorithm and an algorithm of your choice from Ref. 2. Analyze the trajectories with these cluster algorithms. Do the results depend on the choice of the parameters of the cluster algorithm?
5. You will obtain from us a trajectory of a small peptide. Cluster this trajectory using two cluster algorithms which are implemented in gromacs. Visualize the clusters using VMD or pyMol.

3 Literature

1. Lecture notes and exercises.
2. J. Shao , S. W. Tanner ,N. Thompson ,T.E. Cheatham , III, "Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms", *J. Chem. Theory Comput.*, **3**, pp 2312Đ2334 (2007)
3. R. Jarvis, E. Patrick, "Clustering Using a Similarity Measure Based on Shared Near Neighbors", *IEEE Transactions on Computers*, **C-22**, p. 1025-1034 (1973)
4. Chapter 12 in: T. Schlick, "Molecular Modeling and Simulation - an Interdisciplinary Guide", 2nd edition, Springer 2010.
5. Chapters 3 and 4 in: D. Frenkel, B. Smit, "Understanding Molecular Simulation - From Algorithms to Applications", Academic Press, Elsevier (USA) 2002.

Project requirements

- Describe the results of the your project in a report of 5 to 12 pages (font size 11 pt).
- The report should also contain a short description of the theory and the methods. If you implement an algorithm, briefly describe the algorithm.
- Hand in your program along with the report.
- You will present the results of your project in a 20-minute presentation, after which we will have about 10 minutes time for discussion.
- Besides questions which are directly related to the results of the project, the discussion will also cover the theory of the project and the course content which is relevant to the project.
- If you run into problems during the project or have questions, please contact us
  - Bettina Keller, bettina.keller@fu-berlin.de
  - Francesca Vitalini, francesca.vitalini11@gmail.com
- During the lecture hours (Thu, 2.15 - 3.45 pm) one of us will be in the library to answer questions.