# Project 06

## Quasiharmonic analysis / principle component analysis

**Keywords**: Markov chain Monte Carlo, covariance matrix, principle component analysis.
**Deadline**: Please, hand in your report by **Tuesday, 29 July 2014**

## 1 Introduction

Identifying the slowly moving degrees of freedom in a molecular system is one of the most difficult tasks in analyzing a molecular dynamics simulation. Principle component analysis (also referred to as quasi-harmonic analysis) is one approach to answer this question. "The correlation between atomic motions can be expressed in the covariance matrix $\mathbf{C}$ of the positional deviations:

$$C = \text{cov}(\mathbf{x}) = \langle (\mathbf{x} - \langle \mathbf{x} \rangle) (\mathbf{x} - \langle \mathbf{x} \rangle) \rangle \quad (1)$$

where ( ) denote an average over time" (Ref 4). $\mathbf{x}(t)$ is the trajectory of the positions of the systems. The eigenvectors of the covariance matrix represent a set of statistically independent generalized coordinates, which can be ordered according to the associated eigenvalues. The eigenvectors associated to the high lying eigenvalues are then interpreted as the slowly varying degrees of freedom of the system, i.e. the principle components. In this project, you will extend the Markov chain Monte Carlo sampling algorithm you implemented in exercise 7 to an MCMC algorithm which samples a two-dimensional potential energy surface and use it to illustrate the principle component.

## 2 Outline of the project

1. Familiarize yourself with the principle component analysis by reading Ref. 2-6. What is the difference to normal mode analysis?
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 covariance matrix and construct the principle components.
4. You will obtain from us a trajectory of a small peptide. Estimate the covariance matrix and construct the principle components gromacs. Visualize the principle components using VMD or pyMol.

## 3 Literature

1. Lecture notes and exercises.
2. http://www.cs.princeton.edu/picasso/mats/PCA-Tutorial-Intuition_jp.pdf
3. B. R. Brooks, D. Janezic, M. Karplus, "Harmonic Analysis of Large Systems. I. Methodology", *J. Comp. Chem.*, **16**, p. 1522-1542 (1995).
4. M.A. Balsera, W. Wriggers, Y. Oono, K. Schulten, "Principal Component Analysis and Long Time Protein Dynamics", *J. Phys. Chem.*, **100**, p 2567-2572 (1996).
5. A. Amadei, A.B. Linssen, H. J. Berendsen, "Essential dynamics of proteins", *Proteins*, **17**, p. 412-425 (1993).

6. Chapter 12 and 15 in: T. Schlick, “Molecular Modeling and Simulation - an Interdisciplinary Guide”, 2nd edition, Springer 2010.

7. 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.