A coarse graining method for the identification of transition rates between molecular conformations

Susanna Kube

Zuse Institute Berlin

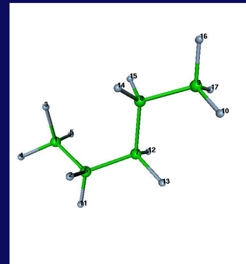
Free University Berlin

DFG Research Center
“Matheon”

Overview

molecular geometry, $q \in \mathbb{R}^{3n}$

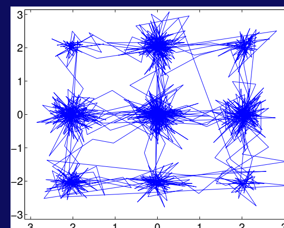
$\Downarrow$ *choose essential degrees of freedom*

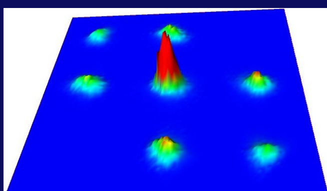


dihedral space $[-\pi, \pi]^d$

$\Downarrow$ *discretization*

ansatz functions $\xrightarrow{\textit{HMC/MD}}$ transition matrix





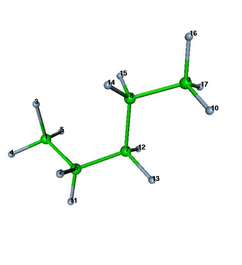
$\Downarrow$ *PCCA+*

conformations $\xrightarrow{\textit{aggregation}}$ conformational propagator

Overview

molecular geometry, $q \in \mathbb{R}^{3n}$

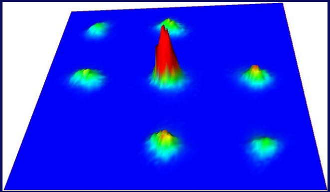
$\Downarrow$ *choose essential degrees of freedom*

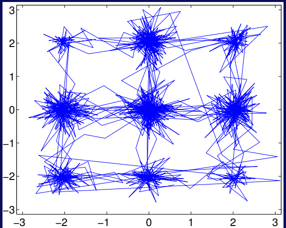


dihedral space $[-\pi, \pi]^d$

$\Downarrow$ *discretization*

ansatz functions $\stackrel{HMC/MD}{\Longrightarrow}$ transition matrix

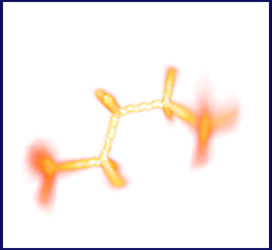




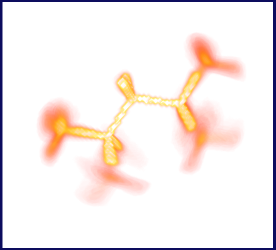
$\Downarrow$ *PCCA+*

conformations $\stackrel{aggregation}{\Longrightarrow}$ conformational propagator

Conformational Propagator

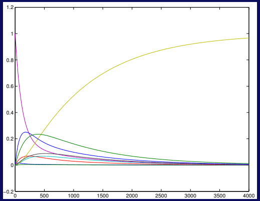


(g+/t)



(transition state)

(t/g+)



life times, transition proabilities/rates?

Transition Probabilities

Markov chains:

- discrete time sample path $\{X_n\}_{n\in\mathbb{N}}$ (uniform time steps)
- finite state space $E = \{1, \dots, N\}$
- stochastic transition probability matrix

$$P(i,j) = \frac{\#(i \to j)}{\sum_{k\in E} \#(i \to k)} \quad i,j \in E$$

Properties:

- irreducibility and positive recurrence $\leftrightarrow$

$$\exists!\ \pi : \quad \pi^\top P = \pi^\top, \sum_{i\in E} \pi(i) = 1$$

- reversibility

$$\pi_i P(i,j) = \pi_j P(j,i), \qquad \forall\, i,j \in E$$

Transition Frequencies

Relative transition frequencies

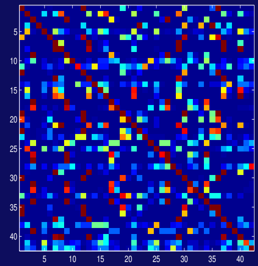
$$\overline{P} = DP \quad \text{(symmetric)}, \quad D = \text{diag}(\pi)$$

- $\overline{P}(i,j) = \frac{\#(i \to j)}{L-1}, \quad i,j \in E, \; L = \text{chain length}$
- $\sum_{i,j=1}^{N} \overline{P}(i,j) = 1, \qquad 0 \leq P(i,j) \leq 1 \quad \forall\, i,j \in E$
- $\overline{P}e = \pi$

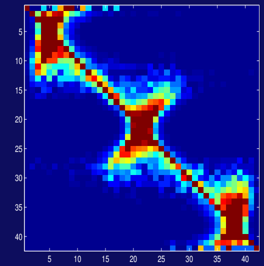
The transition frequencies can be computed directly from the simulation without knowing the stationary distribution!

Crisp Clustering

Reordering of states such that there are n_C blocks $\mathcal{C} = \{C_1, \ldots, C_{n_C}\}$.



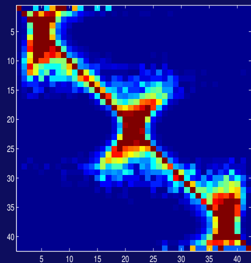
$\longrightarrow$



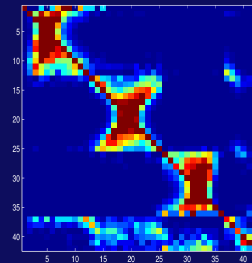
Grade of membership $\chi(i,j) \in [0,1]$:

$$\chi(i,j) = \begin{cases} 1 & \text{if state } i \text{ belongs to cluster } j \\ 0 & \text{else} \end{cases}$$

Soft Clustering



$\longleftrightarrow$



There are states which cannot be assigned uniquely to one of the clusters.

Grade of membership $\chi(i,j) \in (0,1)$: state i belongs to cluster j with probability $\chi(i,j)$

- $\sum_{j=1}^{n_C} \chi(i,j) = 1 \quad \forall\, i \in E$
- $0 \leq \chi(i,j) \leq 1 \quad \forall\, i \in E,\, j \in \mathcal{C}$

Cluster Algorithm

Robust Perron Cluster Analysis (PCCA+)
(Deuflhard, Weber, 2004)

$$\boxed{\chi = X\,A}$$

- X: eigenvectors of P

$$P\,X = \lambda\,X, \quad \lambda \approx 1$$

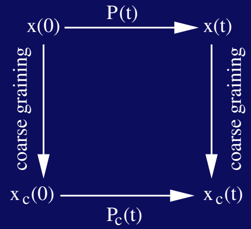
- A: non-singular transformation matrix

Propagation of Densities

- $x_i(t)$: probability to be in state i at time t

$$\sum_{i=1}^{N} x_i(t) = 1, \quad 0 \leq x_i(t) \leq 1$$

- propagation of densities: $\boxed{\mathbf{x}(t) = P^\top(t)\mathbf{x}(0)}$
- Chapman-Kolmogorov: $P(nt) = P^n(t)$
- goal: coarse graining – propagation of densities $\mathbf{x}(t) \in \mathbb{R}^{n_c}$, $n_c \ll N$

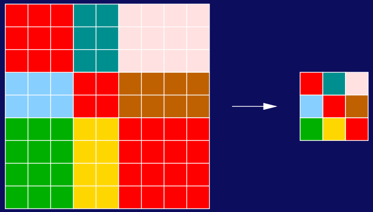


Crisp Coarse Graining

Clustering of states:

- frequencies:

$$\overline{P}_c(k,l) = \sum_{i \in C_k, j \in C_l} \overline{P}(i,j) = \chi(:,k)^\top \overline{P} \chi(:,l)$$



- probabilities:

$$P_c = \tilde{D}^{-1} \overline{P}_c = \tilde{D}^{-1} \chi^\top D P \chi$$

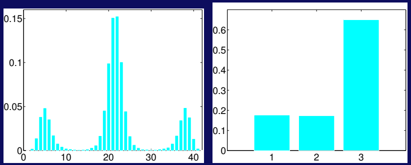
Coarse Graining Operators

$$P_c = \tilde{D}^{-1}\chi^\top D P \chi := I^\top P R^\top$$

Definition:

- restriction $R : \mathbb{R}^N \mapsto \mathbb{R}^{nC}$: $\quad R = \chi^\top$
- interpolation $I : \mathbb{R}^{nC} \mapsto \mathbb{R}^N$: $\quad I = D\chi\tilde{D}^{-1}$

Transformation of densities: $\quad x^c = Rx^f, \quad x^f = Ix^c$



Main goal – commutative diagram: $\quad RP^\top(t)x(0) = P_c^\top(t)Rx(0)$

: (Not satisfied for a hard clustering!

Soft Coarse Graining

Define

$$\boxed{\hat{P}_c := (RI)^{-\top} I^\top P R^\top}$$

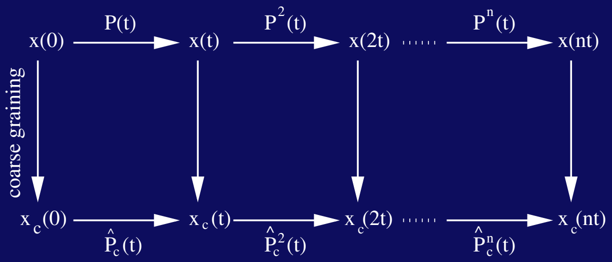
Properties:

- $\hat{P}_c$ is diagonalizable:
$$\hat{P}_c = A^{-1} \Theta A$$
- invariant density: $\hat{P}_c^\top \tilde{\pi} = \tilde{\pi}, \quad \tilde{\pi} = R\pi,\ \pi = I\tilde{\pi}$
- correct row sum: $\sum_{j=1}^{n_C} \hat{P}_c(i,j) = 1 \quad \forall\, i \in \mathcal{C}$
- positivity: $\sim \{0 \leq \hat{P}_c(i,j) \leq 1 \quad \forall\, i,j \in \mathcal{C}\}$
- $RP^\top(t)x(0) = \hat{P}_c^\top R x(0)$

  (change of propagation and restriction)

- extension to time steps nt, $n \in \mathbb{N}$: $\quad x(nt) = P^{n\top}(t)x(0)$

$$\boxed{RP^{n\top(t)}x(0) = \hat{P}_c^{n\top}(t)Rx(0)}$$



Idea:

- generalization of the coarse graining process to arbitrary times $t > 0$
- description of the dynamic behavior in terms of the infinitesimal generator Q

$$\dot{x}(t) = Q^\top x(t)$$

Coarse Grained Kinetics

$$Q = \lim_{t \mapsto 0+} \frac{P(t) - id}{t}$$

Master equation:

$$\dot{x}(t) = Q^\top x(t) \quad \rightarrow \quad x(t) = \exp(tQ^\top) x(0)$$

Define

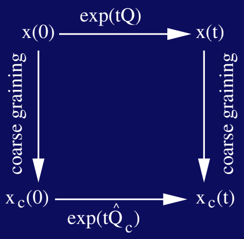
$$\boxed{\hat{Q}_c := (RI)^{-\top} I^\top Q R^\top}$$

Properties:

- $\hat{Q}_c$ is diagonalizable:

$$\hat{Q}_c = A^{-1} \Lambda A$$

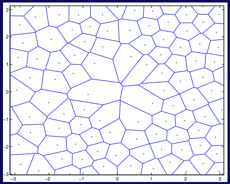
- invariant density: $\hat{Q}_c^\top \tilde{\pi} = 0, \quad \tilde{\pi} = R\pi, \, \pi = I\tilde{\pi}$
- conservation of mass: $\sum_{j=1}^{n_C} \hat{Q}_c(i,j) = 0 \quad \forall i \in \mathcal{C}$
- positivity: $\sim \{0 \le \hat{Q}_c(i,j) \quad \forall i,j \in \mathcal{C}, \, i \neq j\}$

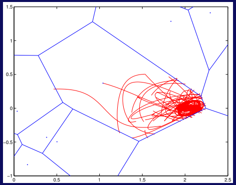


commutative!

Example

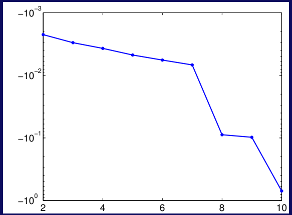
1. Discretization
2. Molecular Simulation
    - Merck Molecular Force Field
    - generation of 3000 points per cell according to the Boltzmann distribution by hybrid Monte-Carlo sampling with umbrella strategies and Gelman-Rubin convergence indicator
    - propagation by MD until the trajectories leave their starting cell

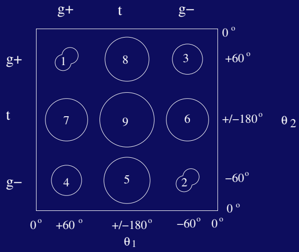


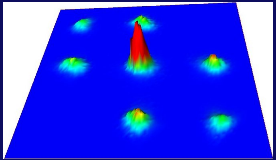


Example

3. Computation of Q
4. Cluster Analysis







$$\text{weights} = \{0.0036, 0.0032, 0.0640, 0.0680, 0.1248, 0.1232, 0.1140, 0.1567, 0.3426\}$$

$$h^{-1} = \{10.39, 11.38, 361.68, 334.35, 198.91, 227.31, 172.40, 235.74, 283.64\} ps$$

References

P. Deuflhard, Ch. Schütte: *Molecular Conformation Dynamics and Computational Drug Design.* In J. M. Hill and R. Moore, editors, *Applied Mathematics Entering the 21th Century.* ICIAM 2003, Sydney, Australia, 2004.

P. Deuflhard, M. Weber: *Robust Perron Cluster Analysis in Conformation Dynamics.* In M. Dellnitz, S. Kirkland, M. Neumann, and Ch. Schütte, editors, *Lin. Alg. Appl. - Special Issues on Matrices and Mathematical Biology*, Vol. 398C, pp 161-184. Elsevier Journals, 2005.

M. Weber: *Meshless Methods in Conformation Dynamics.* PhD thesis, FU Berlin, Februar 2006. Verlag Dr. Hut, München, 2006.

S. Kube, M. Weber: *Coarse Grained Molecular Kinetics.* ZIB-Report 06-35.

Project Group

ZIB Scientific Computing:

Dept. Numerical Analysis and Modelling:
Peter Deuflhard, Marcus Weber, Susanna Kube

Dept. Scientific Visualization:
Hans-Christian Hege, Daniel Baum, Johannes Schmidt-Ehrenberg

Cooperation:

FU Biocomputing:
Christof Schütte, Wilhelm Huisinga, Evelyn Dittmer

DFG Research Center “Matheon”