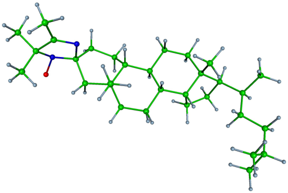
Modeling Principles in Conformation Dynamics

Susanna Kube



YMM, Bonn, July 2007

Zuse Institute Berlin

DFG Research Center

MATHEON

ZIB Scientific Computing:

Dept. Numerical Analysis and Modelling:

Peter Deuflhard, Marcus Weber, Susanna Kube, Isabel Reinecke

Cooperations:

FU Berlin:

- Biocomputing: Christof Schütte
- Applied Mathematics: Caroline Lasser
- Computational Molecular Biology: Frank Noé

DFG Research Center **“Matheon”**

Molecular Dynamics

Modelling of molecules in classical MD:

$$H(q,p) = \frac{1}{2} p^\top M^{-1} p + V(q)$$

$$V = V_{\text{bond}} + V_{\text{angle}} + V_{\text{torsion}} + V_{\text{Coulomb}} + V_{\text{VdW}}$$

$q \in \mathbb{R}^{3s}$: positions of all atoms, $p \in \mathbb{R}^{3s}$: momenta

Corresponding equations of motion:

$$\dot{q} = M^{-1} p, \qquad \dot{p} = -\nabla V$$

Formal solution: $(q(t+\tau), p(t+\tau)) = \Phi^\tau(q(t), p(t), \tau)$

Ensemble

Canonical ensemble:

constant number of particles n

constant volume v

constant temperature T

Invariant density (Boltzmann):

$$\mu(q,p) = \underbrace{\frac{1}{Z_p}\exp\left(-\frac{\beta}{2}p^\top M^{-1}p\right)}_{=\eta(p)}\underbrace{\frac{1}{Z_q}\exp(-\beta V(q))}_{=\pi(q)}, \quad \beta = 1/k_B\mathcal{T}$$

Hamiltonian dynamics with randomized momenta:

$$q_{n+1} = \Pi_q \Phi^\tau(q_n, p_n), \quad p_n \sim \eta(p)$$

Time scales

MD simulation:

1 ns = 10^{-9}s

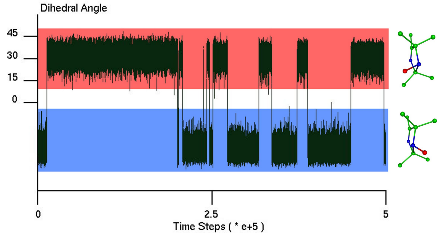
protein folding:

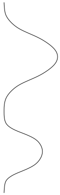
1 ms = 10^{-3}s

Comparison: 1 s $\leftrightarrow$ 278 h (11.5 days)

multiscale approach: conformation dynamics

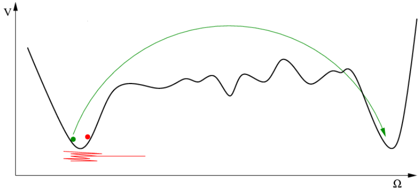
Metastable dynamical systems



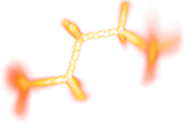


stationary density?

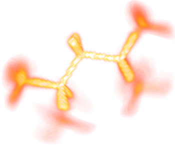
The trajectory gets trapped in valleys of the potential energy surface.



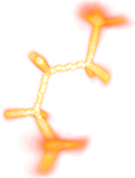
Example: Pentane



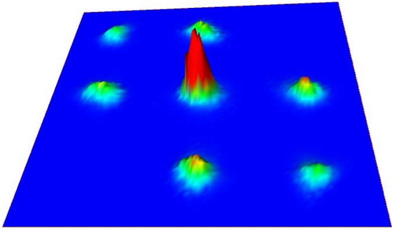
(g+/t)

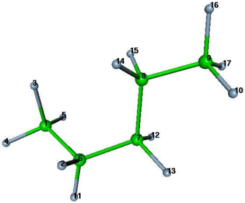


(transition state)



(t/g+)





Conformations as almost invariant sets

Transfer operator:

$$T^\tau u(q) = \int_{\mathbb{R}^d} u(\Pi_q \Phi^{-\tau}(q,p))\eta(p)\, dp$$

[Schütte, Fischer, Huisinga, Deuflhard 1998]

- Probability to be within A
  $w(A) = \int_A \pi(q) dq = \langle \chi_A, \chi_A \rangle_\pi$
- Probability to move from A $\rightarrow$ B during time τ
  $w(A, B, \tau) = \langle T^\tau \chi_A, \chi_B \rangle_\pi / w(A)$
- Probability to stay within A during time τ
  $w(A, A, \tau) = \langle T^\tau \chi_A, \chi_A \rangle_\pi / w(A)$

Conformations: almost invariant subsets of the position space Ω

$$w(A, A, \tau) \approx 1 \quad \leftrightarrow \quad T^\tau \chi_A \approx \chi_A$$

Eigenvalue Problem

$$T^\tau u = \lambda u, \qquad \lambda \approx 1$$

Galerkin approach:

$$u(q) = \sum_{i=1}^{N} \alpha_i \varphi_i(q)$$

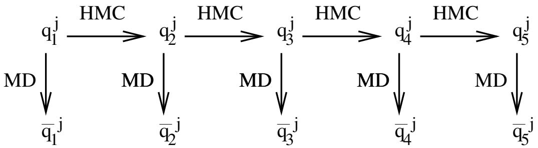
$$\sum_{i=1}^{N} \alpha_i \langle T^\tau \varphi_i, \varphi_j \rangle_\pi = \lambda \sum_{i=1}^{N} \alpha_i \langle \varphi_i, \varphi_j \rangle_\pi, \quad \forall j$$

$$\boxed{\overline{P}\alpha = \overline{S}\alpha\lambda \qquad \text{(symmetric)}}$$

$$\sum_{i=1}^{N} \alpha_i \frac{\langle T^\tau \varphi_i, \varphi_j \rangle_\pi}{\langle \varphi_j \rangle_\pi} = \lambda \sum_{i=1}^{N} \alpha_i \frac{\langle \varphi_i, \varphi_j \rangle_\pi}{\langle \varphi_j \rangle_\pi}, \quad \forall j$$

$$\boxed{P\alpha = S\alpha\lambda \qquad \text{(stochastic)}}$$

Approximation of matrix elements



Approximation of matrices by Monte Carlo importance sampling:

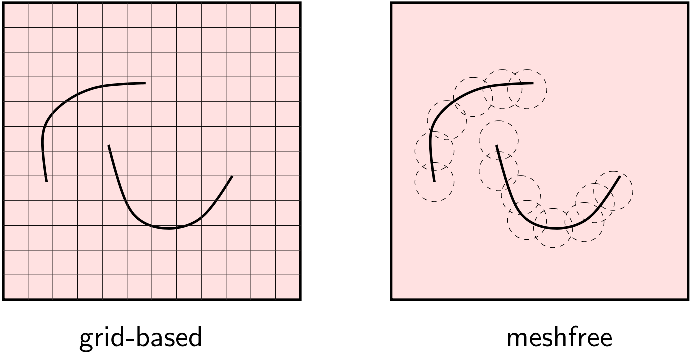
$$\overline{S}(j,i) = \langle \varphi_j, \varphi_i \rangle_\pi = \int_\Omega \varphi_i(q)\varphi_j(q)\pi(q)\,dq \approx \frac{1}{n_j}\sum_{k=1}^{n_j} \varphi_i(q_k^{(j)}),$$

$$\overline{P}(j,i) = \langle T^\tau \varphi_j, \varphi_i \rangle_\pi \int_\Omega T^\tau \varphi_i(q)\varphi_j(q)\pi(q)\,dq \approx \frac{1}{n_j}\sum_{k=1}^{n_j} \varphi_i(\overline{q}_k^{(j)})$$

Note:

- the points $q_k^{(j)}$ are distributed according to $\varphi_j(q)\pi(q)$

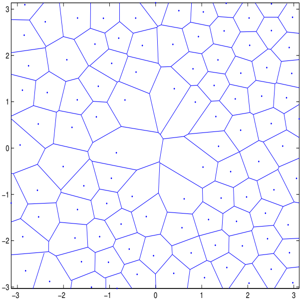
Curse of dimensionality



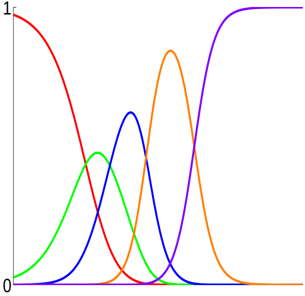
partition of unity

$$\sum_{i=1}^{N} \varphi_i(q) = 1, \quad \varphi_i(q) \geq 0 \qquad \forall q \in \Omega$$

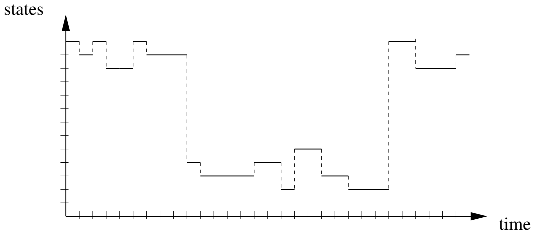
Voronoi discretization



radial basis functions



Markov chains

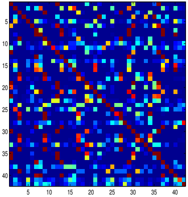


- discrete time sample path $\{X_n\}_{n\in\mathbb{N}}$ (uniform time steps)
- finite state space $E = \{1, \dots, N\}$
- stochastic transition probability matrix

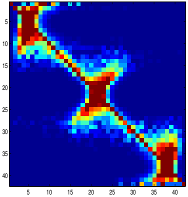
$$P(i,j) = \frac{\#(i \to j)}{\sum_{k\in E} \#(i \to k)} \quad i,j \in E$$

Crisp Clustering

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



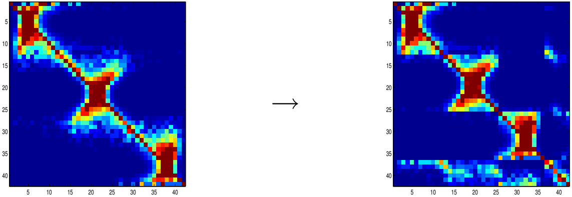
$\longrightarrow$



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



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 $\chi(i,j)$

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

Cluster Algorithm

Robust Perron Cluster Analysis (PCCA+)

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

- $X = [\mathbf{x}_1, \ldots, \mathbf{x}_{n_C}]$: eigenvectors of P

$$P\,\mathbf{x}_k = \lambda_k\,\mathbf{x}_k, \quad \lambda_k \approx 1$$

- $\mathcal{A}$: non-singular transformation matrix
- χ: membership vectors
  - partition of unity, positiv
  - $\forall\, k = \{1, \ldots, n_C\} \exists\, i \in \{1, \ldots, N\} : \chi_k(i) = 1$

[Deuflhard, Weber, Lin. Alg. Appl. 2004]

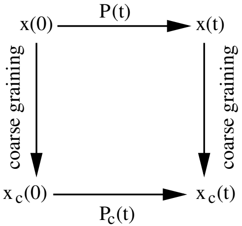
Propagation of densities

$$\boxed{\mathbf{x}(t) = P^\top(t)\mathbf{x}(0)}$$

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

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

- coarse graining: propagation of densities $\mathbf{x}_c(t) \in \mathbb{R}^{n_c}$, $n_c \ll N$



Coarse graining

Transition probabilities between conformations:

$$P_c = (\underbrace{D_c^{-2}\chi^\top D^2 S\chi}_{\mathcal{S}})^{-1}\,\underbrace{D_c^{-2}\chi^\top D^2 P\chi}_{\mathcal{P}} = (\chi^\top D^2 S\chi)^{-1}\chi^\top D^2 P\chi$$

[Kube, Weber; JCP 2007]

correct propagator, but small negative entries!

objective function in PCCA+:

$$\boxed{I(\mathcal{A}) = \mathrm{trace}(\mathcal{S}) \rightarrow \max}$$

Perturbation theory

$$\widetilde{P} = P + E, \qquad \widetilde{P}\widetilde{X} = \widetilde{X}\widetilde{\Lambda}$$

Schur decomposition:

$$[X_1, X_2]^H P [X_1, X_2] = \begin{pmatrix} L_1 & H \\ 0 & L_2 \end{pmatrix}, \quad [X_1, X_2]^H E [X_1, X_2] = \begin{pmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{pmatrix}$$

Perturbation bound:

$$\| \sin \Theta(\mathcal{X}_1, \widetilde{\mathcal{X}}_1) \| < C \| E_{21} \|$$

Error Estimation

- Row-wise correlated random matrices:

$$\mathbb{E}[E(i,j)] = 0 \quad \text{and} \quad \mathbb{E}[E(i,j)E(k,l)] = \delta_{ik} C_i(j,l)$$

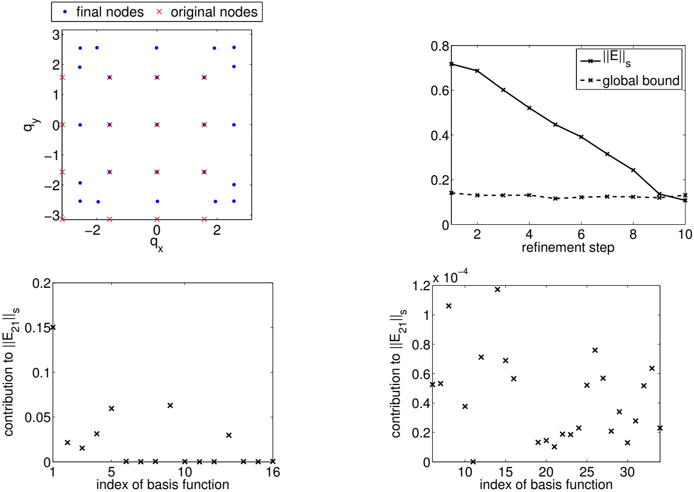
- stochastic norm ([G. W. Stewart, 1990]):

$$\|E\|_s^2 \equiv \mathbb{E}(\|E\|_F^2)$$

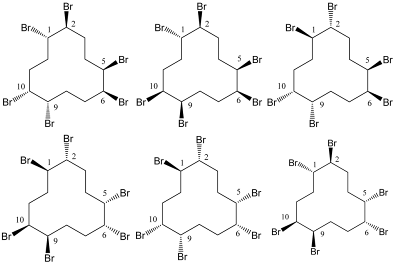
$$\|E_{21}\|_s^2 = \sum_{k=1}^{N} \|\overline{X}_2(k,:)\|_2^2 \text{trace}(\overline{X}_1^H C_k \overline{X}_1)$$

- probability distribution: $E(i,:) \sim \text{Dir}(\alpha_i)$
- parameter estimation from sampling data (Maximum likelihood estimator)

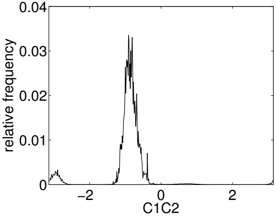
Equilibration of errors and sampling effort



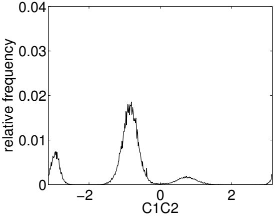
Example: Hexabromocyclododecane



probability for the angle between the bromine atoms to be in gauche or anti position



non-adaptive



adaptive

Finally

Thank you for your attention!

Further information

http://www.zib.de/Numerik/DrugDesign/index.en.html

Open positions:

- Matheon A4 ("Towards a mathematics of biomolecular flexibility"), PhD or Postdoc