Variational Approach to Markov Processes (VAMP)

Identification of molecular order parameters and states from nonreversible MD simulations

Fabian Paul

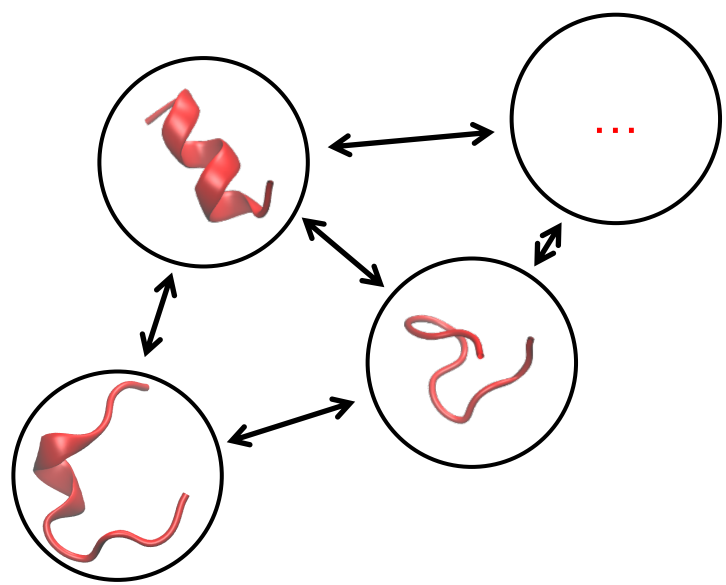
Computer Tutorial in Markov Modeling

18-FEB-2020

Recap: the spectral theory of MSMs

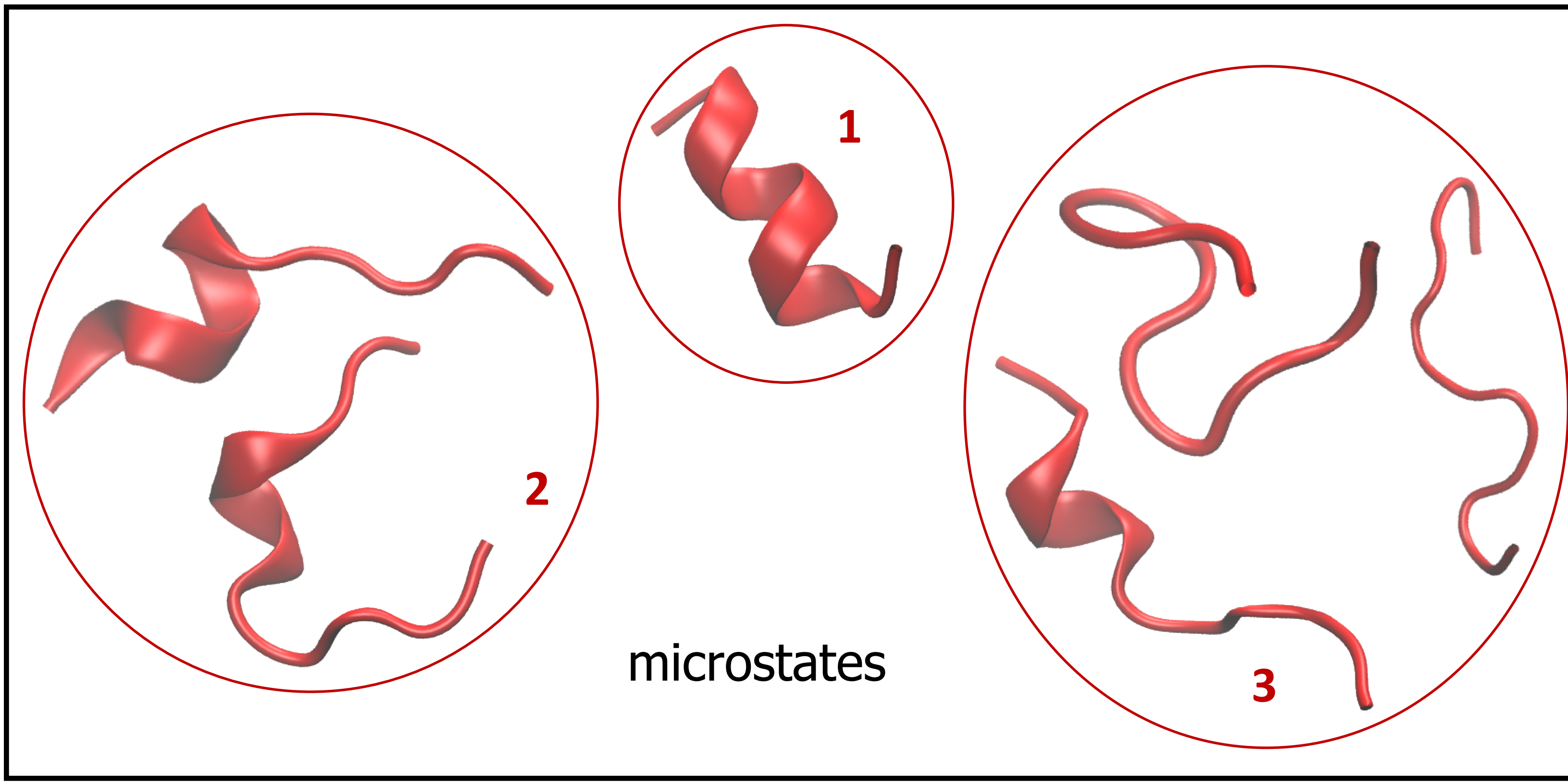
- A Markov state model consists of:
  1. a set of states $\{s_i\}_{i=1,\dots N}$
  2. (conditional) transition probabilities between these states

$$T_{ij} = \mathbb{P}(s(t+\tau) = j \mid s(t) = i)$$



Markov state models: estimation

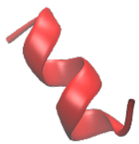
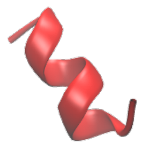
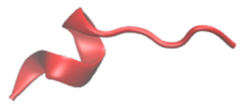
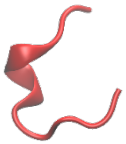
- Markov model estimation starts with:
  grouping of geometrically[1] or kinetically[2] related conformations into *clusters* or *microstates*



[1] Prinz *et al.*, *J. Chem. Phys.* **134**, 174105 (2011)
[2] Pérez-Hernández, **Paul**, *et al.*, *J. Chem. Phys.* **139**, 015102 (2013)

Markov state models: estimation

- We then assign every conformation in a MD trajectory to a microstate.

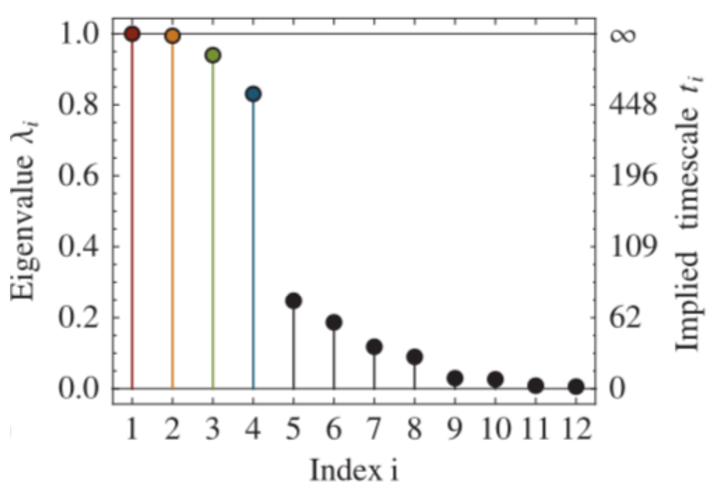
time t	τ	2τ	3τ	4τ	5τ	6τ	7τ
trajectory							
microstate s	1	1	2	3	3	2	3

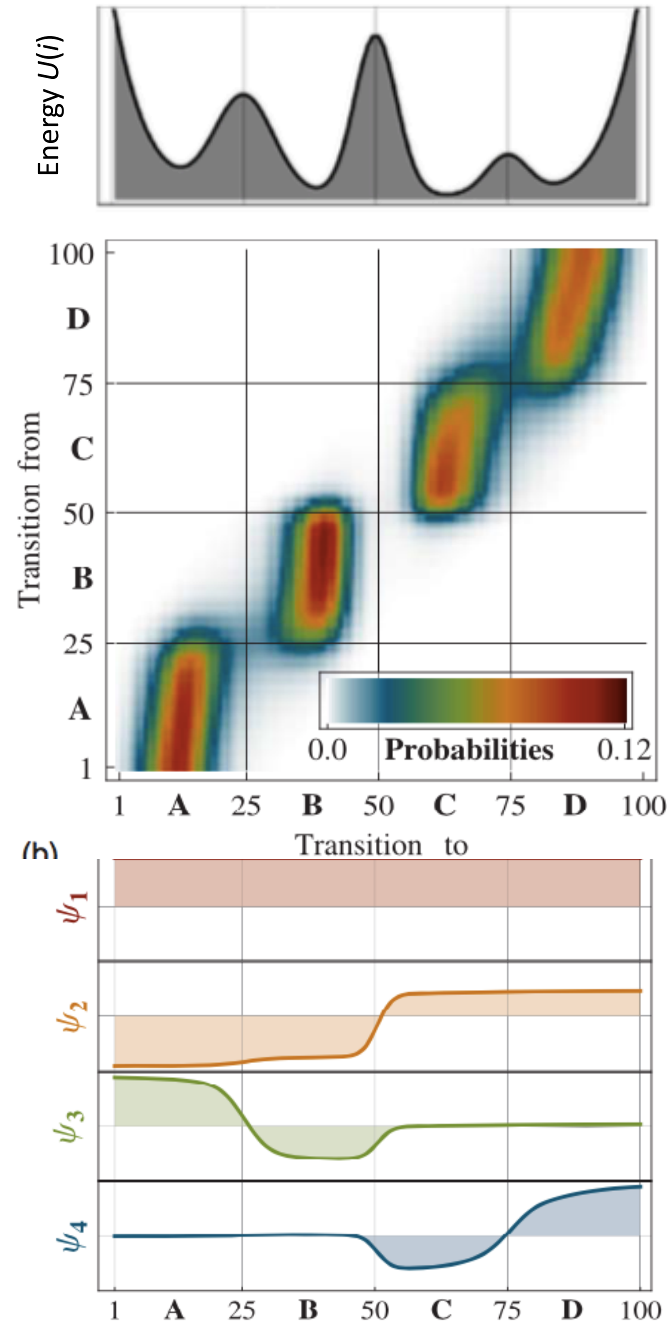
- We count transitions between microstates and tabulate them in a count matrix $\mathbf{C}$
  e. g. $C_{11} = 1$, $C_{12} = 1$, $C_{23} = 2$, …
- We estimate the transition probabilities T_{ij} from $\mathbf{C}$.
  - Naïve estimator: $\hat{T}_{ij} = C_{ij} / \sum_k C_{ik}$
  - Maximum-likelihood estimator [1]

[1] Prinz *et al.*, *J. Chem. Phys.* **134**, 174105 (2011)
[2] Pérez-Hernández, **Paul**, *et al.*, *J. Chem. Phys.* **139**, 015102 (2013)

The spectrum of a reversible T matrix

- The large eigenvalues of the transition matrix and their corresponding eigenvectors encode the information about the slow molecular processes.
- Flat regions of the eigenvectors allow to identify the metastable states.





Prinz *et al.*, *J. Chem. Phys.* **134**, 174105 (2011)

Both MSMs and TICA make use of the same spectral method

The spectral method (working with eigenvalue and eigenvector) is not limited to Markov state models.

- Estimation of MSMs

$$T(\tau) = \frac{C_{ij}(\tau)}{C_i}$$

- In matrix notation

$$\mathbf{T}(\tau) = \mathbf{C}(0)^{-1}\mathbf{C}(\tau)$$

- Eigenvalue problem:
  $\mathbf{T}(\tau)\mathbf{v} = \lambda\mathbf{v} \Leftrightarrow \mathbf{C}(0)^{-1}\mathbf{C}(\tau)\mathbf{v} = \lambda\mathbf{v} \Leftrightarrow \mathbf{C}(\tau)\mathbf{v} = \lambda\mathbf{C}(0)\mathbf{v}$
- The last equation is known as the TICA problem. All equations generalize to the case where $\mathbf{C}(0)$ and $\mathbf{C}(\tau)$ are not count matrices, but correlation matrices.
- The indices i, j don't longer refer to states but to *features*.

Schwantes, Pande, *J. Chem. Theory Comput.* **9** 2000 (2013)
Pérez-Hernández *et al* ,*J. Chem. Phys.*, **139** 015102 (2013)

VAC and VAMP

Variational approach to conformational dynamics VAC (Rayleigh-Ritz for classical dynamics)

Any autocorrelation is bounded by the system-specific number $\hat{\lambda}$, that is related to the system-specific autocorrelation time $\hat{t}$ by $\hat{\lambda} = e^{-\tau/\hat{t}}$.

$$\mathrm{acf}(\psi;\tau) := \frac{\sum_t^{T-\tau} \psi\big(x(t)\big)\psi(x(t+\tau))}{\sum_t^{T-\tau} \psi\big(x(t)\big)\psi(x(t))} = \frac{\langle \psi, \mathrm{T}\psi\rangle_\pi}{\langle \psi, \psi\rangle_\pi} \leq \hat{\lambda}$$

- The maximum is achieved if ψ is an eigenfunction of T.

Proof:

Expand ψ in an (orthonormal) eigen-basis of T:

$$\psi(x) = \sum_i c_i\, \phi_i(x), \qquad \langle \psi, \psi\rangle_\pi = \sum_i c_i^2 > 0$$

$$\langle \psi, \mathrm{T}\psi\rangle_\pi - \hat{\lambda}\langle \psi, \psi\rangle_\pi = \sum_i c_i^2\, \lambda_i - \sum_i c_i^2\, \hat{\lambda} = \sum_i c_i^2 \left(\lambda_i - \hat{\lambda}\right) \leq 0$$

- If $\hat{\lambda}$ is $\max_i \lambda_i$ the largest of T's eigenvalues, the inequality holds.
- Result can only be zero if $c_i = 0$ for $i \neq j$ and $\lambda_j = \max_i \lambda_i \Rightarrow \psi(x) \propto \phi_{\max}(x)$
- Remark: the variational approach generalizes to the optimization of multiple eigenfunctions. $\hat{\lambda}$ is replaced by the sum of the eigenvalues $R_k = \sum_{i=1}^{k} \lambda_i$

Noé, Nüske, *SIAM Multiscale Model. Simul.* **11**, 635 (2013)

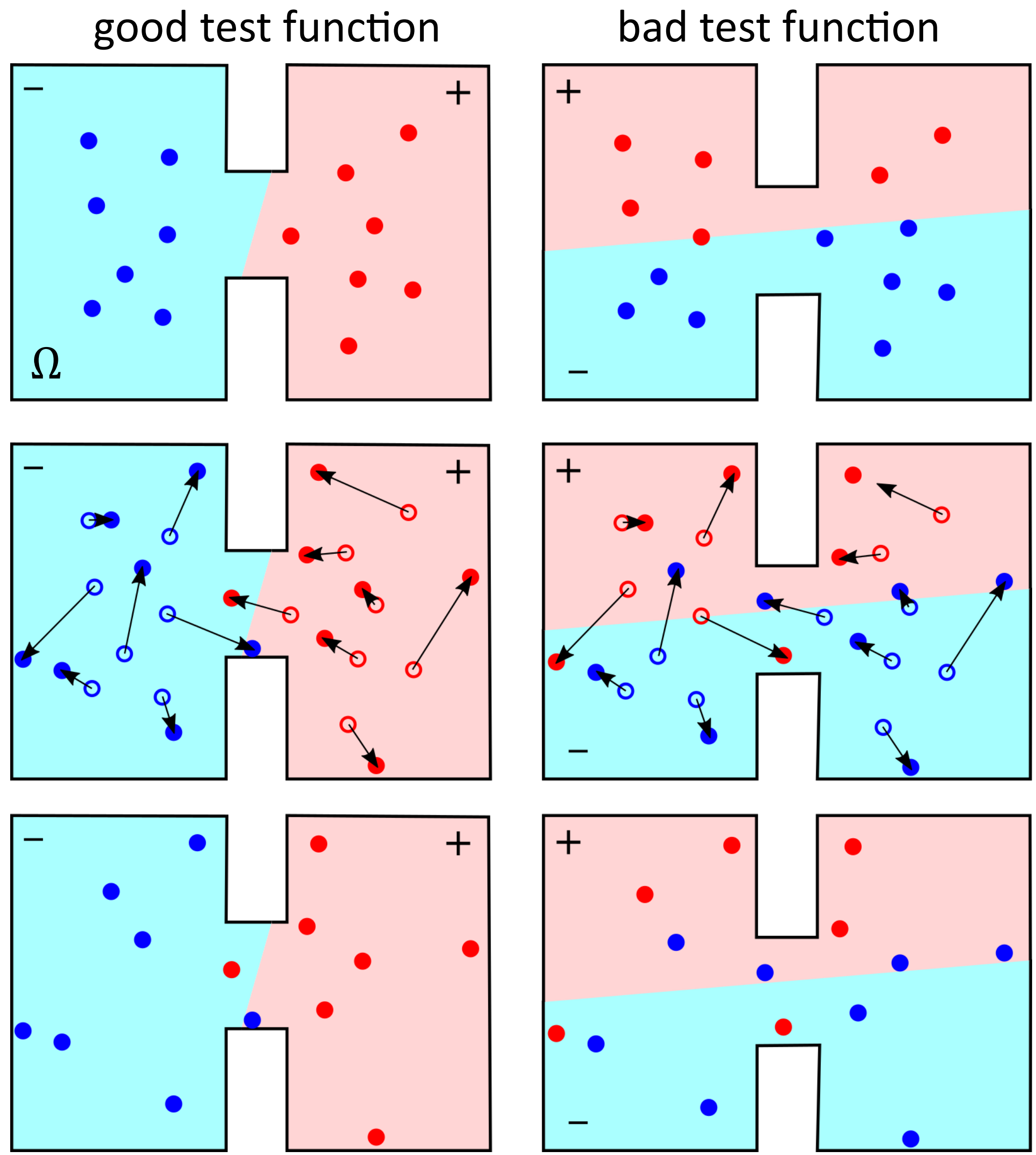
Interpretation of variational principle

1. Pick some test function $\chi_{\text{test}}(\mathbf{x})$ and pick some test conformations $\mathbf{x}_{i,\text{inital}}$ distributed according to equilibrium distribution π

2. Propagate $\mathbf{x}_{i,\text{inital}}$ with the the MD integrator. Call result $\mathbf{x}_{i,\text{final}}$.

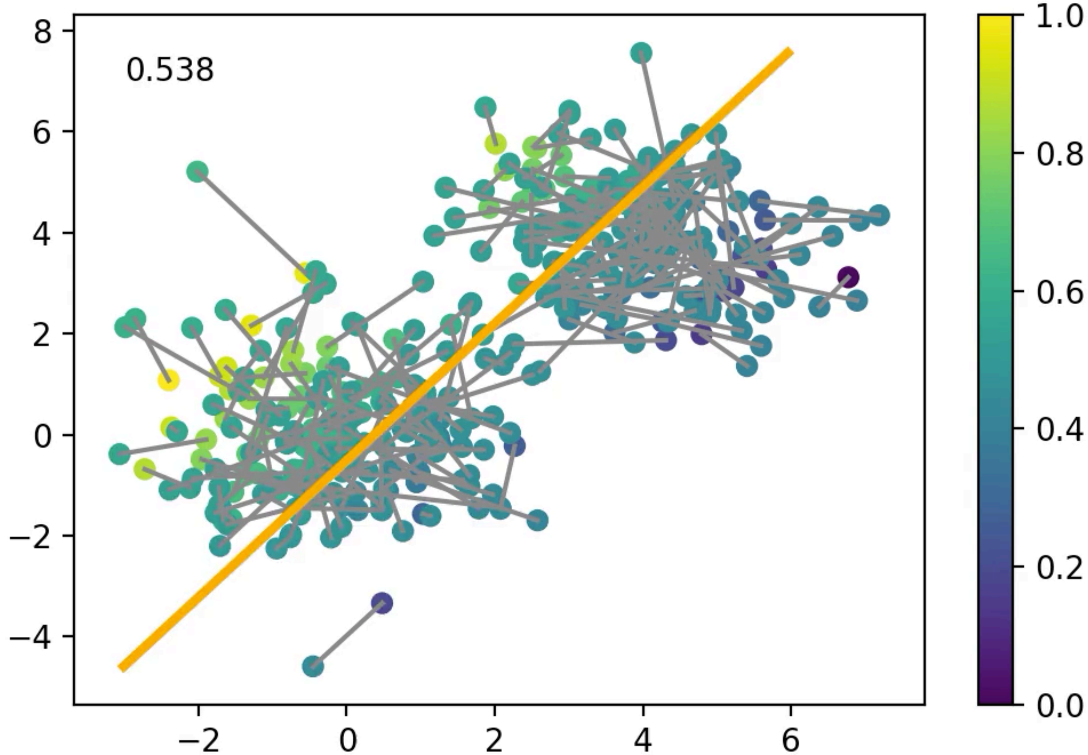
3. Correlate $\chi_{\text{test}}(\mathbf{x}_{\text{inital}})$ with $\chi_{\text{test}}(\mathbf{x}_{\text{final}})$.

$$\text{score}=\frac{\sum_{i=1}^{N}(\chi(\mathbf{x}_{i,\text{inital}})-\overline{\chi})\cdot(\chi(\mathbf{x}_{i,\text{final}})-\overline{\chi})}{\sum_{i=1}^{N}(\chi(\mathbf{x}_{i,\text{inital}})-\overline{\chi})\cdot(\chi(\mathbf{x}_{i,\text{inital}})-\overline{\chi})}$$

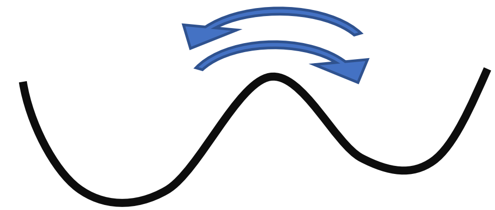


Gradient-based optimization of function parameters

Parameters $\mathbf{p}$ of $\chi_{\mathrm{test}}(\mathbf{x};\mathbf{p})$ can be optimized with gradient-based techniques. Make use of the gradient of the VAC or VAMP score, the gradient of the test function and off-the-shelf optimizers such as ADAM or BFGS.



Reversible dynamics



- **In equilibrium**, every trajectory is as probable as its time-reversed copy

$$\mathbb{P}(s(t+\tau)=j \text{ and } s(t)=i)=\mathbb{P}(s(t+\tau)=i \text{ and } s(t)=j)$$

$$\mathbb{P}(s(t+\tau)=j \mid s(t)=i)\mathbb{P}_{\mathrm{eq}}(s(t)=i)=\mathbb{P}(s(t+\tau)=i \mid s(t)=j)\mathbb{P}_{\mathrm{eq}}(s(t)=j)$$

$$\pi_i T_{ij}=\pi_j T_{ji}$$

- In mathematician's notation $\langle \mathbf{e}_i, \mathbf{T}\mathbf{e}_j \rangle_\pi = \langle \mathbf{e}_j, \mathbf{T}\mathbf{e}_i \rangle_\pi$ where $\langle \mathbf{x}, \mathbf{y} \rangle_\pi = \sum_i x_i y_i \pi_i$
- $\mathbf{T}$ is a symmetric matrix w.r.t. to a non-standard scalar product.
- $\mathbf{T}$ has real eigenvalues and eigenvectors (linear algebra I).

Prinz *et al.*, *J. Chem. Phys.* **134**, 174105 (2011)

The problem with nonreversible systems

- $R_k = \sum_{i=1}^{k} \lambda_i$ where λ_i are the true eigenvalues.
- For nonreversible dynamics $\langle \mathbf{e}_i, \mathbf{T}\mathbf{e}_j \rangle_\pi \neq \langle \mathbf{e}_j, \mathbf{T}\mathbf{e}_i \rangle_\pi$
- There might not even be a well-defined $\boldsymbol{\pi}$.
- Eigenvalues and eigenvectors will be complex.
- Variational principle doesn't work. $\mathrm{acf}(\psi) \leq \hat{\lambda} \in \mathbb{C}$ makes no sense. One can't order complex numbers on a line.
  - Optimization of models not possible
  - Feature selection not possible

- Is there any way to fix this? Can we maybe find some other operator that is related to dynamics and that is symmetric?

A possible solution: VAMP
Variational approach to Markov processes

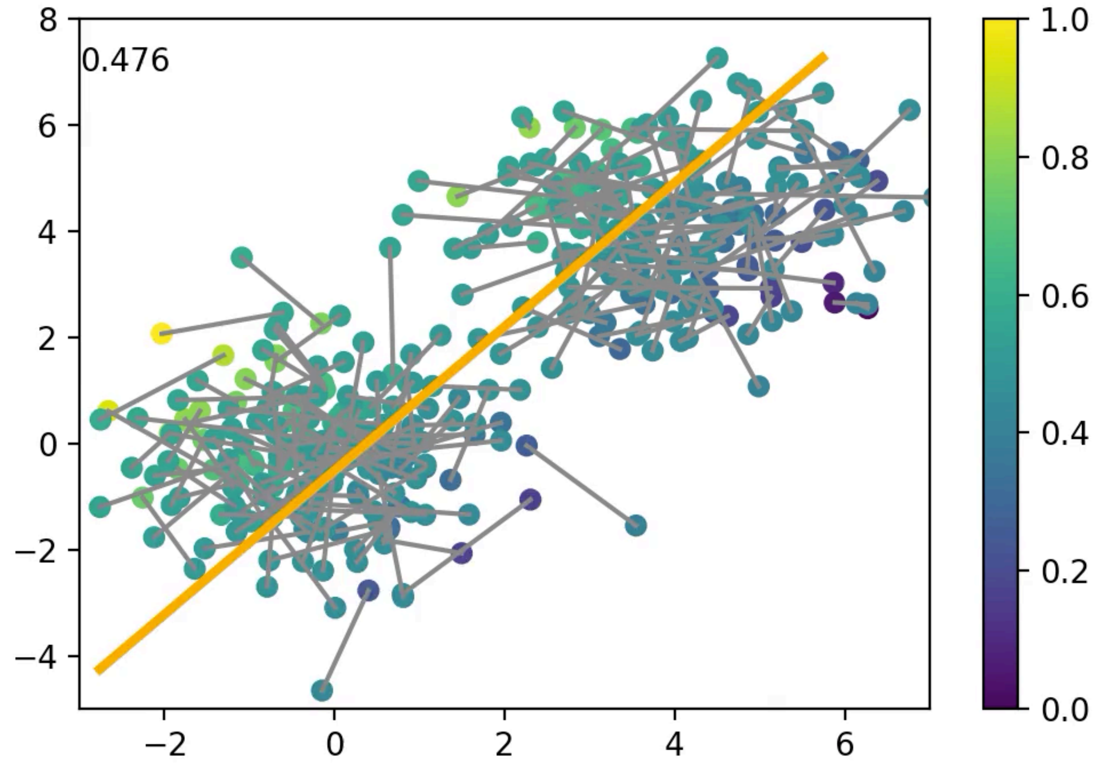
- Introduce the "backward" transition matrix
$$\mathbf{T}_{\mathrm{b}} := \mathbf{C}(N)^{-1}\mathbf{C}(-\tau) = \mathbf{C}(N)^{-1}\,\mathbf{C}^{\top}(\tau)$$
i.e. estimate MSM/TICA from time-reversed data, where
$$C_{ij}(-\tau) := \sum_{t=\tau}^{N} f_i\big(x(t-\tau)\big) f_j(x(t))$$
$$C_{ij}(N) := \sum_{t=\tau}^{N} f_i(x(t)) f_j(x(t))$$
- Introduce the forward-backward transition matrix $\mathbf{T}_{\mathrm{fb}} \coloneqq \mathbf{T}\mathbf{T}_{\mathrm{b}}$ and $\mathbf{T}_{\mathrm{bf}} := \mathbf{T}_{b}\mathbf{T}$
- Can show that $\mathbf{T}_{\mathrm{fb}}$ and $\mathbf{T}_{\mathrm{bf}}$ are symmetric *without any reference to a stationary vector* (symmetry is built into the matrices).
- Eigenvalues and eigenvectors of $\mathbf{T}_{fb}$ and $\mathbf{T}_{bf}$ are real.
- They fulfill a variational principle $\left\|\mathbf{C}^{-1/2}(0)\mathbf{C}(\tau)\mathbf{C}(\mathrm{N})^{-1/2}\right\| \leq R$

Wu, Noé, *J. Nonlinear Sci.* **30**, 23 (2020)   Klus, S. *et al, J. Nonlinear Sci.,* **28**, 1 (2018)

Cross-validation

- The model parameters (in this example parameters of the line and steepness of the transition) were optimized for a particular realization of the dynamics.

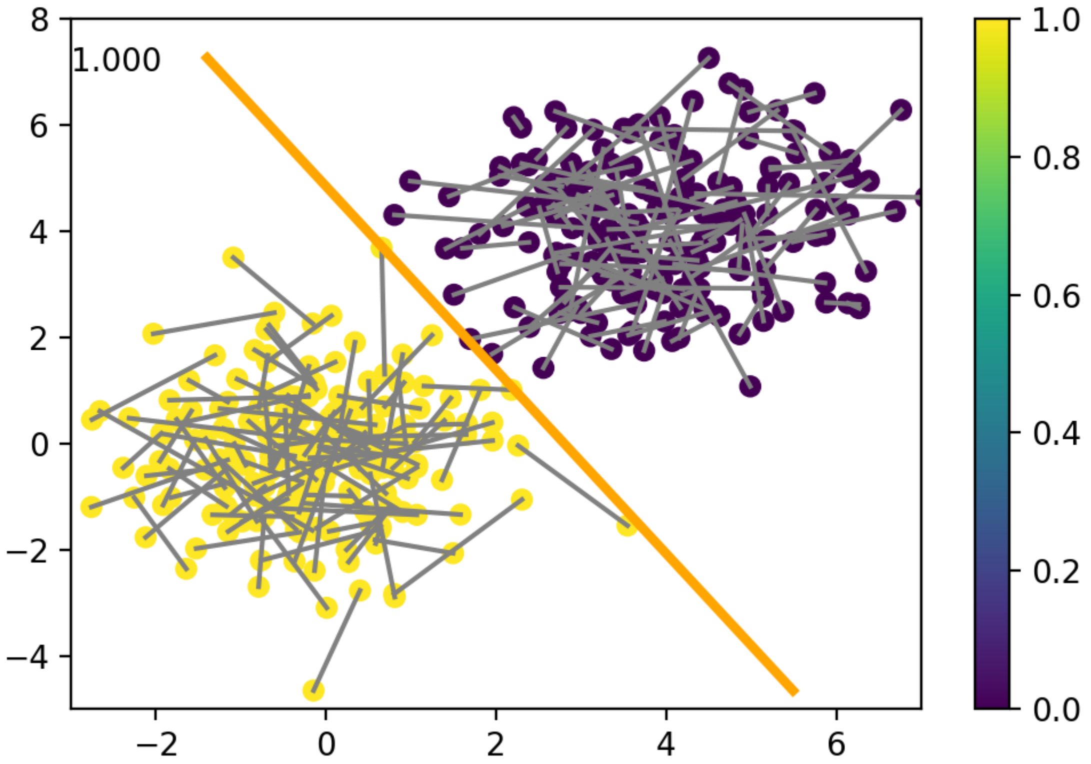
- Didn't we say that the eigenfunctions and eigenvalues were an intrinsic property of the molecular system?
- So the eigenfunctions should be the same if we repeat the analysis with a second simulation of the same system.



Cross-validation

- The model parameters (in this example parameters of the line and steepness of the transition) were optimized for a particular realization of the dynamics.

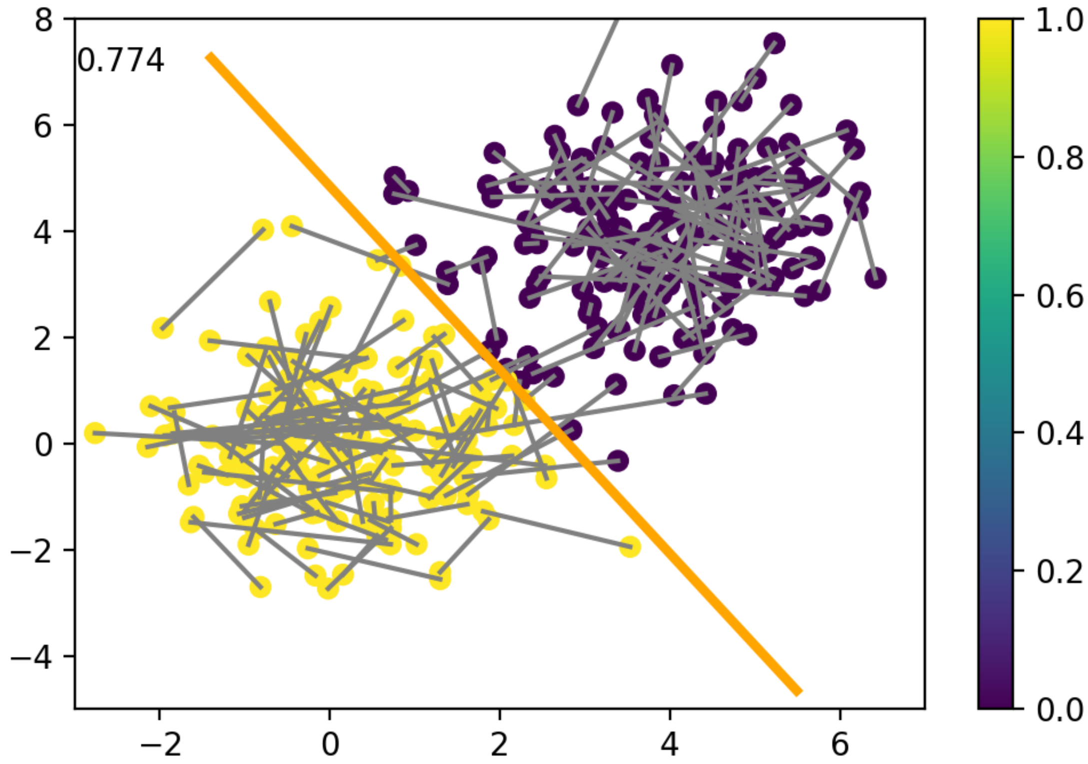
- Didn't we say that the eigenfunctions and eigenvalues were an intrinsic property of the molecular system?
- So the eigenfunctions should be the same if we repeat the analysis with a second simulation of the same system.



Cross-validation

- Ideally, we want to tell if the solution is robust at a single glance by measuring the robustness with one number.

- The VAMP score or VAC score (also called GRMQ[1]) lends itself to this task.
- Keep all the **trained** model parameters fixed (here the line parameters and the steepness of the transition), plug in new data and recompute the **test** autocorrelation.
- The **test** autocorrelation will be lower in general, which means that the original model was fit to noise (**overfit**).



[1] McGibbon, Pande, *J Chem Phys.,* **142** 124105 (2015)

Cross-validation

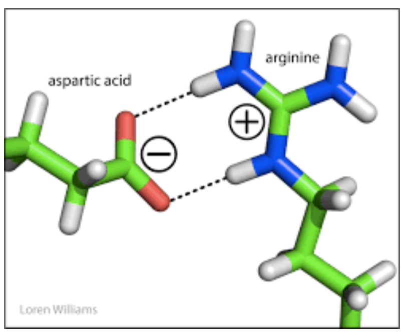
- Reporting a test-score that was computed from independent realizations is the gold standard.
- Independent realizations can be expensive to sample.
- Do the approximate *k*-fold (hold-out) cross-validation.
  - Split all data into **training** set and **test** sets.
  - Optimize the model parameters with the training set and test the parameters with test sets.
  - Repeat for *k* different divisions of the data.

- *k*-fold cross-validation can be tricky with highly autocorrelated time series data!

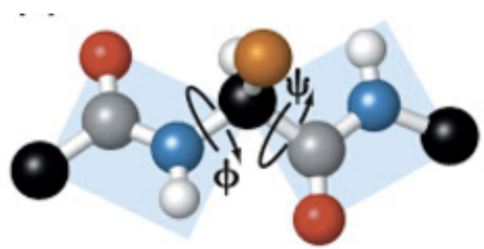
Applications

Application: feature selection

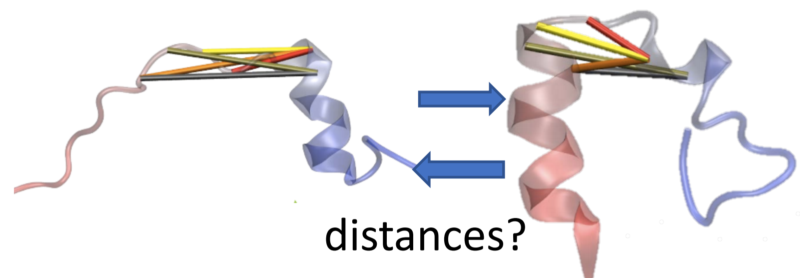
- variational principle: the higher the score the better
- Compare **test** scores for different selections of molecular features. Which selection gives best score?



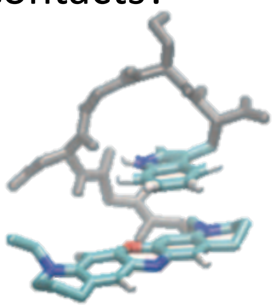
contacts?



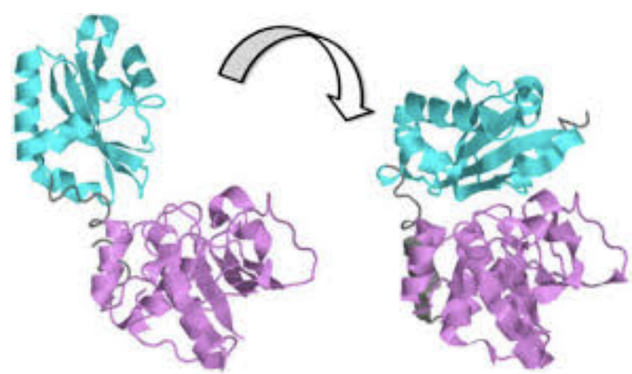
dihedrals?



distances?



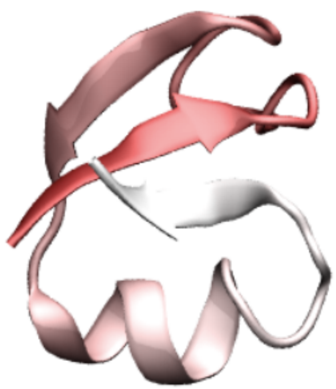
chemical intuition?



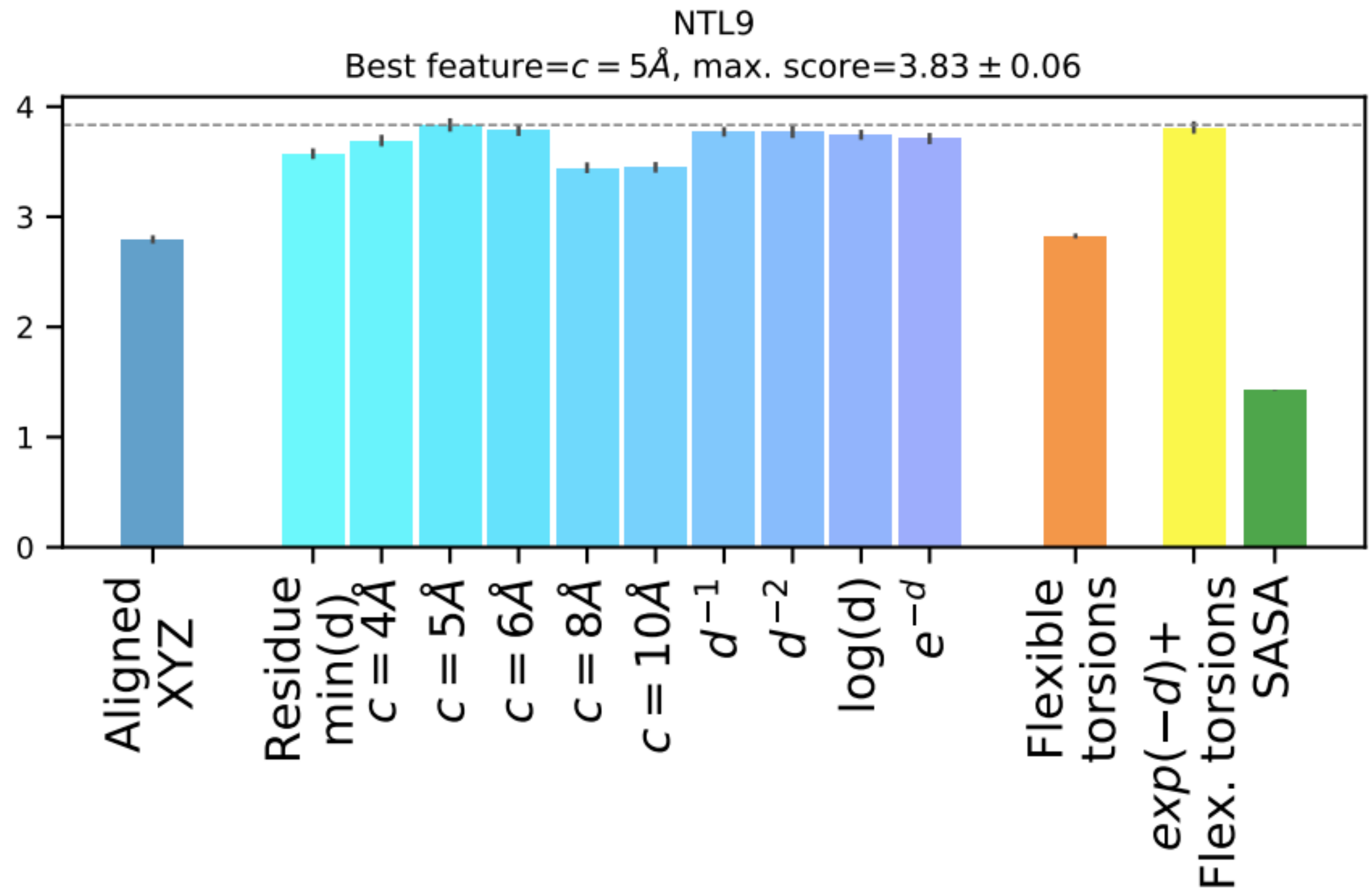
rigid body approximation?

side chain flips?

Application: feature selection

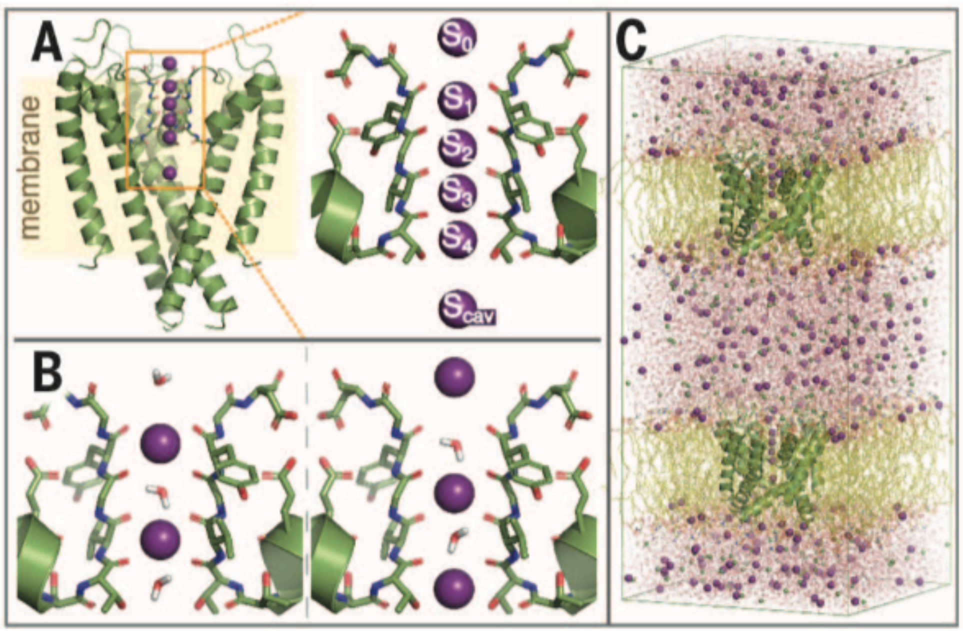


NTL9

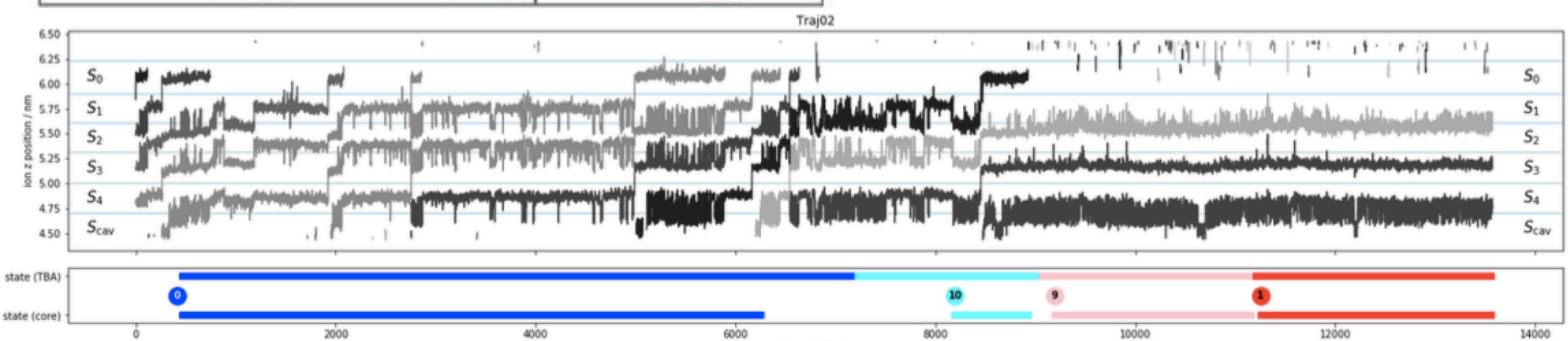


Scherer *et al J. Chem. Phys.* **150**, 194108 (2019)

Application: ion channel non-equilibrium MD



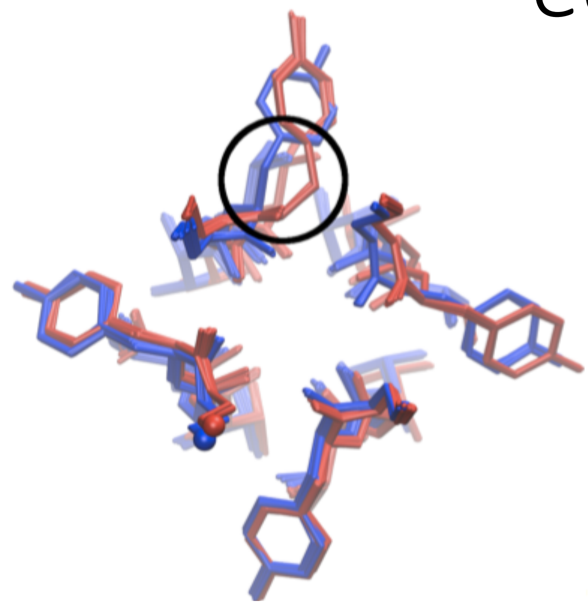
Analysis of MD simulation data of the "controversial" direct-knock-on conduction mechanism in the KcsA potassium channel.
Ions a constantly inserted at one side of the membrane and deleted at the other side.

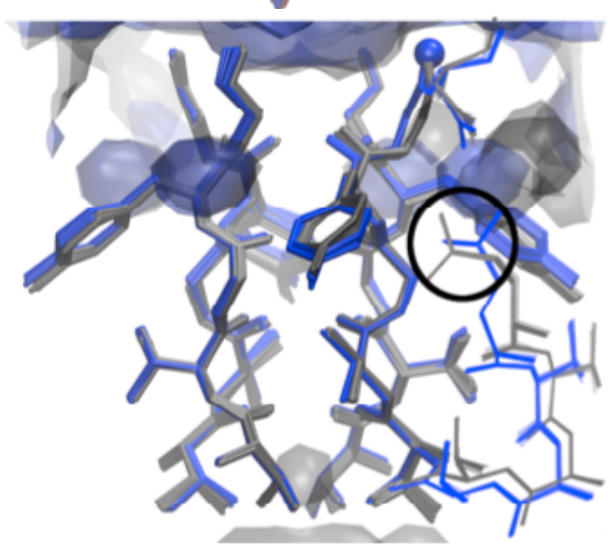


Paul *et al*, *J. Chem. Phys.* **MMMK**, 164120 (2019).

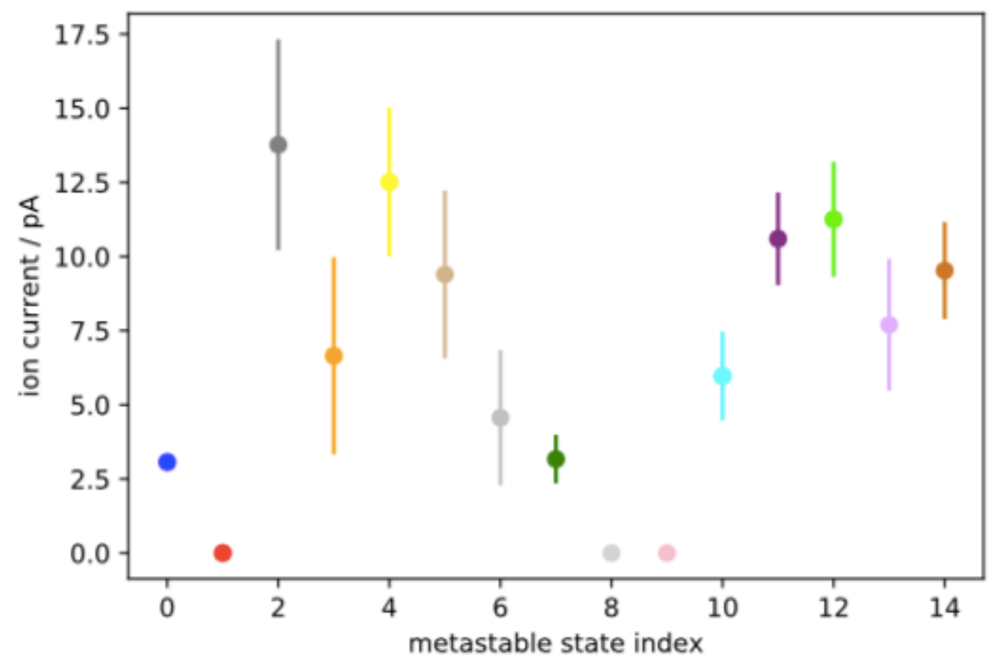
Fig1 and data: Köpfer *et al., Science*, **346**, 352 (2014).

Application: ion channel non-equilibrium MD





By clustering in the VAMP space, we identified 15 different states that differ structurally near the selectivity filter and differ in their conductivity.

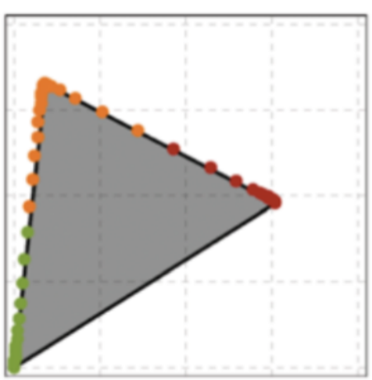


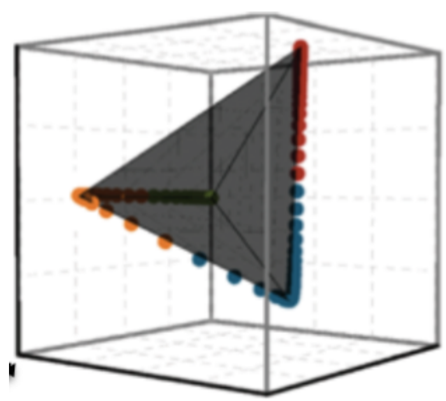
Summary and conclusion

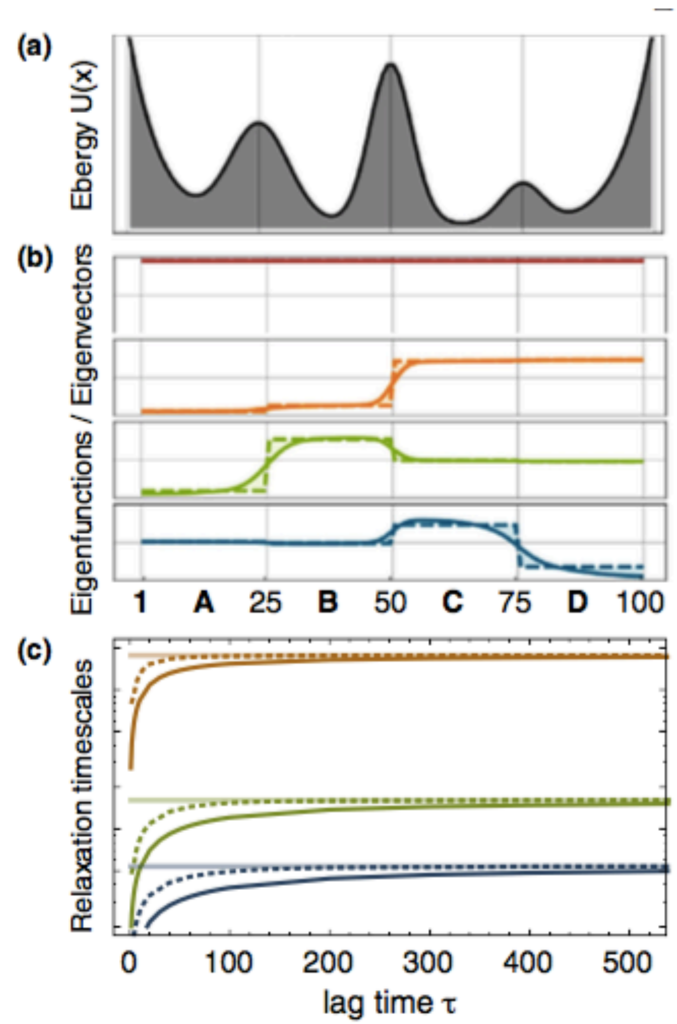
- VAC and VAMP are two variational principles that allow to approximate the true eigenfunctions of the dynamical system (VAC) or its restricted singular functions (VAMP) by using optimization.
- VAMP even works in non-equilibrium settings, if the dynamics is driven by external forces or if the sampling is so limited, that transitions in both the forward and backward directions are not available.
- VAMP can be used for feature selection and to model the slow reaction coordinates with enormously complicated functions (see talk tomorrow).

From order parameters to states to MSMs

- PCCA = Perron-cluster cluster analysis
- Motivating observation:
  the set of all MD data projected onto the dominant eigenvectors $\{\, \mathbf{v}(\mathbf{x}) \mid \mathbf{x} \in \text{data} \,\}$ form a simplex
- In 2-D simplex=triangle
  In 3-D simplex=tetrahedron
  ...



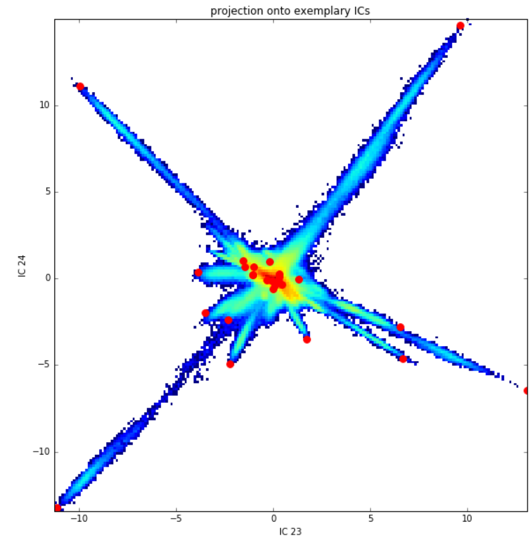




Deuflhard, Weber. *Linear Algebra Appl.,* 398 **161**, (2005).
Weber, Galliat. Tech. Rep. **02-12**, *KZZ* (2002).

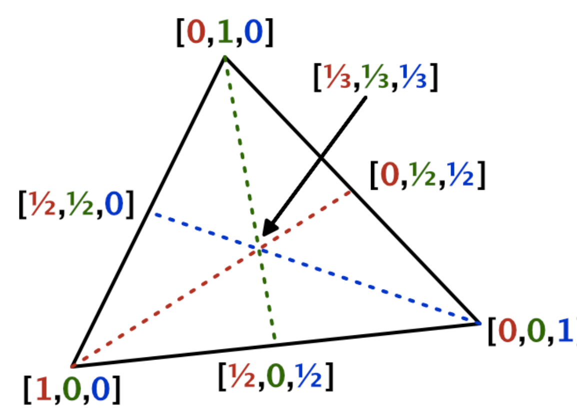
From order parameters to states to MSMs

- I: PCCA only needs the eigenvectors
- II: TICA (and VAMP) provide eigenvectors
- I&II → We can do PCCA in TICA or VAMP space.

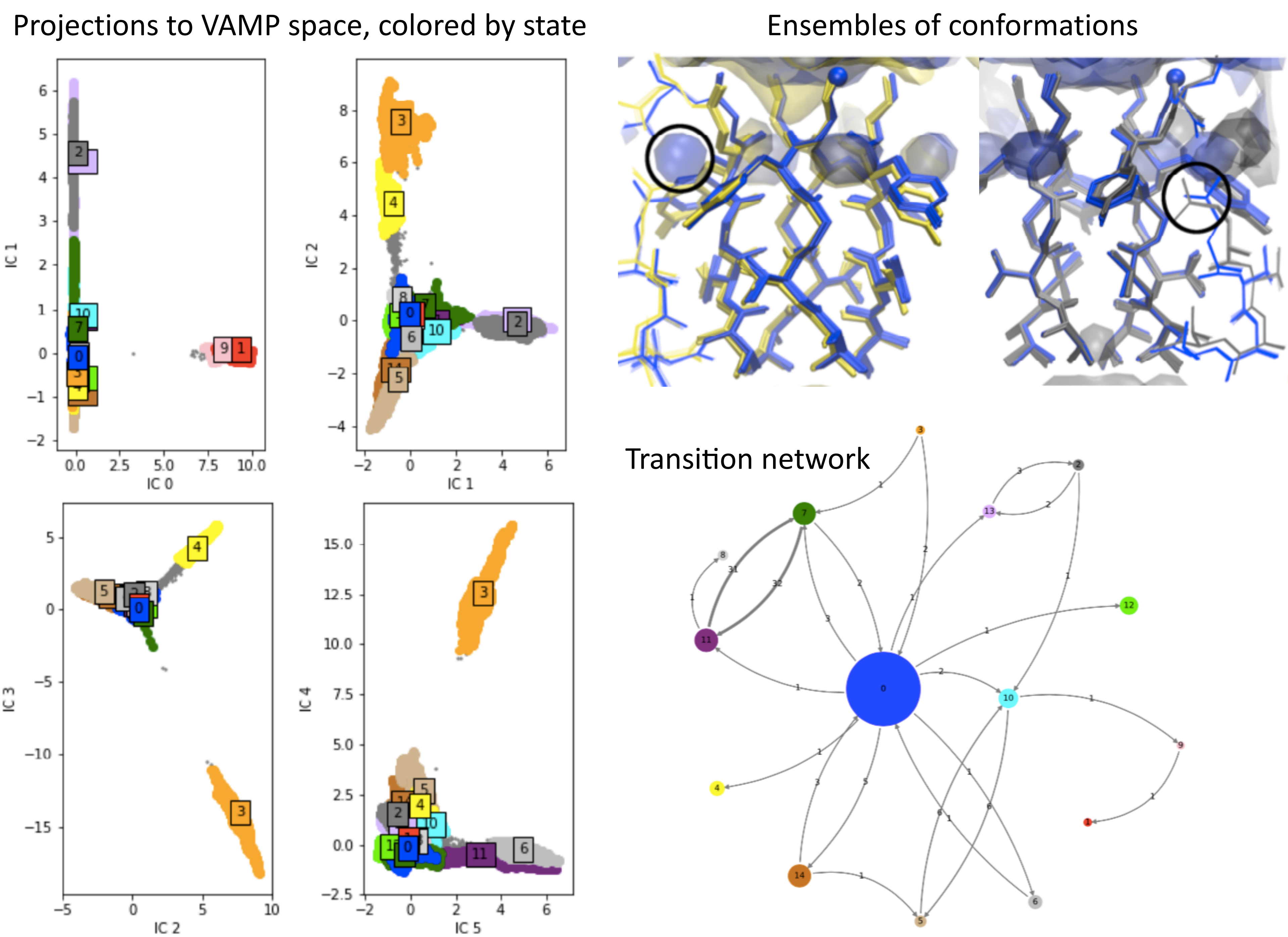


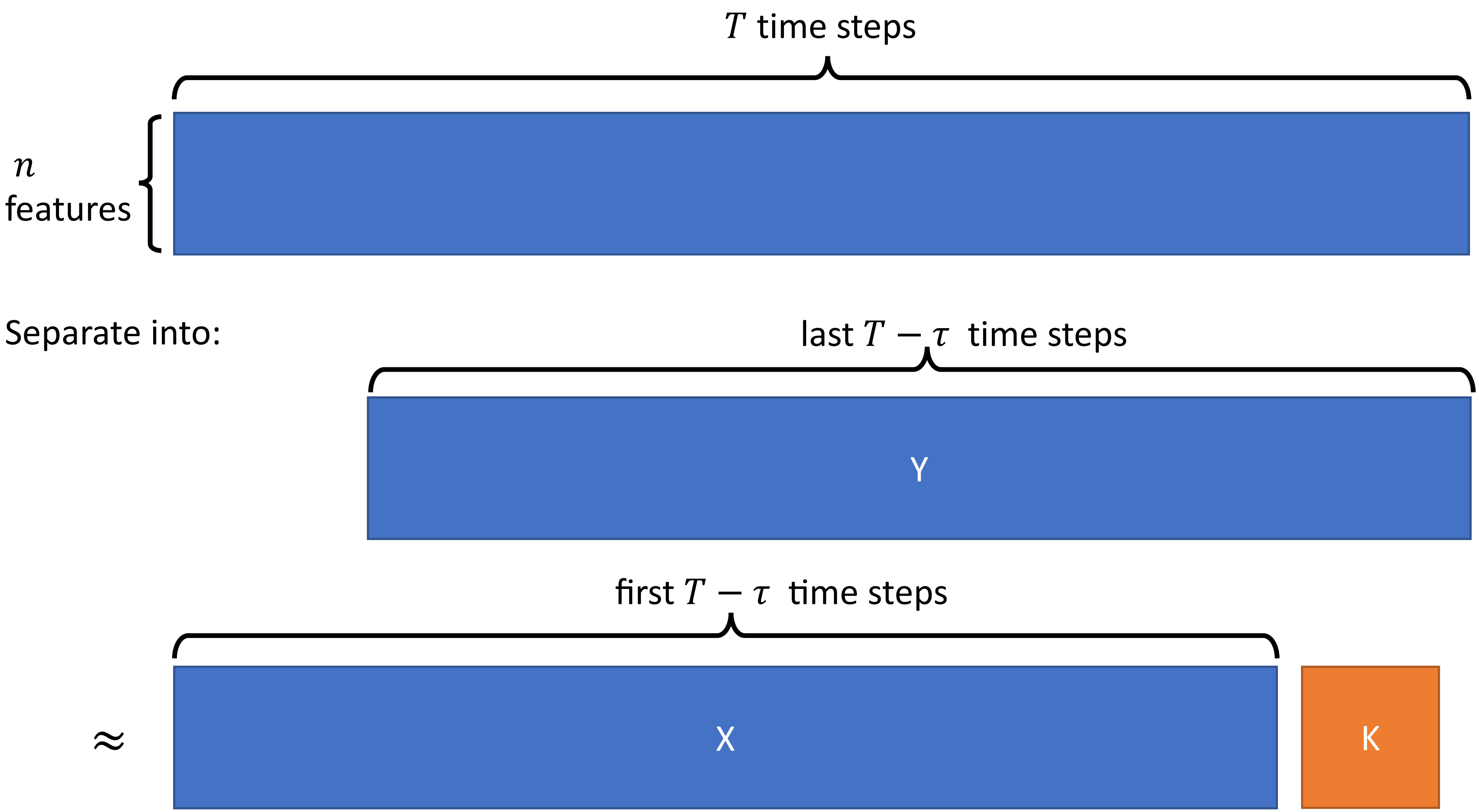
Steps of the PCCA algorithm:

1. Find the N-1 most distant points (the *vertices*) in the N-dimensional eigenspace.
2. Compute barycentric coordinates of every MD frame with respect to the N-1 vertices.



Weber, Galliat, Tech. Rep. **02-12**, *KZZ* (2002).


Projections to VAMP space, colored by state
Ensembles of conformations
Transition network
IC 0
IC 1
IC 2
IC 3
IC 4
IC 5



$$\mathbf{K} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\,\mathbf{Y} = \mathbf{C}(0)^{-1}\,\mathbf{C}(\tau)$$

$$\mathbf{K}\,\mathbf{v} = \lambda\,\mathbf{v} \Leftrightarrow \mathbf{C}(\tau)\,\mathbf{v} = \lambda\,\mathbf{C}(0)\,\mathbf{v}$$

Do a dimensionality reduction by keeping only the dominant eigenmodes.

VAMP is all about the eigendecomposition of the forward-backward transition matrix

$$
\begin{aligned}
T_{fb} := T_f T_b &= C_{00}^{-1} C_{01} C_{11}^{-1} C_{01}^{\top} \\
&= (X^T X)^{-1} X^{\top} Y (Y^{\top} Y)^{-1} Y^{\top} X
\end{aligned}
$$

For the sake of notational simplicity, I have defined $C_{00} := X^T X$, $C_{11} := Y^T Y$, and $C_{01} := X^T Y$ without normalization.

Theorem: T_{fb} has a real-valued spectrum.

Proof: Introduce the co-ordinate transformed features $\tilde{X} := X C_{00}^{-\frac{1}{2}} = X(X^{\top} X)^{-\frac{1}{2}}$ and $\tilde{Y} := Y C_{11}^{-\frac{1}{2}} = Y(Y^{\top} Y)^{-\frac{1}{2}}$. This choice leads to

$$
\begin{aligned}
\tilde{C}_{00} &:= \tilde{X}^T \tilde{X} = \mathbb{I} \\
\tilde{C}_{11} &:= \tilde{Y}^T \tilde{Y} = \mathbb{I} \\
\tilde{C}_{01} &:= \tilde{X}^T \tilde{Y} = C_{00}^{-\frac{1}{2}} C_{01} C_{11}^{-\frac{1}{2}}
\end{aligned}
$$

The new matrix $\tilde{T}_{fb}$ in the new co-ordinates is

$$
\tilde{T}_{fb} := \tilde{C}_{00}^{-1} \tilde{C}_{01} \tilde{C}_{11}^{-1} \tilde{C}_{10} = \tilde{C}_{01} \tilde{C}_{10} = \tilde{X}^{\top} \tilde{Y} \tilde{Y}^{\top} \tilde{X}
$$

Obviously, this matrix is symmetric. Therefore $\tilde{T}_{fb}$ has a real-valued spectrum.

To complete the proof, one has to show that $\tilde{T}_{fb}$ has the same eigenvalues as T_{fb}. The eigenvectors of $\tilde{T}_{fb}$ can be easily found from the eigenvectors of T_{fb} by a linear transform. Let v be an eigenvector of $\tilde{T}_{fb}$ with the corresponding eigenvalue λ.

$$\begin{aligned} \tilde{T}_{fb}v &= \lambda v && \text{(1a)} \\ \Leftrightarrow \tilde{C}_{01}\tilde{C}_{10}v &= \lambda v && \text{(1b)} \\ \Leftrightarrow C_{00}^{-\frac{1}{2}}C_{01}C_{11}^{-\frac{1}{2}}C_{11}^{-\frac{1}{2}}C_{01}^{\top}C_{00}^{-\frac{1}{2}}v &= \lambda v && \text{(1c)} \end{aligned}$$

Set $w := C_{00}^{-\frac{1}{2}}v$, then we find from the left hand side of 1a.

$$\begin{aligned} \tilde{T}_{fb}v &= C_{00}^{-\frac{1}{2}}C_{01}C_{11}^{-1}C_{01}^{\top}C_{00}^{-\frac{1}{2}}C_{00}^{\frac{1}{2}}w \\ &= C_{00}^{-\frac{1}{2}}C_{01}C_{11}^{-1}C_{01}^{\top}w \\ &= C_{00}^{\frac{1}{2}}T_{fb}w \end{aligned}$$

From the right hand side of 1a we find

$$\tilde{T}_{fb}v = \lambda C_{00}^{\frac{1}{2}}w$$

Equating left and right sides, we get

$$\begin{aligned} C_{00}^{\frac{1}{2}}T_{fb}w &= \lambda C_{00}^{\frac{1}{2}}w \\ T_{fb}w &= \lambda w \end{aligned}$$

Therefore w is an eigenvector of T_{fb} with the unchanged eigenvalue λ. Since this this hold for all eigenvectors of T_{fb}, this completes the proof.

Markov state models

MSM theory : propagator and generator

- Langevin equation

$$\ddot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x})/m - \gamma\dot{\boldsymbol{x}} + \sqrt{2k_B T\gamma/m}\,\boldsymbol{\eta}_i(t)$$

- Fokker-Planck equation

$$\frac{\partial p(t,\boldsymbol{p},\boldsymbol{x})}{\partial t} = \underbrace{\left(-\frac{\boldsymbol{p}}{m}\cdot\boldsymbol{\nabla}_x + \boldsymbol{\nabla}_p\cdot\big(\gamma\boldsymbol{p} - \boldsymbol{F}(\boldsymbol{x})\big) + \gamma k_B T m\Delta_p\right)}_{A} p(t,\boldsymbol{p},\boldsymbol{x})$$

- Propagator (operator)
  define $\boldsymbol{X} = (\boldsymbol{p},\boldsymbol{x})$

$$\mathcal{P}_\tau[p(t,.)](\boldsymbol{X}) = \exp[\tau A]p(t,.) = p(t+\tau,\boldsymbol{X})$$
$$= \int p(t,\mathbf{Y})p(\mathbf{Y}\to\mathbf{X};\tau)\mathrm{d}Y$$

- Transfer operator
  define $p(t,\boldsymbol{X}) = u(t,\boldsymbol{X})p_B(\boldsymbol{X})$

$$\mathcal{T}[u_t;\tau](\mathbf{X}) := \frac{1}{p_B(\mathbf{X})}\int u_t(\mathbf{Y})p_B(\mathbf{Y})p(\mathbf{Y}\to\mathbf{X};\tau)\mathrm{d}Y$$

Integration of the equations of

$$c_1 = e^{-\gamma\delta t}, \quad c_2 = \gamma^{-1}(1 - c_1),$$

$$c_3 = \sqrt{k_B T(1 - c_1^2)}$$

Stochastic Position Verlet (SPV)

$$x_{n+1/2} = x_n + \delta t\, M^{-1} p_n / 2;$$

$$p_{n+1} = c_1 p_n - c_2 \nabla U(x_{n+1/2}) + c_3 M^{1/2} R_{n+1};$$

$$x_{n+1} = x_{n+1/2} + \delta t\, M^{-1} p_{n+1}/2;$$

The Method of Brunger-Brooks-Karplus (1982) (BBK)

$$p_{n+1/2} = (1 - \delta t\gamma/2) p_n - \delta t\, \nabla U(x_n)/2 + \sqrt{\delta t k_B T \gamma} M^{1/2} R_n / 2;$$

$$x_{n+1} = x_n + \delta t\, M^{-1} p_{n+1/2};$$

$$p_{n+1} = [p_{n+1/2} - \delta t\, \nabla U(x_{n+1})/2 + \sqrt{\delta t k_B T \gamma} M^{1/2} R_{n+1}/2]/(1 + \delta t\gamma/2);$$

Euler-Maruyama

$$x_{n+1} = x_n - \delta t M^{-1} \nabla U(x_n) + \sqrt{2 k_B T \delta t} M^{-1/2} R_n;$$

cited from: *Leimkuhler, Matthews, Applied Mathematics Research eXpress*, **2013**, 34 (2013)

MSM theory : transfer operator

$$\mathcal{T}[u_t;\tau](\mathbf{X}) := \frac{1}{p_B(\mathbf{X})}\int u_t(\mathbf{Y})p_B(\mathbf{Y})p(\mathbf{Y}\to\mathbf{X};\tau)\mathrm{d}y$$

$$u_{t+\tau}(\mathbf{X}) = \mathcal{T}_{\mathrm{slow}}[u_t;\tau](\mathbf{X}) + \mathcal{T}_{\mathrm{fast}}[u_t;\tau](\mathbf{X})$$

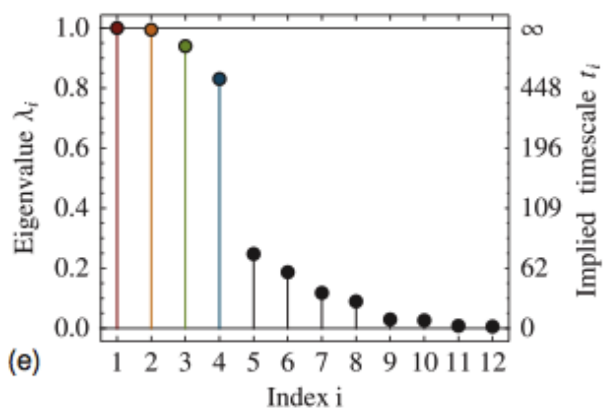
$$\mathcal{T}_{\mathrm{slow}}[u_t;\tau](\mathbf{X}) = \sum_i \lambda_i(\tau)\psi_i(\mathbf{X})\int \psi_i(\mathbf{Y})p_B(\mathbf{Y})u_t(\mathbf{Y})\mathrm{d}y = \sum_i \lambda_i(\tau)\psi_i(\mathbf{X})\langle\psi_i,u_t\rangle_{p_B}$$

$$T_{ij} = \frac{\left\langle \chi_j,\mathcal{T}[\chi_i]\right\rangle_{p_B}}{\left\langle \chi_j,\chi_i\right\rangle_{p_B}} = \frac{\iint \chi_i(\mathbf{x})p_B(\boldsymbol{Y})p(\mathbf{Y}\to\mathbf{X};\tau)\chi_j(\mathbf{X})\mathrm{d}x\mathrm{d}y}{\int \chi_j(\mathbf{X})\chi_i(\mathbf{X})p_B(\mathbf{X})\mathrm{d}y} = \frac{\mathrm{cov}(\chi_j,\chi_i;\tau)}{\mathrm{cov}(\chi_i,\chi_i;0)}$$

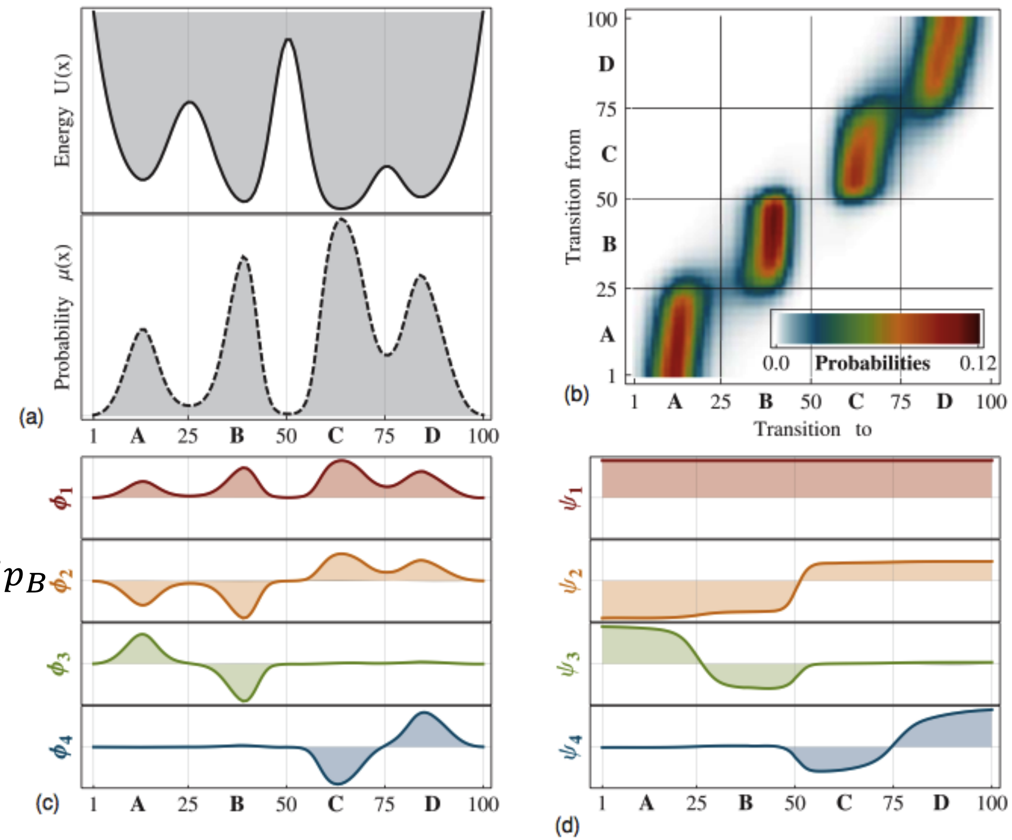
Prinz *et al*. *J. Chem. Phys.* **134**, 174105 (2011)
Sarich *et al SIAM Multiscale Model. Simul*. **8**, 1154 (2010).

MSM: spectral properties

time scales:



processes:



ϕ_i (left) ψ_i (right)

$$\mathcal{T}_{\text{slow}}[u_t;\tau](\mathbf{X}) = \sum_i \lambda_i(\tau)\psi_i(\mathbf{X})\langle\psi_i,u_t\rangle_{p_B}$$

$$\underbrace{\mathcal{T}_s\circ\cdots\circ\mathcal{T}_s}_{n\text{ times}}\, u_t = \sum_i \lambda_i^n(\tau)\psi_i\langle\psi_i,u_t\rangle_{p_B}$$

for MSM:

$$\boldsymbol{p}^T(n\tau) = \sum_i \lambda_i^n \boldsymbol{\phi}_i\,[\boldsymbol{\psi}_i\cdot\boldsymbol{p}(0)]$$

Prinz *et al.*, *J. Chem. Phys.* **134**, 174105 (2011)
Sarich *et al., SIAM Multiscale Model. Simul*. **8**, 1154 (2010).