VAMP Theory

Moritz Hoffmann

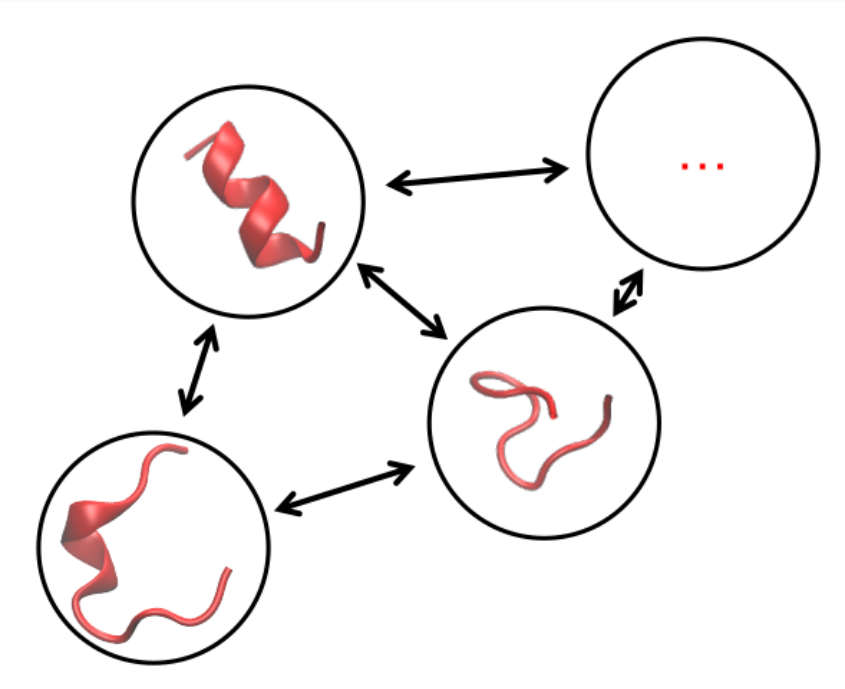
February 21, 2022

Recap: The spectral theory of MSMs

A Markov state model consists of....

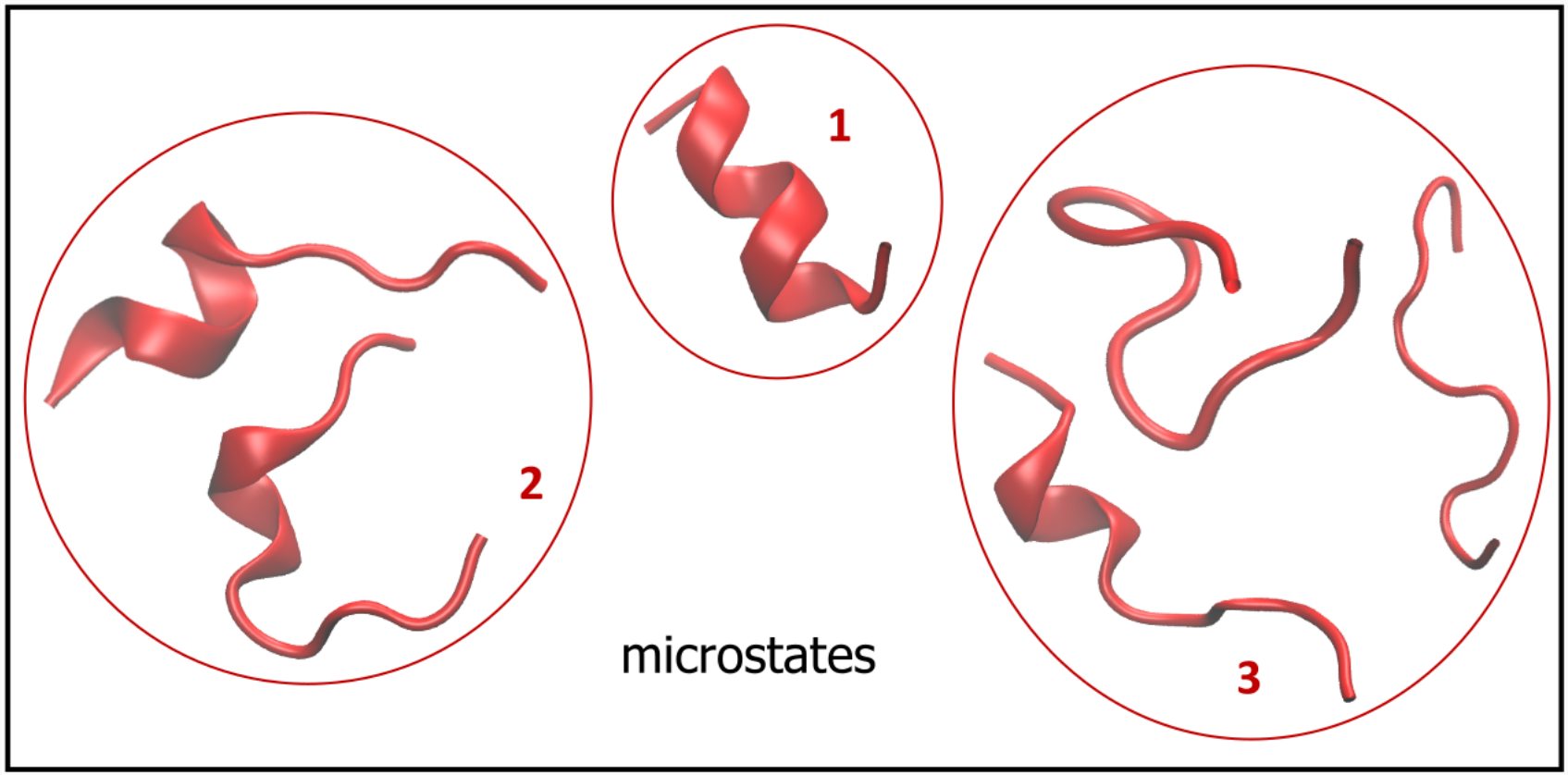
- A set of states $\{s_i\}_{i=1,\ldots,N}$,
- (conditional) transition probabilities between these states

$$P_{ij} = \mathbb{P}(s_{t+\tau} = j \mid s_t = i)$$



Markov state models: Estimation

Markov state model estimation starts with grouping related conformations into *clusters* or *microstates*.



Markov state models: Estimation

We 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

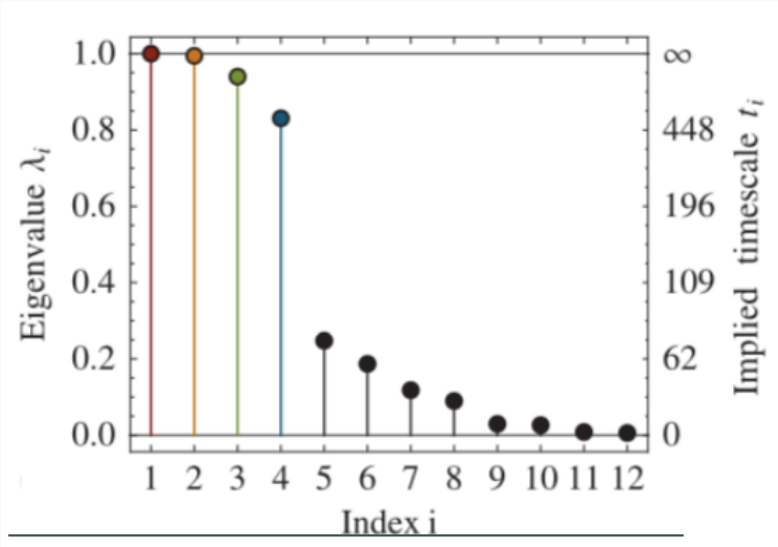
- We count transitions between microstates and tabulate them in a count matrix C.
- We estimate the transition probabilities P_{ij} from C
  - Naive estimator $\hat{P}_{ij} = C_{ij} / \sum_k C_{ik}$
  - Maximum-likelihood estimator[1][2]

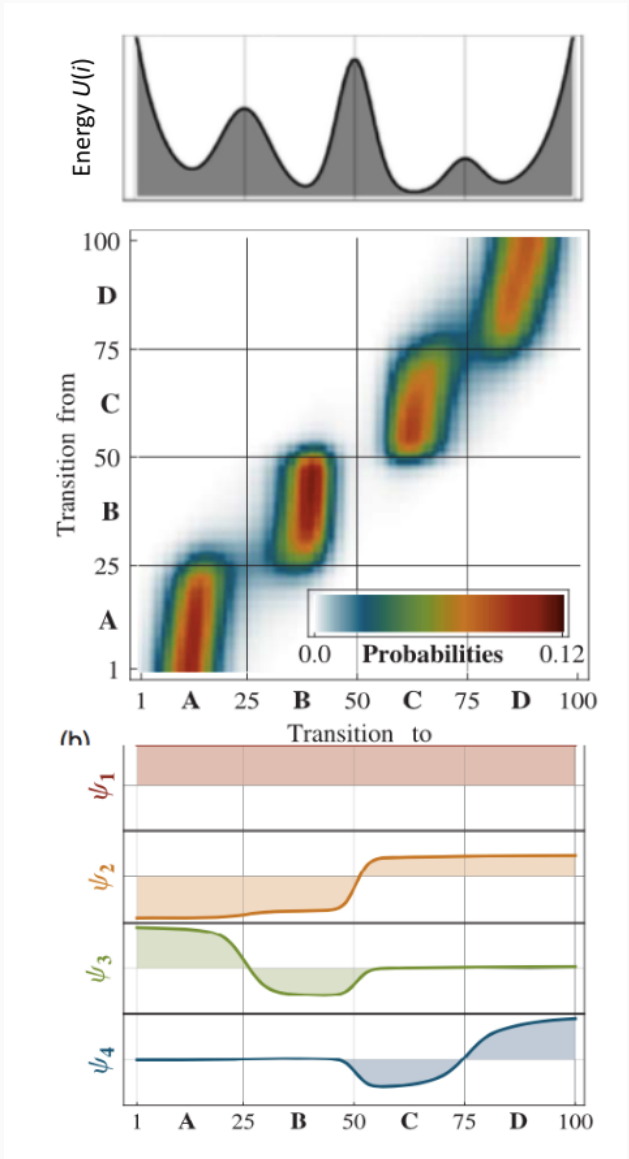
[1] Prinz et al., J. Chem. Phys. 134, 174105 (2011)

[2] Pérez-Hernández, Paul, et al., J. Chem. Phys. 139, 015102 (2013)

The spectrum of a reversible MSM[3]

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





[3] Prinz et al., J. Chem. Phys. 134, 174105 (2011)

Both MSM and TICA are based on the same spectral method

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

- Estimation of MSMs
$$T(\tau) = C_{ij}(\tau)/C_i$$
- In matrix notation
$$T(\tau) = C(0)^{-1}C(\tau)$$
- Eigenvalue problem:
$$T(\tau)\mathbf{v} = \lambda\mathbf{v} \Leftrightarrow C(0)^{-1}C(\tau)\mathbf{v} = \lambda\mathbf{v} \Leftrightarrow C(\tau)\mathbf{v} = \lambda C(0)\mathbf{v}$$
- The last equation is known as the TICA problem. All equations generalize to the case where $C(0)$ and $C(\tau)$ 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)

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

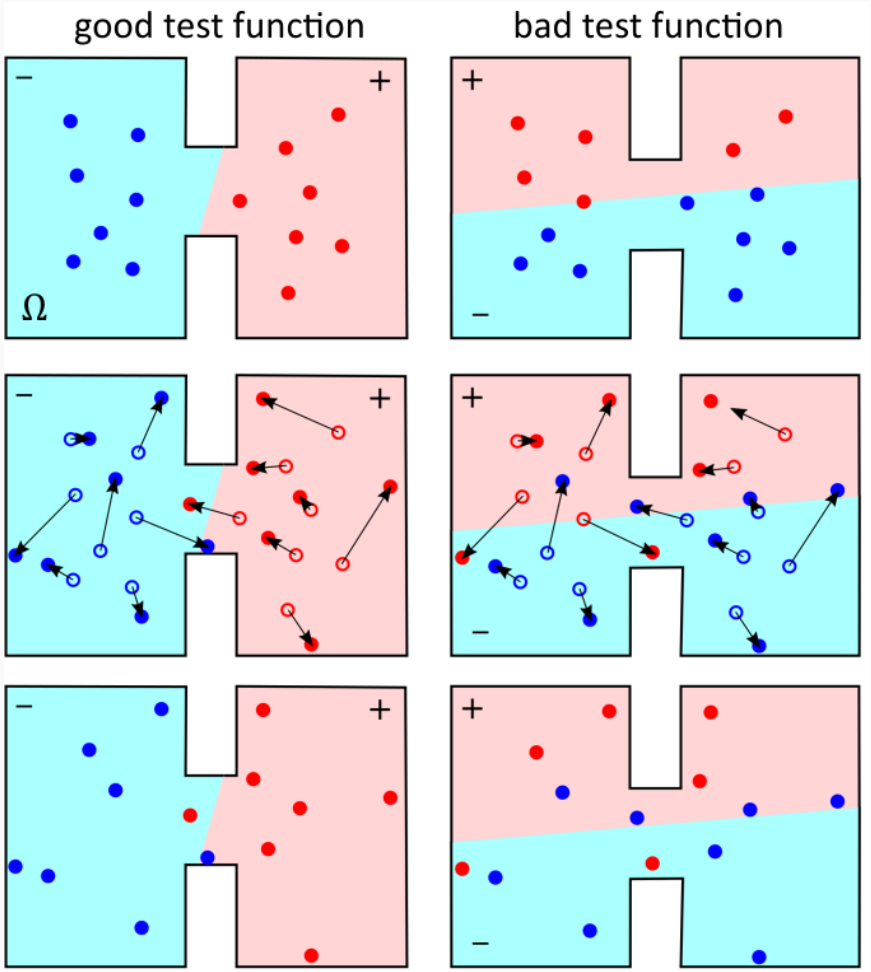
$$\mathrm{acf}(\psi, \tau) := \frac{\langle \psi, T\psi \rangle_\pi}{\langle \psi, \psi \rangle_\pi} \leq \hat{\lambda}.$$

Multiple eigenfunctions by replacing $\hat{\lambda}$ with $R_k = \sum_k \lambda_k$.

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

Interpretation of variational principle

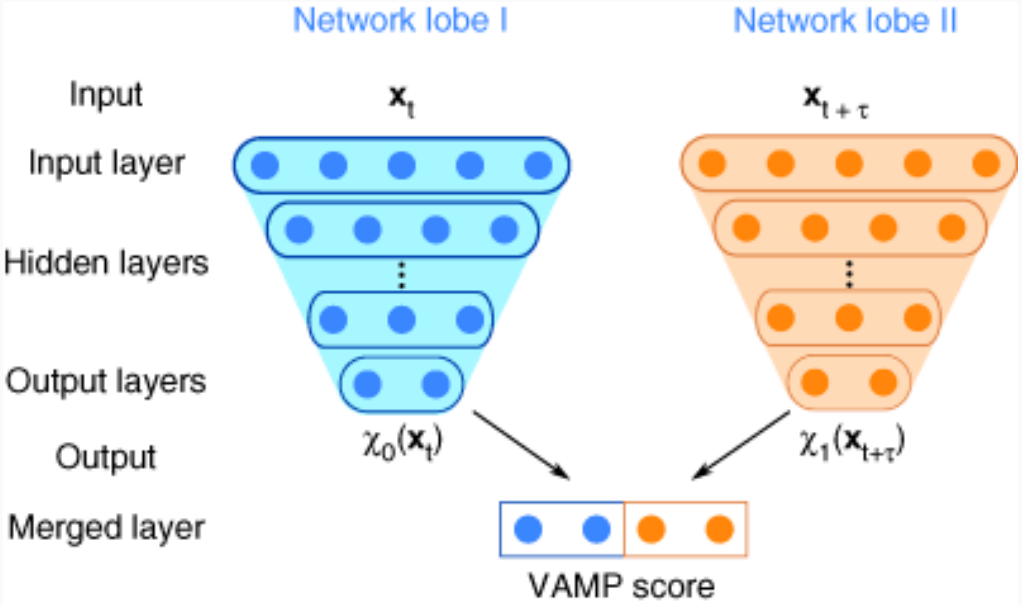
1. Pick some test function $\chi_{\text{test}}(x)$ and pick some test conformations $x_{i,\text{initial}} \sim \pi$
2. Propagate $x_{i,\text{initial}}$ with the MD integrator, call it $x_{i,\text{final}}$
3. Correlate $\chi_{\text{test}}(x_{i,\text{initial}})$ with $\chi_{\text{test}}(x_{i,\text{final}})$



$$\text{score} = \frac{\langle \chi(x_{\text{initial}}) - \bar{\chi}, \chi(x_{\text{final}}) - \bar{\chi} \rangle}{\langle \chi(x_{\text{initial}}) - \bar{\chi}, \chi(x_{\text{initial}}) - \bar{\chi} \rangle}$$

Gradient-based optimization of function parameters

Parameters p of $\chi_{\mathrm{test},\mathsf{p}}$ can be optimized with gradient-based methods like ADAM or gradient descent by using the gradient of the VAC score.



Reversible dynamics

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

$$\mathbb{P}(s_{t+\tau} = j \wedge s_t = i) = \mathbb{P}(s_{t+\tau} = i \wedge s_t = j)$$
$$\mathbb{P}(s_{t+\tau} = j \mid s_t = i)\mathbb{P}_{\mathrm{eq}}(s_t = i) = \mathbb{P}(s_{t+\tau} = i \mid s_t = j)\mathbb{P}_{eq}(s_t = j)$$
$$\pi_i T_{ij} = \pi_j T_{ji}$$

- In other notation $\langle e_i, Te_j \rangle_\pi = \langle e_j, Te_i \rangle_\pi$ with $\langle x, y \rangle_\pi = \sum_i x_i x_j \pi_i$
- So T is symmetric with respect to $\langle \cdot, \cdot \rangle_\pi$
- That means that T has real eigenvalues and eigenvectors.

Nonreversible dynamics

- $R_k = \sum_i^k \lambda_i$ where λ_i are the true eigenvalues
- For nonreversible dynamics $\langle e_i, Te_j \rangle_\pi \neq \langle e_j, Te_i \rangle_\pi$
- Eigenvalues and eigenvectors are in general complex
- Variational principle does not work: $\mathrm{acf}(\psi) \leq \hat{\lambda} \in \mathbb{C}$ makes no sense

How to fix this?

A possible solution: Variational approach to Markov processes (VAMP)

- Introduce the "backward" transition matrix

$$T_b := C(N)^{-1}C(-\tau) = C(N)^{-1}C^\top(\tau)$$

  i.e. estimate MSM/TICA from time-reversed data, where

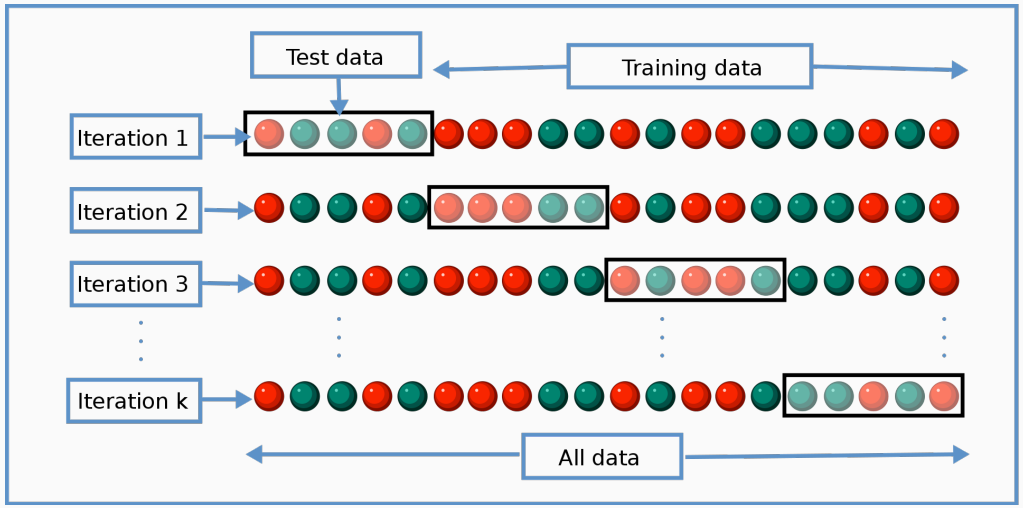
$$C_{ij}(-\tau) = \sum_{t=\tau}^{N} f_i(x(t-\tau))f_j(x(t)), \quad C_{ij}(N) \quad = \sum_{t=\tau}^{N} f_i(x(t))f_j(x(t))$$

- Introduce forward=backward transition matrix $T_{fb} := TT_b$
- One can show that T_{fb} (and T_{bf}) are symmetric matrices
- Therefore eigenvalues and eigenvectors are real
- They fulfill a variational principle $\|C^{-1/2}(0)C(\tau)C(N)^{-1/2}\| \leq R$.

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

Cross-validation

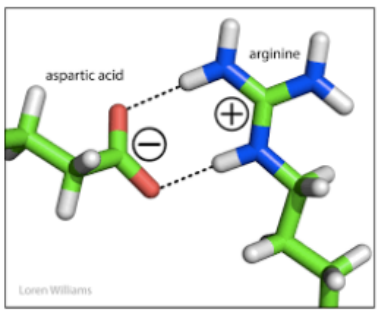
- Reporting a test-score that was computed from independent realizations
- Independent realizations can be expensive to sample.
- To the approximate $k-$ *fold* (hold-out) cross-validation:
  - Split all data into **training** and **validation** sets
  - Optimize the model parameters with training set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



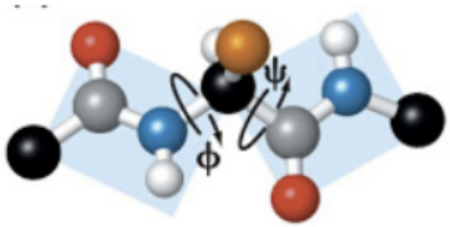
- k-fold CV can be tricky with highly autocorrelated time series data!

Application: feature selection

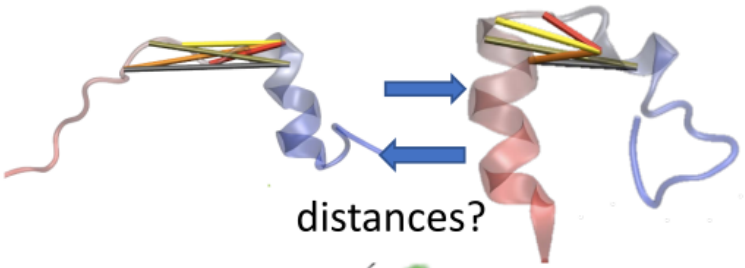
- Variational principle: higher score is better
- Compare test scores for different selections of molecular features



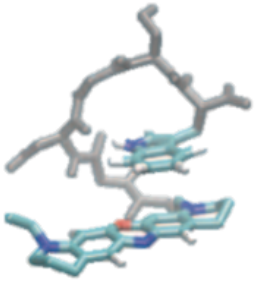
contacts?



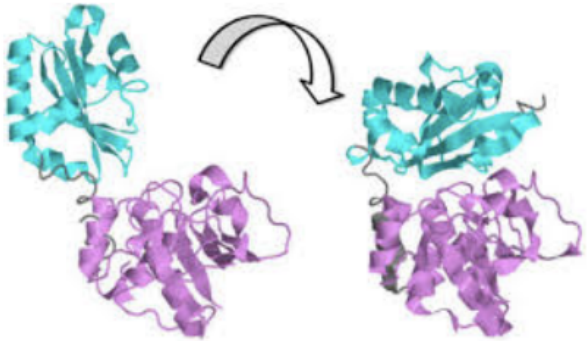
dihedrals?



distances?



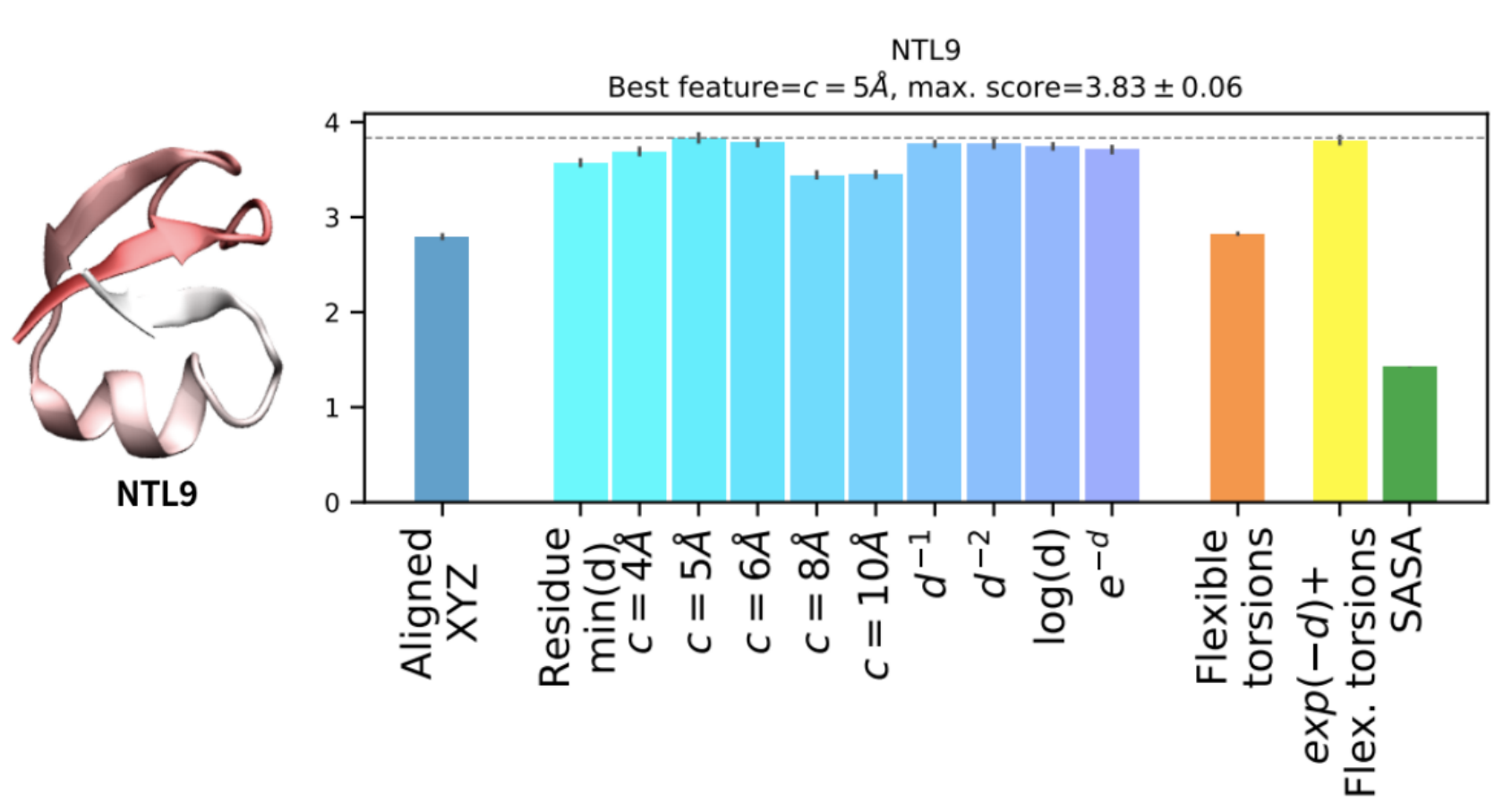
chemical intuition?



rigid body approximation?

side chain flips?

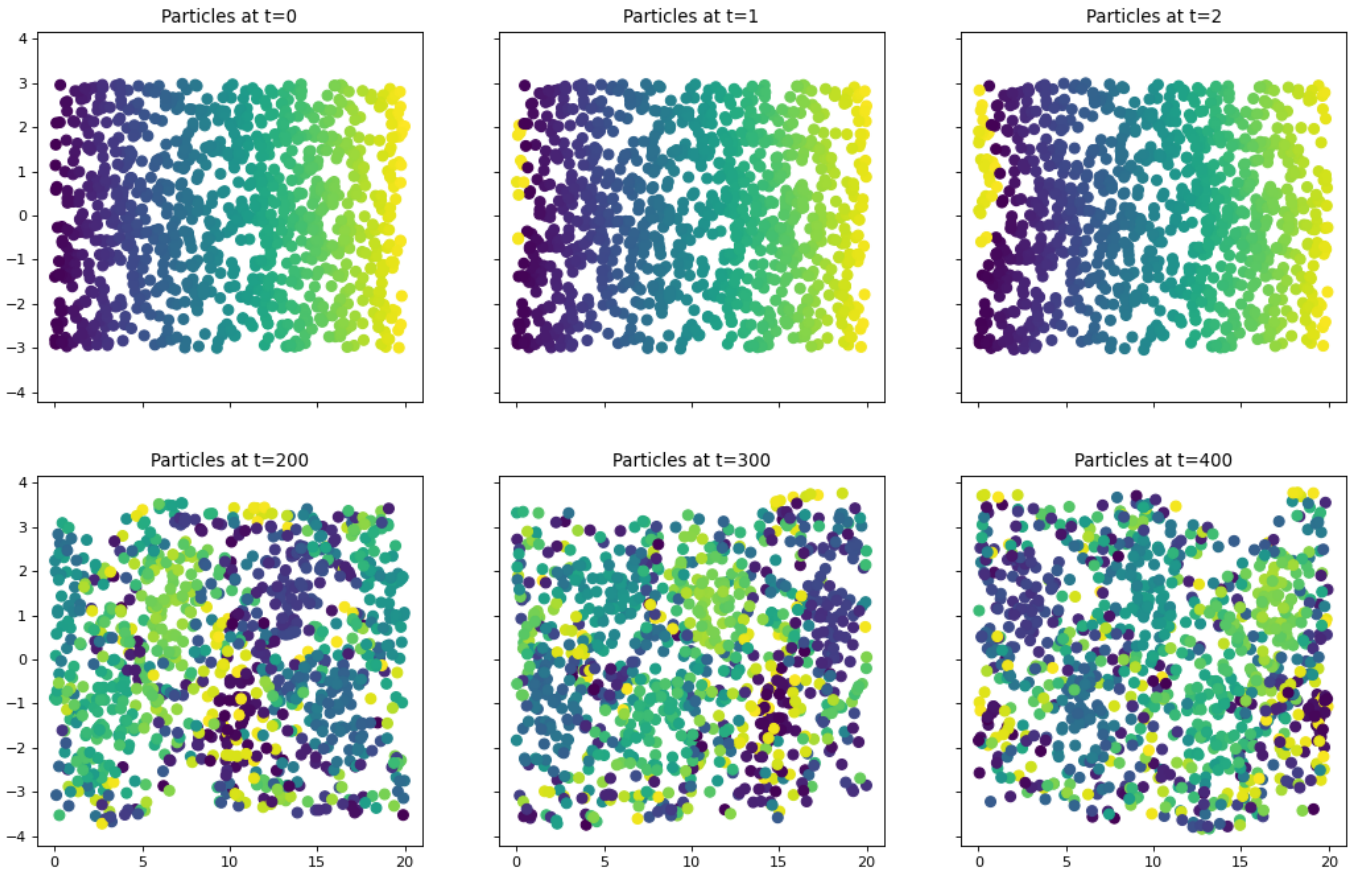
Application: feature selection



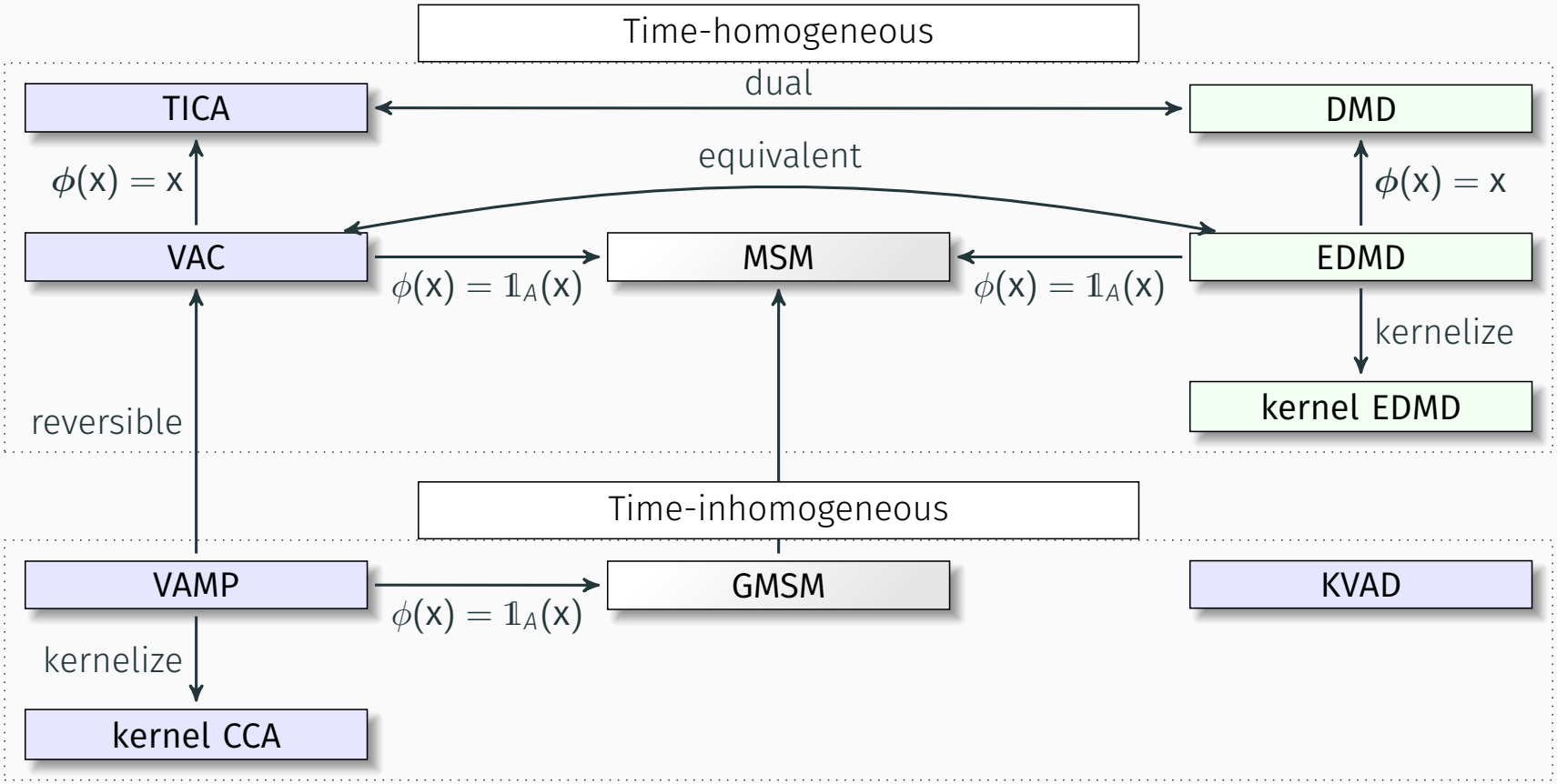
Scherer et al J. Chem. Phys. 150, 194108 (2019)

Coherent sets with VAMPNets

Optimize feature functions by maximizing VAMP score to find coherent sets.



Methods in context



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 optimization
- VAMP even works in non-equilibrium settings
- VAMP can be used for feature selection via gradient-based methods

Questions?

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.