

MAX PLANCK INSTITUTE
FOR DYNAMICS OF COMPLEX
TECHNICAL SYSTEMS
MAGDEBURG


DATA-DRIVEN MODELING OF
COMPLEX PHYSICAL SYSTEMS

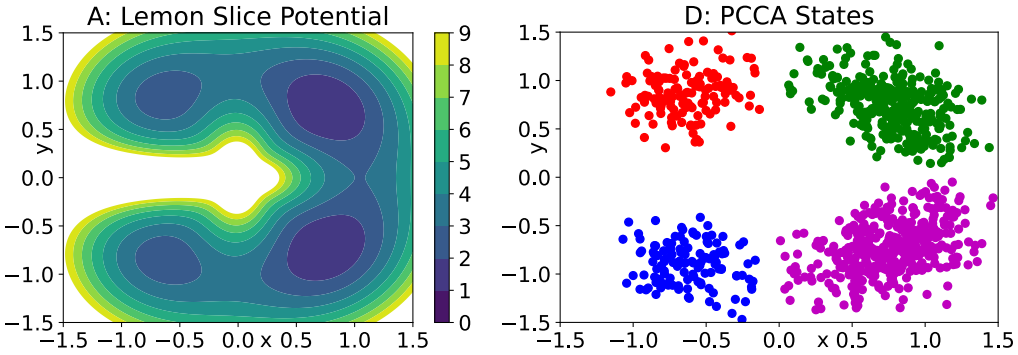
Kernel Methods for Koopman-based Modeling in Molecular Simulation

Feliks Nüske

April 11, 2024

Metastable Systems

- **Goal:** Automatically analyse *metastable* systems based on simulation data.
- Example: Langevin Dynamics $dX_t = -\nabla V(X_t)\,\mathrm{d}t + \sqrt{2kT}\,\mathrm{d}W_t$.



Outline

1. The Koopman Operator and EDMD
2. Variational Approach
3. Generator Learning
4. Kernel Methods and Random Features

Extended Dynamic Mode Decomposition (EDMD)

Idea: Track a dynamical system X_t through the lens of **observables** by means of a **linear map**:

Extended Dynamic Mode Decomposition (EDMD)

Idea: Track a dynamical system X_t through the lens of **observables** by means of a **linear map**:

- Choose finitely many observables:

$$\psi(x) = \left[\psi_1(x) \cdots \psi_n(x)\right]^T.$$

Extended Dynamic Mode Decomposition (EDMD)

Idea: Track a dynamical system X_t through the lens of **observables** by means of a **linear map**:

- Choose finitely many observables:

$$\psi(x) = \begin{bmatrix}\psi_1(x) \cdots \psi_n(x)\end{bmatrix}^T.$$

- Generate transformed snapshot matrices (x_k, y_k separated by **lag time** t):

$$\Psi(\mathbf{X}) = \begin{bmatrix}\psi(x_1) & | & \cdots & | & \psi(x_m)\end{bmatrix} \in \mathbb{R}^{n\times m},$$
$$\Psi(\mathbf{Y}) = \begin{bmatrix}\psi(y_1) & | & \cdots & | & \psi(y_m)\end{bmatrix} \in \mathbb{R}^{n\times m}.$$

Extended Dynamic Mode Decomposition (EDMD)

Idea: Track a dynamical system X_t through the lens of **observables** by means of a **linear map**:

- Choose finitely many observables:

$$\psi(x) = \begin{bmatrix}\psi_1(x) \cdots \psi_n(x)\end{bmatrix}^T.$$

- Generate transformed snapshot matrices (x_k, y_k separated by **lag time** t):

$$\Psi(\mathbf{X}) = \begin{bmatrix}\psi(x_1) & | & \cdots & | & \psi(x_m)\end{bmatrix} \in \mathbb{R}^{n\times m},$$
$$\Psi(\mathbf{Y}) = \begin{bmatrix}\psi(y_1) & | & \cdots & | & \psi(y_m)\end{bmatrix} \in \mathbb{R}^{n\times m}.$$

- Solve **regression** problem (EDMD):

$$\begin{aligned}\mathbf{K}^t &= \operatorname{argmin}_{K\in\mathbb{R}^{n\times n}} \|\Psi(\mathbf{Y}) - \mathbf{K}^T\Psi(\mathbf{X})\|_F \\ &= (\Psi(\mathbf{X})^T\Psi(\mathbf{X}))^{-1}(\Psi(\mathbf{X})^T\Psi(\mathbf{Y})).\end{aligned}$$

Williams et al, JNLS (2015)

Statistical Limit

- Infinite Data Limit:

$$\frac{1}{m}\left[\Psi(\mathbf{X})^T\Psi(\mathbf{X})\right]_{ij} \to \int_{\mathbb{X}} \psi_i(x)\,\psi_j(x)\,\mathrm{d}\rho(x),$$
$$\frac{1}{m}\left[\Psi(\mathbf{X})^T\Psi(\mathbf{Y})\right]_{ij} \to \int \psi_i(x)\,\mathbb{E}[\psi_j(\mathcal{X}_t)|\mathcal{X}_0 = x]\,\mathrm{d}\rho(x).$$

- EDMD learns a finite-dimensional projection of the **Koopman operator**:

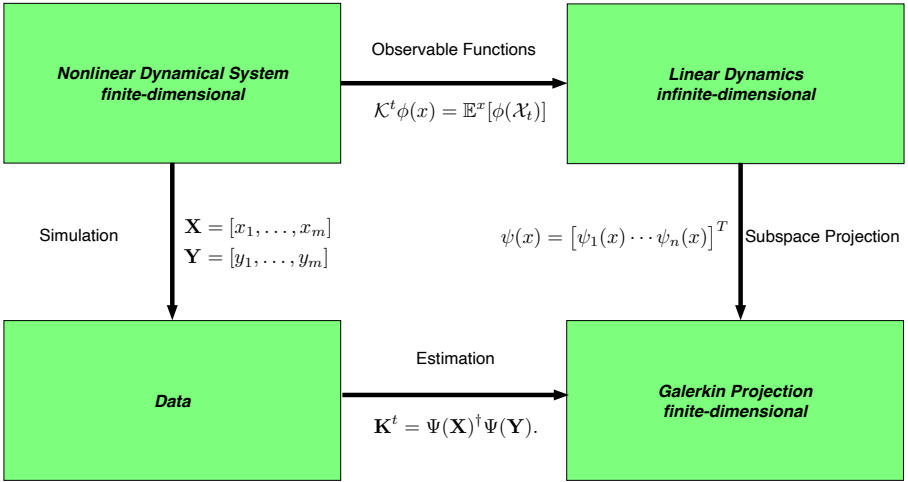
$$\mathcal{K}^t\phi(x) = \mathbb{E}[\phi(\mathcal{X}_t)|\mathcal{X}_0 = x].$$

Reviews:
Klus, **FN**, et al, JNLS, 2018,
Mauroy, Suzuki, Mezic (eds), *Koopman operator in systems and control*, Springer 2020,

Berry, Giannakis, Harlim, Notices of the AMS, 2020.

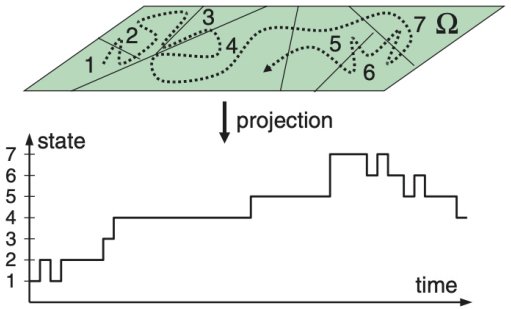
The Koopman Approach

Main Idea: **lifting** into an **infinite-dimensional** space where the dynamics become linear.



Example: MSM

Special case of EDMD framework for piece-wise constant basis functions



$$\psi_i(x) = \begin{cases} 1 & x \in S_i \\ 0 & \text{else} \end{cases},$$

$$(\Psi(\mathbf{X})^T \Psi(\mathbf{X}))_{ij} = c_i \delta_{ij},$$

$$(\Psi(\mathbf{X})^T \Psi(\mathbf{Y}))_{ij} = c_{ij}.$$

from Prinz et al, JCP (2011)

Outline

Variational Principle

- We now consider **reversible** systems with invariant measure μ; Koopman operator $\mathcal{K}^t$ is then **self-adjoint**.

Variational Principle

- We now consider **reversible** systems with invariant measure μ; Koopman operator $\mathcal{K}^t$ is then **self-adjoint**.
- To understand metastability, we need to compute the *dominant eigenvalues* of the Koopman operator: $1 = \lambda_0(t) \geq \ldots \geq \lambda_M(t)$.

Variational Principle

- We now consider **reversible** systems with invariant measure μ; Koopman operator $\mathcal{K}^t$ is then **self-adjoint**.
- To understand metastability, we need to compute the *dominant eigenvalues* of the Koopman operator: $1 = \lambda_0(t) \geq \ldots \geq \lambda_M(t)$.
- **Rayleigh variational principle** (RVP):

$$\sum_{i=0}^{M} \left\langle \phi_i,\, \mathcal{K}^t \phi_i \right\rangle_\mu =: \mathcal{R}(\phi) \leq \sum_{i=0}^{M} \lambda_i(t) \tag{1}$$

$$\left\langle \phi_k,\, \phi_l \right\rangle_\mu = \delta_{kl}. \tag{2}$$

Variational Principle

- We now consider **reversible** systems with invariant measure μ; Koopman operator $\mathcal{K}^t$ is then **self-adjoint**.
- To understand metastability, we need to compute the *dominant eigenvalues* of the Koopman operator: $1 = \lambda_0(t) \geq \ldots \geq \lambda_M(t)$.
- **Rayleigh variational principle** (RVP):

$$\sum_{i=0}^{M} \left\langle \phi_i,\, \mathcal{K}^t \phi_i \right\rangle_\mu =: \mathcal{R}(\phi) \leq \sum_{i=0}^{M} \lambda_i(t) \tag{1}$$

$$\left\langle \phi_k,\, \phi_l \right\rangle_\mu = \delta_{kl}. \tag{2}$$

- $R(\psi)$ can be used as an **objective function**, subject to the constraint (2).
- All quantities in (1-2) can be **estimated from simulation data**:

$$\left\langle \phi_i,\, \phi_i \right\rangle_\mu \approx \frac{1}{m} \sum_{k=1}^{m} \phi_i(x_k)\phi_i(x_k), \quad \left\langle \phi_i,\, \mathcal{K}^t \phi_i \right\rangle_\mu \approx \frac{1}{m} \sum_{k=1}^{m} \phi_i(x_k)\phi_i(y_k).$$

Linear Variational Approach

- Applied to a finite-dimensional subspace $\mathrm{span}\{\psi_j\}_{j=1}^n$:

$$\phi_i = \sum_{j=1}^{n} \mathbf{v}_{ji}\psi_j,$$

- leads to generalized eigenvalue problem for Gramian matrices:

$$\mathbf{C}^t\mathbf{V} = \mathbf{C}^0\mathbf{V}\Lambda, \qquad \mathbf{C}^t = \frac{1}{m}\sum_{l=1}^{m}\psi_i(x_k)\psi_j(y_k), \qquad \mathbf{C}^0 = \frac{1}{m}\sum_{l=1}^{m}\psi_i(x_k)\psi_j(x_k).$$

Noé and **FN**, SIAM MMS, 2013, **FN**, Keller, et al, JCTC, 2014
TICA: Pérez-Hernández et al, JCP, 2013,
Deep Learning: Mardt et al, Nat. Commun., 2018

Non-reversible systems: Wu and Noé, JNLS, 2020

Outline

Generator Learning (gEDMD)

- Consider the **generator** of the Koopman operator:

$$\mathcal{L}\psi = \lim_{t\to 0} \frac{1}{t}\left[\mathcal{K}^t - \mathrm{Id}\right]\psi, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}^t\psi(x) = \mathcal{L}\mathcal{K}^t\psi(x).$$

Generator Learning (gEDMD)

- Consider the **generator** of the Koopman operator:

$$\mathcal{L}\psi = \lim_{t\to 0} \frac{1}{t}\left[\mathcal{K}^t - \mathrm{Id}\right]\psi, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}^t\psi(x) = \mathcal{L}\mathcal{K}^t\psi(x).$$

- For SDEs, generator is known in analytical form:

$$\mathcal{L}\psi = -\nabla V \cdot \nabla\psi + k_B T \Delta\psi.$$

Generator Learning (gEDMD)

- Consider the **generator** of the Koopman operator:

$$\mathcal{L}\psi = \lim_{t\to 0} \frac{1}{t}\left[\mathcal{K}^t - \mathrm{Id}\right]\psi, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}^t\psi(x) = \mathcal{L}\mathcal{K}^t\psi(x).$$

- For SDEs, generator is known in analytical form:

$$\mathcal{L}\psi = -\nabla V \cdot \nabla\psi + k_B T \Delta\psi.$$

- Learn a data-driven matrix model analogous to EDMD:

$$\Psi(\mathbf{X}) = \begin{bmatrix} \psi_1(x_1) & \dots & \psi_1(x_m) \\ \vdots & \ddots & \vdots \\ \psi_n(x_1) & \dots & \psi_n(x_m) \end{bmatrix}, \qquad \mathcal{L}\Psi(\mathbf{X}) = \begin{bmatrix} \mathcal{L}\psi_1(x_1) & \dots & \mathcal{L}\psi_1(x_m) \\ \vdots & \ddots & \vdots \\ \mathcal{L}\psi_n(x_1) & \dots & \mathcal{L}\psi_n(x_m) \end{bmatrix},$$

- matrix model:

$$\mathbf{L} = \left(\Psi(\mathbf{X})\Psi(\mathbf{X})^\top\right)^{-1}\left(\Psi(\mathbf{X})\mathcal{L}\Psi(\mathbf{X})^\top\right).$$

Coarse Grained Generator

- Choose a **coarse graining (CG)** map: $\xi : \mathbb{R}^d \mapsto \mathbb{R}^m$, $m \leq d$.

Coarse Grained Generator

- Choose a **coarse graining (CG)** map: $\xi : \mathbb{R}^d \mapsto \mathbb{R}^m$, $m \leq d$.
- Let $\mathcal{P}$ be the conditional expectation with respect to ξ:

$$\mathcal{P}\phi(z) = \frac{1}{\nu(z)}\mathbb{E}[\phi(x)|\xi(x) = z].$$

Coarse Grained Generator

- Choose a **coarse graining (CG)** map: $\xi : \mathbb{R}^d \mapsto \mathbb{R}^m$, $m \leq d$.
- Let $\mathcal{P}$ be the conditional expectation with respect to ξ:

$$\mathcal{P}\phi(z) = \frac{1}{\nu(z)}\mathbb{E}[\phi(x)|\xi(x) = z].$$

- Projected generator $\mathcal{L}^\xi = \mathcal{P}\mathcal{L}\mathcal{P}$ is **again the generator** of an **effective SDE** on $\mathbb{R}^m$.

$$d\mathcal{Z}_t = b^\xi(\mathcal{Z}_t)\,\mathrm{d}t + \sigma^\xi(\mathcal{Z}_t)\,\mathrm{d}W_t.$$

Legoll and Lelièvre, *Nonlinearity* (2010), Zhang, Hartmann, Schütte, *Faraday Disc.* (2016)

Learning the Reduced Generator

- Given a CG map ξ, choose basis functions **on CG space** $\psi_i = \psi_i(\xi(x))$.

[1] Zhang et al, *Faraday Disc.* (2016), [2] Klus, **FN**, Peitz, et al, *Physica D* (2020)

Learning the Reduced Generator

- Given a CG map ξ, choose basis functions **on CG space** $\psi_i = \psi_i(\xi(x))$.
- We have the fundamental identity [1]

$$\langle \psi_i,\, \mathcal{L}\psi_j \rangle_\mu = \left\langle \psi_i,\, \mathcal{L}^\xi \psi_j \right\rangle_\nu .$$

[1] Zhang et al, *Faraday Disc.* (2016), [2] Klus, **FN**, Peitz, et al, *Physica D* (2020)

Learning the Reduced Generator

- Given a CG map ξ, choose basis functions **on CG space** $\psi_i = \psi_i(\xi(x))$.
- We have the fundamental identity [1]

$$\langle \psi_i, \mathcal{L}\psi_j \rangle_\mu = \left\langle \psi_i, \mathcal{L}^\xi \psi_j \right\rangle_\nu .$$

- Therefore, gEDMD **simultaneously** provides a matrix approximation of $\mathcal{L}$ and $\mathcal{L}^\xi$. [2].

[1] Zhang et al, *Faraday Disc.* (2016), [2] Klus, **FN**, Peitz, et al, *Physica D* (2020)

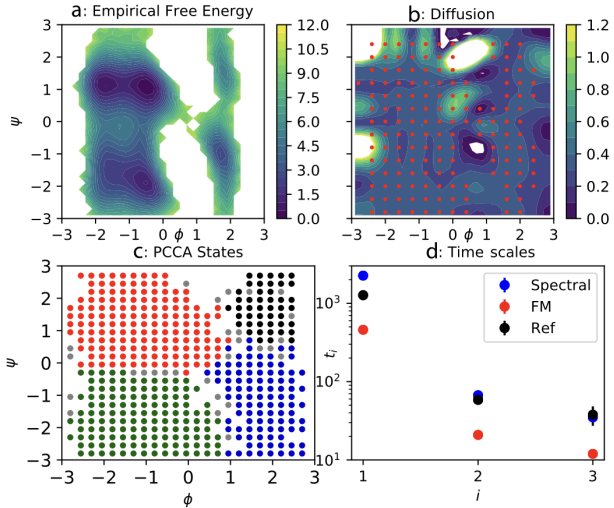
System Identification

- Effective diffusion can be learned using models for the CG generator models:

$$\left\langle \mathcal{L}_\theta^\xi \psi_i,\, f_j \right\rangle_\nu = -\frac{1}{2} \int \nabla \psi_i(z) \cdot A_\theta^\xi(z) \cdot \nabla f_j(z)\, \mathrm{d}\nu.$$

- Can be combined with force matching to complete CG dynamics.

FN, Boninsegna and Clementi, *JCP* (2019)

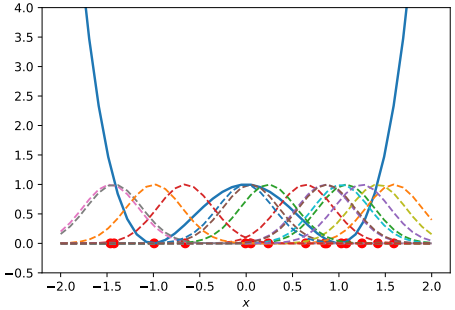


Outline

1. The Koopman Operator and EDMD
2. Variational Approach
3. Generator Learning
4. Kernel Methods and Random Features

Data-driven Basis Sets

- Choosing a good basis set is hard.
- Idea: let the data define the basis.



- Typical choice: **radial basis functions**, e.g.

$$k(x_i, y) = \exp\left(-\frac{1}{2\sigma^2}\|x_i - y\|^2.\right)$$

EDMD and RKHS

- Kernel-based basis functions $k(x_i, \cdot)$ lead to matrices...

$$\mathbf{C}^0(r,s) = \mathbf{K}_X(r,s) = k(x_r, x_s), \qquad \mathbf{C}^t(r,s) = \mathbf{K}_X^t(r,s) = k(y_r, x_s).$$

- ... and the generalized eigenvalue problem:

$$\mathbf{K}_X^t \mathbf{w}_i = \lambda_i(t) \mathbf{K}_X \mathbf{w}_i.$$

- Both matrices scale with the data size $(m \times m)$.

Klus et al, JNLS, 2020; Klus, **FN**, and Hamzi, Entropy, 2020

Intro RFF

- A translation invariant kernel with $k(x, x) = 1$ can be written as a superposition of complex plane waves:

$$k(x, y) = \mathbb{E}^{\omega \sim \rho} \left[e^{-i\omega^T x} \overline{e^{-i\omega^T y}} \right] \approx \frac{1}{p} \sum_{u=1}^{p} e^{-i\omega_u^T x} \overline{e^{-i\omega_u^T y}},$$

  where ρ is the **spectral measure** in frequency space (Bochner's theorem).
- The spectral measure is known for most popular kernels, sampling from ρ is easy.

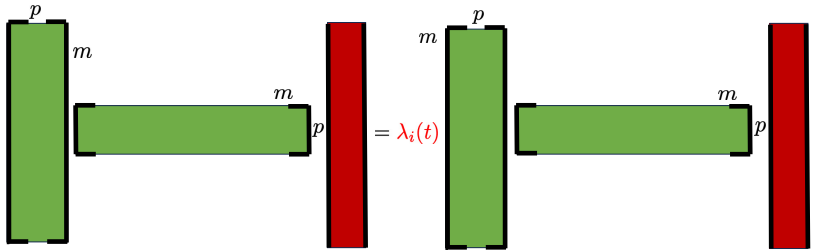
Rahimi and Recht, Advances in Neural Information Processing Systems, 2007

Low-Rank Kernel GEV

- **Low-rank** rep of kernel Koopman GEV:

$$\mathbf{K}_X = [k(x_r, x_s)]_{r,s} \approx \frac{1}{p}\left[\mathbf{M}\mathbf{M}^{\mathrm{H}}\right]_{r,s}, \qquad \mathbf{K}_X^t = [k(y_r, x_s)]_{r,s} \approx \frac{1}{p}\left[\mathbf{M}^t\mathbf{M}^{\mathrm{H}}\right]_{r,s},$$

$$\mathbf{M} = \left[e^{-ix_r^\top \omega_u}\right]_{r,u} \in \mathbb{C}^{m\times p}, \qquad \mathbf{M}^t = \left[e^{-iy_r^\top \omega_u}\right]_{r,u} \mathbb{C}^{m\times p}.$$



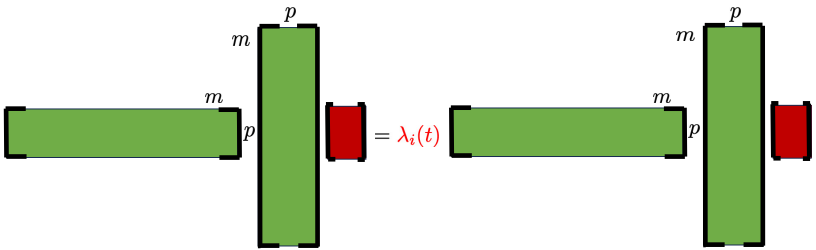
FN and Klus, JCP, 2023

Low-Rank Kernel GEV

- Non-zero eigenvalues can be obtained from dual problem

$$\mathbf{M}^{\mathrm{H}}\mathbf{M}^{t}\mathbf{v}_i = \hat{\lambda}_i(t)\mathbf{M}^{\mathrm{H}}\mathbf{M}\mathbf{v}_i,$$

which is of dimension $p \times p$.



FN and Klus, JCP, 2023

Algorithm

Algorithm 1 RFF-based Spectral Approximation of the Koopman Operator

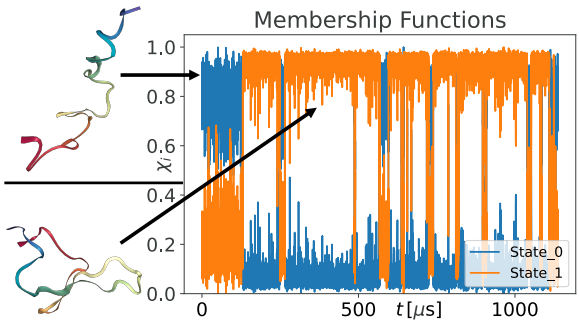
Input: data matrices $\mathbf{X} = [x_1, \ldots, x_m] \in \mathbb{R}^{d \times m}$, $\mathbf{Y} = [y_1, \ldots, y_m] \in \mathbb{R}^{d \times m}$
kernel function k with spectral measure ρ,
number of features p, truncation rule for singular values.

Output: Approximate eigenpairs $(\hat{\lambda}_i(t), \hat{\psi}_i)$.

1: Draw p samples $\{\omega_u\}_{u=1}^{p}$ from the spectral measure ρ.
2: Form matrices $\mathbf{M} = \left[e^{-ix_r^\top \omega_u}\right]_{r,u}$, $\mathbf{M}^t = \left[e^{-iy_r^\top \omega_u}\right]_{r,u}$.
3: Compute SVD of $\mathbf{M}$, choose rank r according to truncation rule: $\mathbf{M} \approx \mathbf{U}\Sigma\mathbf{W}^{\mathrm{H}}$.
4: Form reduced matrix $\mathbf{R} = \mathbf{U}^{\mathrm{H}}\mathbf{M}^t\mathbf{W}\Sigma^{-1}$.
5: Compute eigenpairs of reduced problem $\mathbf{R}\mathbf{u}_i = \hat{\lambda}_i(t)\mathbf{u}_i$.
6: Transform to original RFF basis: $\mathbf{v}_i = \mathbf{W}\Sigma^{-1}\mathbf{u}_i$, $\hat{\psi}_i(x) = \mathbf{v}_i^{\mathrm{H}}\phi_{\mathrm{RFF}}(x)$.

Fip35

- Small protein, 35 amino acids.
- Gaussian kernel on 600 distances and angles.
- Use $p = 1000$ Fourier features.
- Compute leading 2 eigenvalues and eigenfunctions.
- Transform into membership functions indicating metastable states.



Philipp, Schaller, Boshoff, Peitz, **FN**, Worthmann, arxiv 2402.02494, 2024

Acknowledgments

Main Collaborators: Stefan Klus (Heriot-Watt U, Edinburgh), Frank Noé (Freie U Berlin), Cecilia Clementi (Freie U Berlin), Karl Worthmann (TU Ilmenau), Sebastian Peitz (U Paderborn)

References:

- **FN**, Keller, Pérez-Hernández, Mey, Noé, *Variational Approach to Molecular Kinetics*, JCTC 10 (4), 1739-1752, 2014
- Klus, **FN**, Hamzi, *Kernel-Based Approximation of the Koopman Generator and Schrödinger Operator*, Entropy, 22, 0722, 2020
- **FN** and Klus, *Efficient Approximation of Molecular Kinetics using Random Fourier Features*, Journal of Chemical Physics 159, 074105, 2023
- Philipp, Schaller, Boshoff, Peitz, **FN**, Worthmann, *Extended Dynamic Mode Decomposition: Sharp bounds on the sample efficiency*, arxiv 2402.02494, 2024