

MAX PLANCK INSTITUTE
FOR DYNAMICS OF COMPLEX
TECHNICAL SYSTEMS
MAGDEBURG


DATA-DRIVEN MODELING OF
COMPLEX PHYSICAL SYSTEMS

Dimensionality Reduction and Metastability Analysis using the Koopman Operator

Feliks Nüske

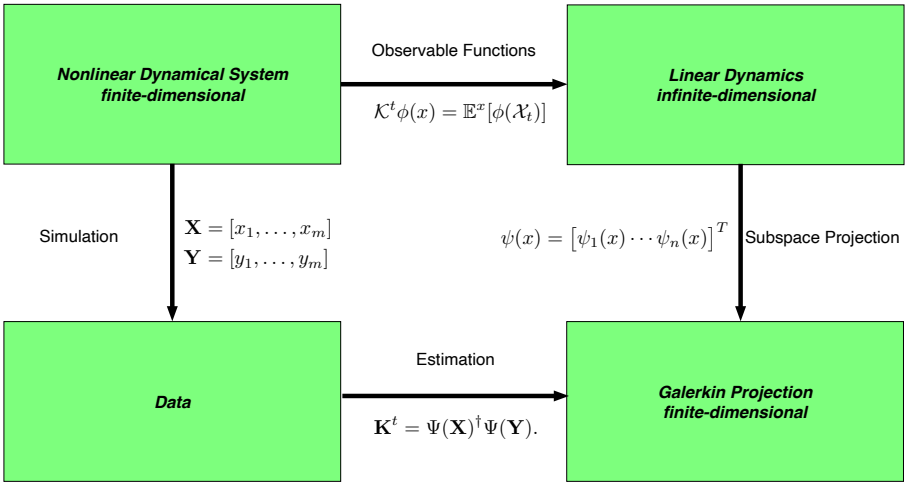
September 7, 2023

Outline

1. The Koopman Operator and EDMD
2. Variational Approach (VAC)
3. EDMD and Kernels
4. Random Fourier Features for EDMD

The Koopman Approach

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



Extended Dynamic Mode Decomposition (EDMD)

Idea: Track the dynamics of a set of **observables** by a **linear map**

Extended Dynamic Mode Decomposition (EDMD)

Idea: Track the dynamics of a set of **observables** by 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 the dynamics of a set of **observables** by 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 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 the dynamics of a set of **observables** by 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 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, *J. Nonlinear Sci.* (2015)

EDMD as a Statistical Method

- Assume x_k are **drawn from distribution** ρ:

$$\frac{1}{m}\left[\Psi(\mathbf{X})^T\Psi(\mathbf{X})\right]_{ij} = \frac{1}{m}\sum_{k=1}^{m}\psi_i(x_k)\psi_j(x_k)$$
$$\xrightarrow{m\to\infty} \mathbb{E}^{\mathcal{X}_0\sim\rho}[\psi_i(\mathcal{X}_0)\psi_j(\mathcal{X}_0)],$$

EDMD as a Statistical Method

- Assume x_k are **drawn from distribution** ρ:

$$\frac{1}{m}\left[\Psi(\mathbf{X})^T\Psi(\mathbf{X})\right]_{ij} = \frac{1}{m}\sum_{k=1}^{m}\psi_i(x_k)\psi_j(x_k) \xrightarrow{m\to\infty} \mathbb{E}^{\mathcal{X}_0\sim\rho}[\psi_i(\mathcal{X}_0)\psi_j(\mathcal{X}_0)],$$

- and similarly

$$\frac{1}{m}\left[\Psi(\mathbf{X})^T\Psi(\mathbf{Y})\right]_{ij} = \frac{1}{m}\sum_{k=1}^{m}\psi_i(x_k)\psi_j(y_k) \xrightarrow{m\to\infty} \mathbb{E}^{\mathcal{X}_0\sim\rho}[\psi_i(\mathcal{X}_0)\psi_j(\mathcal{X}_t)].$$

Williams et al, *J. Nonlinear Sci.* (2015),

EDMD as a Galerkin Method

- Empirical **Gramian** in Hilbert space L^2_ρ:

$$\mathbb{E}^{\mathcal{X}_0\sim\rho}[\psi_i(\mathcal{X}_0)\psi_j(\mathcal{X}_0)] = \int_{\mathbb{X}} \psi_i(x)\,\psi_j(x)\,\mathrm{d}\rho(x) = \langle \psi_i,\, \psi_j \rangle_\rho\,,$$

- Empirical **Stiffness Matrix** in Hilbert space L^2_ρ:

$$\mathbb{E}^{\mathcal{X}_0\sim\rho}[\psi_i(\mathcal{X}_0)\psi_j(\mathcal{X}_t)] = \mathbb{E}^{x\sim\rho}[\psi_i(x)\,\mathbb{E}[\psi_j(\mathcal{X}_t)|\mathcal{X}_0 = x]\,].$$

EDMD as a Galerkin Method

- Empirical **Gramian** in Hilbert space L^2_ρ:

$$\mathbb{E}^{\mathcal{X}_0\sim\rho}[\psi_i(\mathcal{X}_0)\psi_j(\mathcal{X}_0)] = \int_{\mathbb{X}} \psi_i(x)\,\psi_j(x)\,\mathrm{d}\rho(x) = \langle \psi_i,\, \psi_j \rangle_\rho\,,$$

- Empirical **Stiffness Matrix** in Hilbert space L^2_ρ:

$$\mathbb{E}^{\mathcal{X}_0\sim\rho}[\psi_i(\mathcal{X}_0)\psi_j(\mathcal{X}_t)] = \mathbb{E}^{x\sim\rho}[\psi_i(x)\,\mathbb{E}[\psi_j(\mathcal{X}_t)|\mathcal{X}_0 = x]\,].$$

- Using the **Koopman operator**, EDMD becomes a Galerkin method:

$$\mathcal{K}^t\phi(x) = \mathbb{E}[\phi(\mathcal{X}_t)|\mathcal{X}_0 = x]$$
$$\Rightarrow \mathbb{E}^{\mathcal{X}_0\sim\rho}[\psi_i(\mathcal{X}_0)\psi_j(\mathcal{X}_t)] = \int_{\mathbb{X}} \psi_i(x)\,\mathcal{K}^t\psi_j(x)\,\mathrm{d}\rho(x) = \left\langle \psi_i,\, \mathcal{K}^t\psi_j \right\rangle_\rho.$$

Williams et al, *J. Nonlinear Sci.* (2015)

The Koopman Semigroup

- Koopman operators form a **semigroup**:

$$\mathcal{K}^{s+t} = \mathcal{K}^s \mathcal{K}^t$$

- Semigroup property gives rise to **infinitesimal generator**:

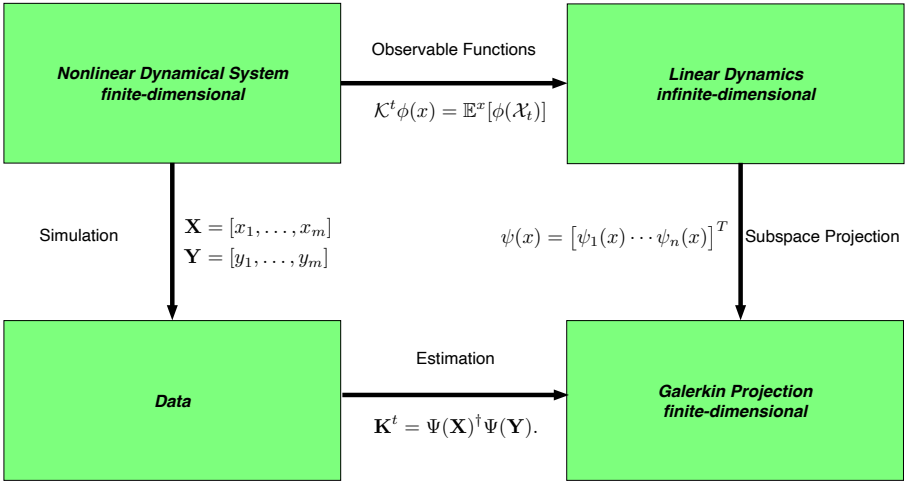
$$\mathcal{L} f = \lim_{t \to 0} \frac{1}{t} (\mathcal{K}^t - \mathrm{Id}) f.$$

- On the space of observables, the dynamics are linear:

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{K}^t \phi = \mathcal{L} \mathcal{K}^t \phi.$$

The Koopman Approach

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



Summary

- Koopman framework uses a **lifting** into an **infinite-dimensional** space where the dynamics become **linear**.
- EDMD approximates a **Galerkin projection** of $\mathcal{K}^t$ in the **statistical limit**.
- EDMD matrix can be used to **forecast** expectations of ob system observables, potentially over **long timescales**.
- Applications in control, model reduction, metastability analysis, ...

Klus, **FN**, et al, *J. Nonlinear Sci.*, 2018,
Mauroy, Suzuki, Mezic (eds), *Koopman operator in systems and control*, Springer 2020,
Klus, **FN**, Peitz et al, *Physica D*, 2020,
Berry, Giannakis, Harlim, *Notices of the AMS*, 2020.

Outline

Variational Principle

- For **reversible** systems, Koopman operator $\mathcal{K}^t$ is **self-adjoint**.

Variational Principle

- For **reversible** systems, Koopman operator $\mathcal{K}^t$ is **self-adjoint**.
- **Rayleigh variational principle** (RVP) for dominant eigenvalues $1 = \lambda_0(t) \geq \ldots \geq \lambda_M(t)$:

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

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

Variational Principle

- For **reversible** systems, Koopman operator $\mathcal{K}^t$ is **self-adjoint**.
- **Rayleigh variational principle** (RVP) for dominant eigenvalues $1 = \lambda_0(t) \geq \ldots \geq \lambda_M(t)$:

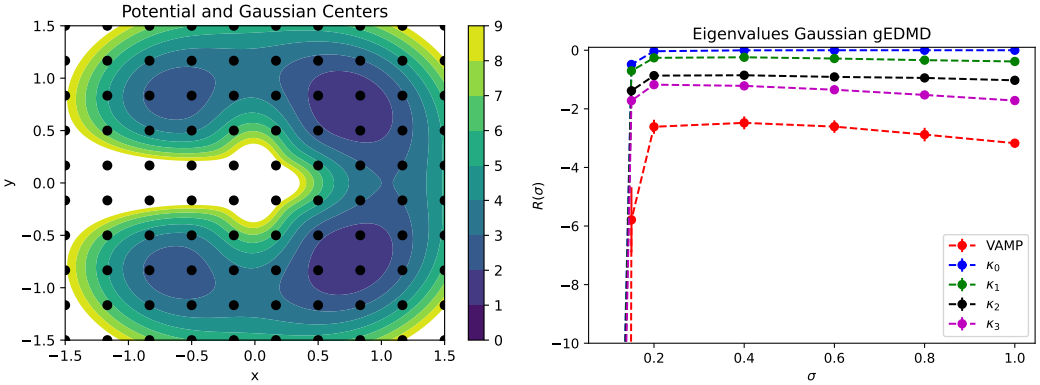
$$\sum_{i=0}^{M} \left\langle \psi_i,\, \mathcal{K}^t \psi_i \right\rangle_\mu =: \mathcal{R}(\psi) \leq \sum_{i=0}^{M} \lambda_i(t) \tag{1}$$

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

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

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

Cross-Validation



Linear VAC: Noé and **FN**, *SIAM MMS*, 2013, **FN**, Keller, et al, *JCTC*, 2014,
TICA: Pérez-Hernández et al, *J. Chem. Phys.*, 2013, **Deep Learning**: Mardt et al, *Nat. Commun.*, 2018
Non-reversible systems: Wu and Noé, *J. Nonlinear Sci.*, 2020

Perspectives

- Learning from **short trajectories** and **renewal equations**,
  Noé, Schütte, et al, *PNAS* (2009)
  Cao, Huang, et al, *J. Chem. Phys.* 2020
- Models for the generator, **system identification**,
  Boninsegna, **FN**, Clementi, *J. Chem. Phys.* (2018)
  Klus, **FN**, Peitz, et al, *Physica D* (2020)
- Use Koopman models to **enhance exploration**,
  Hruska and Clementi, *J. Chem. Phys* (2018)
  Sidky, Chen, Ferguson, *Modern Physics*, 2020
  Shmilovich and Ferguson, *J Phys. Chem. A* (2023)
- **Error theory** based on concentration inequalities,
  Zhang and Zuazua, *Comptes Rendus Math.* (2022)
  FN et al, *J. Nonlinear Sci.* (2023)

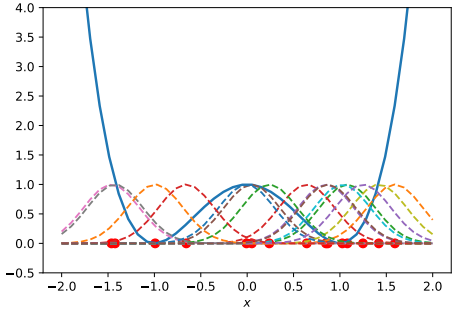
Outline

Data-driven Basis Sets

- Choosing a good basis set is hard.

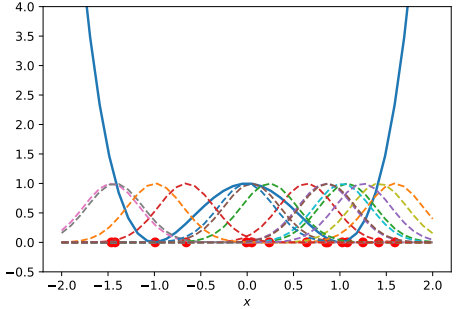
Data-driven Basis Sets

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



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.

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

RKHS

Theorem (Moore-Aronszajn, 1950)

Let $\mathbb{X}$ be a set and $k : \mathbb{X} \times \mathbb{X}$ be a continuous function such that

(*i*) $k(x, y) = k(y, x)$ *(symmetry)*

(*ii*) $\sum_{i,j=1}^{m} \alpha_i \alpha_j k(x_i, x_j) \geq 0$ *(positive definiteness).*

Then k generates a unique Hilbert space $\mathbb{H}$ of continuous functions such that

(*a*) $\mathbb{H} = \overline{\text{span}\{k(x, \cdot),\, x \in \mathbb{X}\}}$,

(*b*) $\langle f,\, k(x, \cdot)\rangle_{\mathbb{H}} = f(x) \quad \forall f \in \mathbb{H}$ *(reproducing property).*

RKHS

Theorem (Moore-Aronszajn, 1950)

Let $\mathbb{X}$ be a set and $k : \mathbb{X} \times \mathbb{X}$ be a continuous function such that

(i) $k(x, y) = k(y, x)$ *(symmetry)*

(ii) $\sum_{i,j=1}^{m} \alpha_i \alpha_j k(x_i, x_j) \geq 0$ *(positive definiteness).*

Then k generates a unique Hilbert space $\mathbb{H}$ of continuous functions such that

(a) $\mathbb{H} = \overline{\text{span}\{k(x, \cdot),\, x \in \mathbb{X}\}}$,

(b) $\langle f,\, k(x, \cdot)\rangle_{\mathbb{H}} = f(x) \quad \forall f \in \mathbb{H}$ *(reproducing property).*

Facts:

- Gaussian RBF kernel is a valid kernel.

RKHS

Theorem (Moore-Aronszajn, 1950)

Let $\mathbb{X}$ be a set and $k : \mathbb{X} \times \mathbb{X}$ be a continuous function such that

(*i*) $k(x, y) = k(y, x)$ *(symmetry)*

(*ii*) $\sum_{i,j=1}^{m} \alpha_i \alpha_j k(x_i, x_j) \geq 0$ *(positive definiteness).*

Then k generates a unique Hilbert space $\mathbb{H}$ of continuous functions such that

(*a*) $\mathbb{H} = \overline{\text{span}\{k(x, \cdot),\, x \in \mathbb{X}\}}$,

(*b*) $\langle f,\, k(x, \cdot)\rangle_{\mathbb{H}} = f(x) \quad \forall f \in \mathbb{H}$ *(reproducing property).*

Facts:

- Gaussian RBF kernel is a valid kernel.
- Many popular kernels correspond to infinite-dimensional RKHSs.

RKHS

Theorem (Moore-Aronszajn, 1950)

Let $\mathbb{X}$ be a set and $k : \mathbb{X} \times \mathbb{X}$ be a continuous function such that

(*i*) $k(x, y) = k(y, x)$ *(symmetry)*

(*ii*) $\sum_{i,j=1}^{m} \alpha_i \alpha_j k(x_i, x_j) \geq 0$ *(positive definiteness).*

Then k generates a unique Hilbert space $\mathbb{H}$ of continuous functions such that

(*a*) $\mathbb{H} = \overline{\operatorname{span}\{k(x, \cdot),\, x \in \mathbb{X}\}}$,

(*b*) $\langle f,\, k(x, \cdot)\rangle_{\mathbb{H}} = f(x) \quad \forall f \in \mathbb{H}$ *(reproducing property).*

Facts:

- Gaussian RBF kernel is a valid kernel.
- Many popular kernels correspond to infinite-dimensional RKHSs.
- These RKHSs are dense in most L^2-spaces.

EDMD and RKHS

- Applying EDMD to kernel-based basis functions $k(x_i, \cdot)$ leads to matrices...

$$\mathbf{K}_{\mathbb{H}}^t = \mathbf{K}_X^{-1}\mathbf{K}_X^t, \qquad \mathbf{K}_X(r,s) = k(x_r, x_s), \qquad \mathbf{K}_X^t(r,s) = k(y_r, x_s).$$

- ... or to the generalized eigenvalue problem:

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

Klus et al, *J. Nonlinear Sci.*, 2020; Klus, **FN**, and Hamzi, *Entropy*, 2020

EDMD and RKHS

- Applying EDMD to kernel-based basis functions $k(x_i, \cdot)$ leads to matrices...

$$\mathbf{K}_{\mathbb{H}}^t = \mathbf{K}_X^{-1}\mathbf{K}_X^t, \qquad \mathbf{K}_X(r,s) = k(x_r, x_s), \qquad \mathbf{K}_X^t(r,s) = k(y_r, x_s).$$

- ... or to the generalized eigenvalue problem:

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

- For infinite data, converges to cross-covariance operators:

$$\mathcal{K}_X f = \int_{\mathbb{X}} f(x)k(x,\cdot)\,\mathrm{d}\mu(x), \qquad \mathcal{K}_X^t f = \int_{\mathbb{X}} \mathcal{K}^t f(x)k(x,\cdot)\,\mathrm{d}\mu(x).$$

Klus et al, *J. Nonlinear Sci.*, 2020; Klus, **FN**, and Hamzi, *Entropy*, 2020

Outline

Motivation

- Recall kernelized eigenvalue problem:

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

- Kernel matrices are often low-rank; require **low-rank decomposition** of $\mathbf{K}_X$.
- Straightforward approach scales as $\mathcal{O}(m^3)$.
- Needs to be solved **many times** to accurately tune kernel parameters.
- Requires efficient low-rank approximation, e.g. Nyström methods, random linear algebra, random Fourier features,...

Intro RFF

- A translation invariant kernel with $k(x, x) = 1$ can be written as

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

  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.

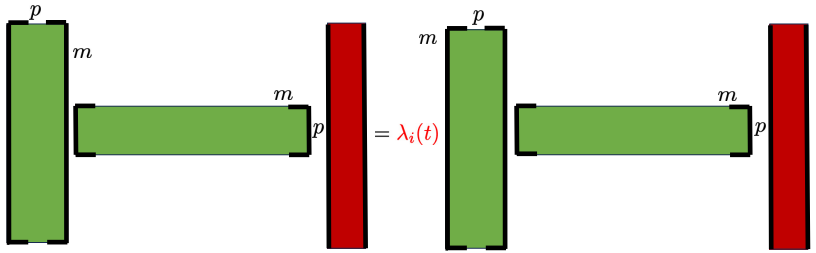
Rahimi and Recht, Advances in Neural Information Processing Systems, 2007

Low-Rank Kernel GEV

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

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

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



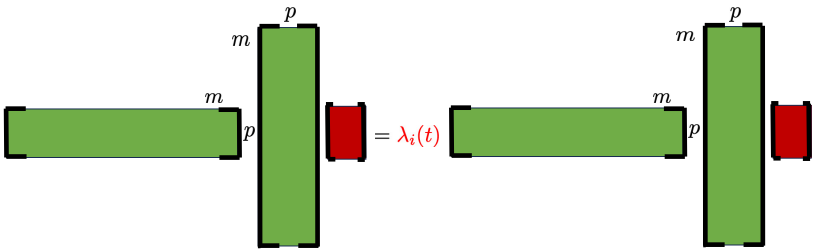
FN and Klus, J. Chem. Phys., 2023

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, J. Chem. Phys., 2023

Low-Rank Kernel GEV

- The dual eigenvalue problem

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

  can be solved by operating only on $\mathbf{M}$, $\mathbf{M}^t$.
- If p is significantly smaller than m, cost is $\mathcal{O}(mp^2)$ instead of $\mathcal{O}(m^3)$.
- Dual GEV is the EDMD eigenvalue problem for a **randomized, complex-valued basis set** of plane waves:

$$\phi_{\mathrm{RFF}}(x) := \begin{bmatrix} \phi_1(x) \\ \vdots \\ \phi_p(x) \end{bmatrix} = \begin{bmatrix} e^{ix^\top \omega_1} \\ \vdots \\ e^{ix^\top \omega_p} \end{bmatrix}.$$

FN and Klus, J. Chem. Phys., 2023

NTL9

Small protein with 39 amino acids, use Gaussian kernel on 666 inter-residue distances, $m = 15.000$ data points in real space.

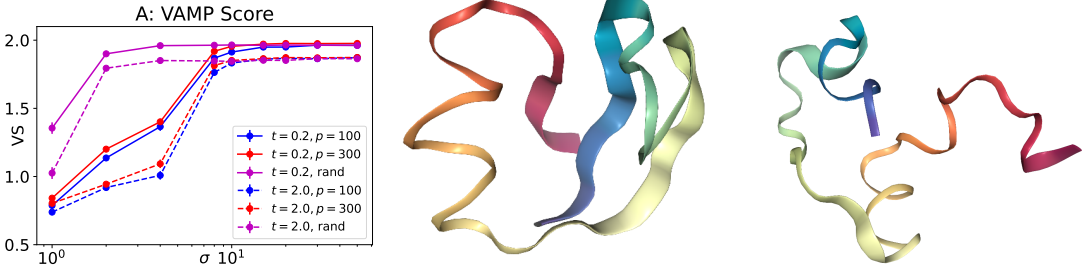


Figure: VAMP Score and Representatives of PCCA States for NTL9

FN and Klus, J. Chem. Phys., 2023

Conclusions

1. Koopman framework provides a linear view of dynamical systems, Galerkin projections can be learned consistently from simulation data.
2. Variational Score (VAMP) is a principled way of choosing model classes, especially for metastable systems.
3. Kernel Methods provide a powerful data-driven basis set with well-understood asymptotic properties.
4. Random Fourier features appear as a promising candidate for efficient low-rank approx. of resulting linear algebra problems.

Acknowledgments

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

References:

- **FN**, Keller, Pérez-Hernández, Mey, Noé, *Variational Approach to Molecular Kinetics*, JCTC 10 (4), 1739-1752, 2014
- Klus, **FN**, Koltai, Wu, Kevrekidis, Schütte, Noé, *Data-driven model reduction and transfer operator approximation*, Journal of Nonlinear Science, 28, 985-1010, 2018
- Klus, **FN**, Peitz, et al, *Data-driven approximation of the Koopman generator: Model reduction, system identification, and control*, Physica D, 406, 132416, 2020
- Klus, **FN**, Hamzi, *Kernel-Based Approximation of the Koopman Generator and Schrödinger Operator*, Entropy 22 (7), 722, 2020
- **FN** and Klus, *Efficient Approximation of Molecular Kinetics using Random Fourier Features*, Journal of Chemical Physics 159, 074105, 2023