MAX PLANCK INSTITUTE
FOR DYNAMICS OF COMPLEX
TECHNICAL SYSTEMS
MAGDEBURG

1998
2023

DATA-DRIVEN MODELING OF
COMPLEX PHYSICAL SYSTEMS

Kernel-based Approximation of the Koopman Operator and Generator

Feliks Nüske

April 5, 2023

Outline

1. Koopman Operators and Generators

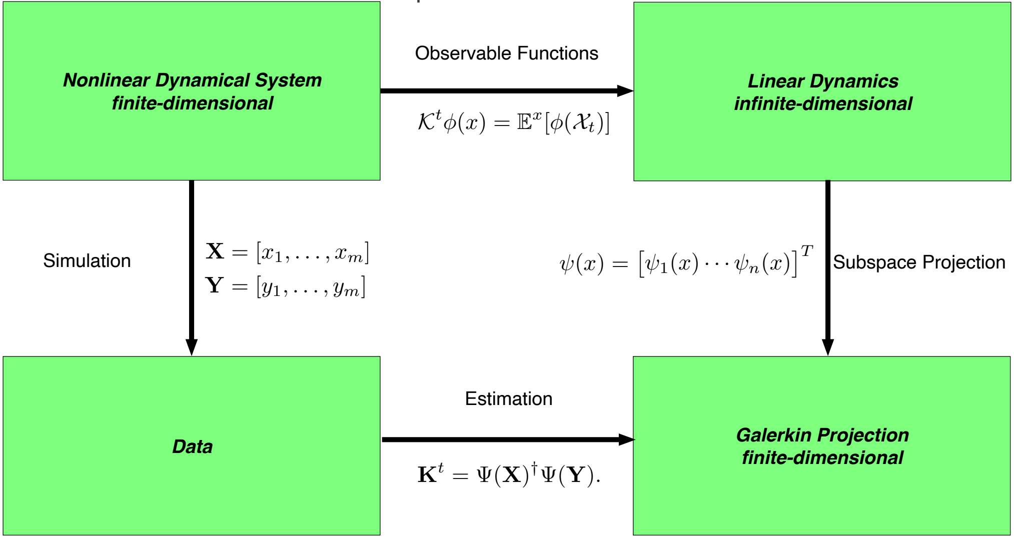
2. EDMD and generator EDMD

3. Representation by Reproduing Kernels

Koopman Theory: Overview

Koopman theory:

lifting into infinite-dimensional feature space + statistical estimation.



Koopman Semigroup

- Dynamical system $\mathcal{X}_t$, continuous in time and space, state space $\mathbb{X}$, invariant measure μ.

Koopman Semigroup

- Dynamical system $\mathcal{X}_t$, continuous in time and space, state space $\mathbb{X}$, invariant measure μ.
- **Koopman operator** for $f \in L^2_\mu(\mathbb{X})$:

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

Koopman Semigroup

- Dynamical system $\mathcal{X}_t$, continuous in time and space, state space $\mathbb{X}$, invariant measure μ.
- **Koopman operator** for $f \in L^2_\mu(\mathbb{X})$:

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

- Semigroup of bounded linear operators: $\mathcal{K}^{t_1+t_2} = \mathcal{K}^{t_1}\mathcal{K}^{t_2}$.

Koopman Semigroup

- Dynamical system $\mathcal{X}_t$, continuous in time and space, state space $\mathbb{X}$, invariant measure μ.
- **Koopman operator** for $f \in L^2_\mu(\mathbb{X})$:

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

- Semigroup of bounded linear operators: $\mathcal{K}^{t_1+t_2} = \mathcal{K}^{t_1}\mathcal{K}^{t_2}$.
- With the **infinitesimal generator**:

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

Koopman Semigroup

- Dynamical system $\mathcal{X}_t$, continuous in time and space, state space $\mathbb{X}$, invariant measure μ.
- **Koopman operator** for $f \in L^2_\mu(\mathbb{X})$:

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

- Semigroup of bounded linear operators: $\mathcal{K}^{t_1+t_2} = \mathcal{K}^{t_1}\mathcal{K}^{t_2}$.
- With the **infinitesimal generator**:

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

- ... we have **linear dynamics** on the domain of $\mathcal{L}$:

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

Generator for SDEs

- For **Stochastic Differential Equations**, $\mathcal{L}$ is known from Ito calculus:

$$d\mathcal{X}_t = b(\mathcal{X}_t)\,\mathrm{d}t + \sigma(\mathcal{X}_t)\,\mathrm{d}W_t, \qquad \Rightarrow \mathcal{L} = b \cdot \nabla + \frac{1}{2}(\sigma\sigma^T) : \nabla^2.$$

- Example: **Overdamped Langevin Dynamics**:

$$d\mathcal{X}_t = -\nabla V(\mathcal{X}_t)\,\mathrm{d}t + \sqrt{2k_B T}\,\mathrm{d}W_t, \qquad \Rightarrow \mathcal{L} = -\nabla V \cdot \nabla + k_B T \Delta.$$

Outline

Extended Dynamic Mode Decomposition (EDMD)

- Choose a finite-dimensional space of **observables** $\psi(x) = \left[\psi_1(x) \cdots \psi_n(x)\right]^T$.

Extended Dynamic Mode Decomposition (EDMD)

- Choose a finite-dimensional space of **observables** $\psi(x) = \left[\psi_1(x) \cdots \psi_n(x)\right]^T$.
- For snapshot matrices $\mathbf{X}$ and $\mathbf{Y}$ ($\mathbf{Y}$ equals $\mathbf{X}$ shifted by t), generate transformed matrices:

$$\Psi(\mathbf{X}) = \left[\psi(x_1) \;\middle|\; \cdots \;\middle|\; \psi(x_m)\right] \in \mathbb{R}^{n \times m},$$
$$\Psi(\mathbf{Y}) = \left[\psi(y_1) \;\middle|\; \cdots \;\middle|\; \psi(y_m)\right] \in \mathbb{R}^{n \times m}.$$

Extended Dynamic Mode Decomposition (EDMD)

- Choose a finite-dimensional space of **observables** $\psi(x) = \left[\psi_1(x) \cdots \psi_n(x)\right]^T$.
- For snapshot matrices $\mathbf{X}$ and $\mathbf{Y}$ ($\mathbf{Y}$ equals $\mathbf{X}$ shifted by t), generate transformed matrices:

$$\Psi(\mathbf{X}) = \left[\psi(x_1) \;\middle|\; \cdots \;\middle|\; \psi(x_m)\right] \in \mathbb{R}^{n \times m},$$
$$\Psi(\mathbf{Y}) = \left[\psi(y_1) \;\middle|\; \cdots \;\middle|\; \psi(y_m)\right] \in \mathbb{R}^{n \times m}.$$

- Solve regression problem:

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

Extended Dynamic Mode Decomposition (EDMD)

- Choose a finite-dimensional space of **observables** $\psi(x) = \left[\psi_1(x) \cdots \psi_n(x)\right]^T$.
- For snapshot matrices $\mathbf{X}$ and $\mathbf{Y}$ ($\mathbf{Y}$ equals $\mathbf{X}$ shifted by t), generate transformed matrices:

$$\Psi(\mathbf{X}) = \left[\psi(x_1) \;\middle|\; \cdots \;\middle|\; \psi(x_m)\right] \in \mathbb{R}^{n \times m},$$
$$\Psi(\mathbf{Y}) = \left[\psi(y_1) \;\middle|\; \cdots \;\middle|\; \psi(y_m)\right] \in \mathbb{R}^{n \times m}.$$

- Solve regression problem:

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

- Consistently approximates Galerkin projection of $\mathcal{K}^t$ for infinite data:

$$\hat{\mathbf{K}}^t \to \mathbf{K}^t = \left[\langle \psi_i, \psi_j \rangle_\mu\right]^{-1} \left[\langle \psi_i, \mathcal{K}^t \psi_j \rangle_\mu\right] =: (\mathbf{C}^0)^{-1} \mathbf{C}^t.$$

Williams et al, *J. Nonlinear Sci.* (2015), Klus, **N.** et al, *J. Nonlinear Sci.* (2018)

Generator EDMD

- Same approach for generator: form matrices $\Psi(\mathbf{X})$ and $\mathcal{L}\Psi(\mathbf{X})$:

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

Generator EDMD

- Same approach for generator: form matrices $\Psi(\mathbf{X})$ and $\mathcal{L}\Psi(\mathbf{X})$:

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

- Solve regression problem:

$$\hat{\mathbf{L}}^{\mathbb{V}} = (\Psi(\mathbf{X})\Psi(\mathbf{X})^T)^{\dagger}((\Psi(\mathbf{X})\mathcal{L}\Psi(\mathbf{X})^T)$$
$$\rightarrow \left[\langle \psi_i, \psi_j \rangle_\mu\right]^{-1} \left[\langle \psi_i, \mathcal{L}\psi_j \rangle_\mu\right] =: (\mathbf{C}^0)^{-1}\mathbf{C}^{\mathcal{L}}.$$

Generator EDMD

- Same approach for generator: form matrices $\Psi(\mathbf{X})$ and $\mathcal{L}\Psi(\mathbf{X})$:

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

- Solve regression problem:

$$\hat{\mathbf{L}}^{\mathbb{V}} = (\Psi(\mathbf{X})\Psi(\mathbf{X})^T)^\dagger((\Psi(\mathbf{X})\mathcal{L}\Psi(\mathbf{X})^T)$$
$$\to \left[\langle \psi_i, \psi_j \rangle_\mu\right]^{-1} \left[\langle \psi_i, \mathcal{L}\psi_j \rangle_\mu\right] =: (\mathbf{C}^0)^{-1}\mathbf{C}^{\mathcal{L}}.$$

Klus, **N.** et al, *Physica D* (2020)

Outline

Model Class

- EDMD approximation depends critically on the selected observables. Choosing the basis sets often requires expert knowledge.
- Popular linear model classes:
  - Indicator functions of a set partitioning (Markov State Models, MSM) [Prinz et al, *J. Chem. Phys.* (2011)]
  - Elementary Descriptors (TICA) [Pérez-Hernández et al, *J. Chem. Phys.* (2013)]
  - Gaussian Basis Sets [**N.**, Keller, et al, *J. Chem. Theory Comput.* (2014)]
- Non-linear model classes:
  - Neural networks [Mardt et al, *Nat. Commun.* (2018)].
  - Tensor Format [**N.**, Gelß, et al, *Physica D* (2021)]
  - Reproducing Kernels (**this talk**).

Reproducing Kernels

Definition

A Hilbert space of continuous functions on $\mathbb{X}$ is a **reproducing kernel Hilbert space** if a function $k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ exists such that

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

(ii) $\langle \phi,\, k(x, \cdot) \rangle_{\mathbb{H}} = \phi(x)$ for all $\phi \in \mathbb{H}$ (reproducing property).

Reproducing Kernels

Definition

A Hilbert space of continuous functions on $\mathbb{X}$ is a **reproducing kernel Hilbert space** if a function $k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ exists such that

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

(ii) $\langle \phi,\, k(x, \cdot)\rangle_{\mathbb{H}} = \phi(x)$ for all $\phi \in \mathbb{H}$ (reproducing property).

- The symmetric, positive-definite function k is called the **kernel**, the map $\Phi : x \mapsto \Phi(x) := k(x, \cdot)$ is called the **feature map** associated the $\mathbb{H}$.

Reproducing Kernels

Definition

A Hilbert space of continuous functions on $\mathbb{X}$ is a **reproducing kernel Hilbert space** if a function $k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ exists such that

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

(ii) $\langle \phi,\, k(x,\cdot)\rangle_{\mathbb{H}} = \phi(x)$ for all $\phi \in \mathbb{H}$ (reproducing property).

- The symmetric, positive-definite function k is called the **kernel**, the map $\Phi : x \mapsto \Phi(x) := k(x,\cdot)$ is called the **feature map** associated the $\mathbb{H}$.
- RKHS can be defined by specifying the kernel. Many kernels generate infinite-dimensional spaces which are dense in L^p_μ under mild conditions.

Reproducing Kernels

Definition

A Hilbert space of continuous functions on $\mathbb{X}$ is a **reproducing kernel Hilbert space** if a function $k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ exists such that

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

(ii) $\langle \phi,\, k(x, \cdot) \rangle_{\mathbb{H}} = \phi(x)$ for all $\phi \in \mathbb{H}$ (reproducing property).

- The symmetric, positive-definite function k is called the **kernel**, the map $\Phi : x \mapsto \Phi(x) := k(x, \cdot)$ is called the **feature map** associated the $\mathbb{H}$.
- RKHS can be defined by specifying the kernel. Many kernels generate infinite-dimensional spaces which are dense in L^p_μ under mild conditions.
- Most popular kernel is the Gaussian RBF kernel

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

Steinwart, Christmann, *Support Vector Machines*, Springer (2008)

(g)EDMD with Kernel

- Mass and stiffness matrices are limits of averaged rank-one operators on $\mathbb{R}^n$:

$$\mathbf{C}^0 = \lim_{m\to\infty} \frac{1}{m} \sum_{k=0}^{m} \psi(x_k) \otimes \psi(x_k), \qquad \mathbf{C}^t = \lim_{m\to\infty} \frac{1}{m} \sum_{k=0}^{m} \psi(x_k) \otimes \psi(y_k).$$

(g)EDMD with Kernel

- Mass and stiffness matrices are limits of averaged rank-one operators on $\mathbb{R}^n$:

$$\mathbf{C}^0 = \lim_{m\to\infty} \frac{1}{m}\sum_{k=0}^{m} \psi(x_k)\otimes\psi(x_k), \qquad \mathbf{C}^t = \lim_{m\to\infty} \frac{1}{m}\sum_{k=0}^{m} \psi(x_k)\otimes\psi(y_k).$$

- Using the feature map, we can define similar operators on $\mathbb{H}$:

$$\mathcal{C}^0 = \int \Phi(x)\otimes\Phi(x)\,\mathrm{d}\mu(x) = \lim_{m\to\infty} \frac{1}{m}\sum_{k=0}^{m} \Phi(x_k)\otimes\Phi(x_k),$$

$$\mathcal{C}^t = \int \Phi(x)\otimes\Phi(y)\,\mathrm{d}\mu_t(x,y) = \lim_{m\to\infty} \frac{1}{m}\sum_{k=0}^{m} \Phi(x_k)\otimes\Phi(y_k),$$

$$\mathcal{C}^{\mathcal{L}} = \int \Phi(x)\otimes\mathcal{L}\Phi(x)\,\mathrm{d}\mu(x) = \lim_{m\to\infty} \frac{1}{m}\sum_{k=0}^{m} \Phi(x_k)\otimes\mathcal{L}\Phi(x_k).$$

Klus et al, *J. Nonlinear Sci.* (2018), Klus, **N.** et al, *Entropy* (2020)

Analysis

- Under mild assumptions, these are Hilbert-Schmidt operators and emulate μ-inner products on $\mathbb{H}$:

$$\langle \phi,\, \psi \rangle_\mu = \langle \phi,\, \mathcal{C}^0 \psi \rangle_{\mathbb{H}}\,, \qquad \langle \phi,\, \mathcal{K}^t \psi \rangle_\mu = \langle \phi,\, \mathcal{C}^t \psi \rangle_{\mathbb{H}}\,, \qquad \langle \phi,\, \mathcal{L} \psi \rangle_\mu = \langle \phi,\, \mathcal{C}^{\mathcal{L}} \psi \rangle_{\mathbb{H}}\,.$$

Analysis

- Under mild assumptions, these are Hilbert-Schmidt operators and emulate μ-inner products on $\mathbb{H}$:

$$\langle \phi,\, \psi \rangle_\mu = \langle \phi,\, \mathcal{C}^0 \psi \rangle_{\mathbb{H}}\,, \qquad \langle \phi,\, \mathcal{K}^t \psi \rangle_\mu = \langle \phi,\, \mathcal{C}^t \psi \rangle_{\mathbb{H}}\,, \qquad \langle \phi,\, \mathcal{L} \psi \rangle_\mu = \langle \phi,\, \mathcal{C}^{\mathcal{L}} \psi \rangle_{\mathbb{H}}\,.$$

- If $\mathbb{H}$ is invariant under $\mathcal{K}^t$, then the operator $(\mathcal{C}^0)^{-1}\mathcal{C}^t$ is well-defined and bounded on $\mathbb{H}$. It can therefore be used for metastability analysis etc.

Klebanov et al, *SIAM J. Math. Data Sci.* (2021),
Phillip, Schaller, Worthmann, Peitz, **N.**, *arxiv* 2301.08637 (2023)

Analysis

- Under mild assumptions, these are Hilbert-Schmidt operators and emulate μ-inner products on $\mathbb{H}$:

$$\langle \phi, \psi \rangle_\mu = \langle \phi, \mathcal{C}^0 \psi \rangle_{\mathbb{H}}, \qquad \langle \phi, \mathcal{K}^t \psi \rangle_\mu = \langle \phi, \mathcal{C}^t \psi \rangle_{\mathbb{H}}, \qquad \langle \phi, \mathcal{L} \psi \rangle_\mu = \langle \phi, \mathcal{C}^{\mathcal{L}} \psi \rangle_{\mathbb{H}}.$$

- If $\mathbb{H}$ is invariant under $\mathcal{K}^t$, then the operator $(\mathcal{C}^0)^{-1}\mathcal{C}^t$ is well-defined and bounded on $\mathbb{H}$. It can therefore be used for metastability analysis etc.
- Finite-data approximations require only kernel evalutations at the data sites. Promising results to small MD systems can be obtained.

Klebanov et al, *SIAM J. Math. Data Sci.* (2021),
Phillip, Schaller, Worthmann, Peitz, **N.**, *arxiv* 2301.08637 (2023)
Klus, **N.** et al, *Entropy* (2020)

Error Estimates

Theorem

Assume the spectrum of $\mathcal{L}$ is discrete. Setting $z_k = (x_k, y_k)$, $k = 1, \ldots, m$, the Hilbert-Schmidt variance of the empirical estimator can be written as

$$\mathbb{E}\left[\|\hat{\mathcal{C}}^t - \mathcal{C}^t\|_{HS}^2\right] = \frac{1}{m}\left[\mathbb{E}_0(t) + 2\sum_{j=1}^{\infty}\frac{d_{j,t}q_j}{1-q_j}\left(1 - \frac{1}{m}\cdot\frac{1-q_j^m}{1-q_j}\right)\right], \tag{1}$$

with

$$q_j = e^{\mu_j \Delta t}, \quad d_{j,t} = \langle c_{j,t}, \psi_j \rangle_\mu, \quad \textit{and} \quad c_{j,t}(x) = \left\langle \Phi_{t,x}, \tilde{\psi}_j \right\rangle_\mu.$$

Phillip, Schaller, Worthmann, Peitz, **N.**, *arxiv* 2301.08637 (2023)

Error Estimates

Theorem

Assume the spectrum of $\mathcal{L}$ is discrete. Setting $z_k = (x_k, y_k)$, $k = 1, \ldots, m$, the Hilbert-Schmidt variance of the empirical estimator can be written as

$$\mathbb{E}\left[\|\hat{\mathcal{C}}^t - \mathcal{C}^t\|_{HS}^2\right] = \frac{1}{m}\left[\mathbb{E}_0(t) + 2\sum_{j=1}^{\infty} \frac{d_{j,t} q_j}{1 - q_j}\left(1 - \frac{1}{m} \cdot \frac{1 - q_j^m}{1 - q_j}\right)\right], \quad (1)$$

with

$$q_j = e^{\mu_j \Delta t}, \quad d_{j,t} = \langle c_{j,t}, \psi_j \rangle_\mu, \quad \text{and} \quad c_{j,t}(x) = \left\langle \Phi_{t,x}, \tilde{\psi}_j \right\rangle_\mu.$$

- First term is the i.i.d. variance, the second part yields an asymptotic variance for $m \to \infty$.

Phillip, Schaller, Worthmann, Peitz, **N.**, *arxiv* 2301.08637 (2023)

Error Estimates

Theorem

Assume the spectrum of $\mathcal{L}$ is discrete. Setting $z_k = (x_k, y_k)$, $k = 1, \dots, m$, the Hilbert-Schmidt variance of the empirical estimator can be written as

$$\mathbb{E}\left[\|\hat{\mathcal{C}}^t - \mathcal{C}^t\|_{HS}^2\right] = \frac{1}{m}\left[\mathbb{E}_0(t) + 2\sum_{j=1}^{\infty} \frac{d_{j,t} q_j}{1 - q_j}\left(1 - \frac{1}{m} \cdot \frac{1 - q_j^m}{1 - q_j}\right)\right], \tag{1}$$

with

$$q_j = e^{\mu_j \Delta t}, \quad d_{j,t} = \langle c_{j,t}, \psi_j \rangle_\mu, \quad \textit{and} \quad c_{j,t}(x) = \left\langle \Phi_{t,x}, \tilde{\psi}_j \right\rangle_\mu .$$

- First term is the i.i.d. variance, the second part yields an asymptotic variance for $m \to \infty$.
- Can be combined with concentration inequalities to obtain probabilistic bounds for $\hat{\mathcal{C}}^0, \hat{\mathcal{C}}^t$ and $(\hat{\mathcal{C}}^0)^{-1}\hat{\mathcal{C}}^t$.

Phillip, Schaller, Worthmann, Peitz, **N.**, *arxiv* 2301.08637 (2023)

Thanks

Thank you for your attention!

Joint work with: Stefan Klus (U Surrey), Sebastian Peitz (UPB), Karl Worthmann, Manuel Schaller, Friedrich Phillip (TU Ilmenau)

Main Papers:

- Klus, **Nüske**, Peitz, et al, Data-driven approximation of the Koopman generator: Model reduction, system identification, and control, *Physica D: Nonlinear Phenomena*, 406, 132416 (2020)
- Klus, **Nüske**, Hamzi, Kernel-Based Approximation of the Koopman Generator and Schrödinger Operator, *Entropy* 23(2), 134 (2020)
- Phillip, Schaller, Worthmann, Peitz, **Nüske**, Error bounds for kernel-based approximations of the Koopman operator, *arxiv* 2301.08637 (2023)