The Distributional Koopman Operator

SIAM DS25, Data Driven Modeling of Dynamical Systems via the Transfer Operator

Maria Oprea, May 12th 2025

Joint work with

Yunan Yang

Goenka Family Assistant Professor in Mathematics at Cornell University

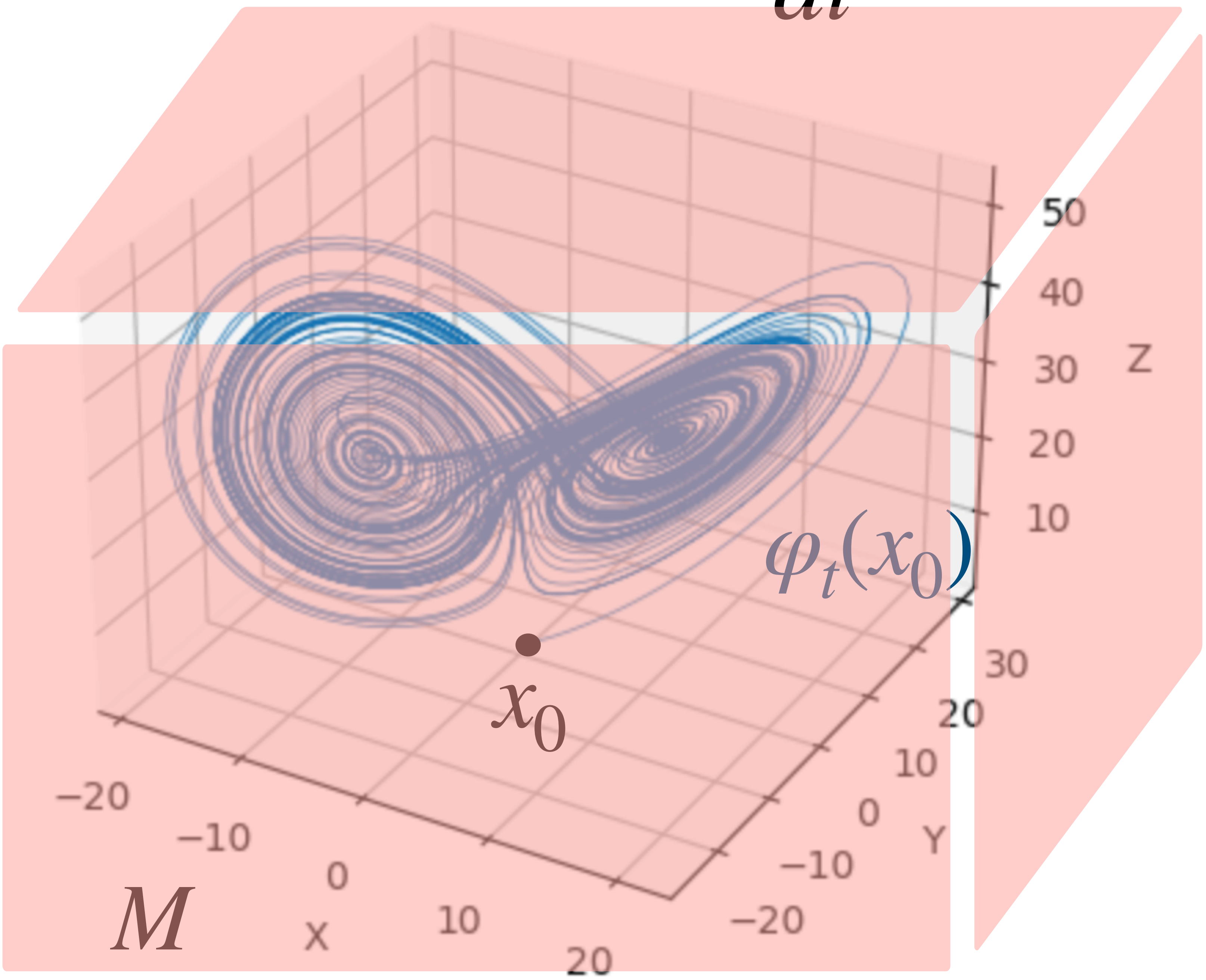
Alex Townsend

Associate Professor at Cornell University

The Koopman operator K_t

Dynamics on M:

$$\varphi : \mathbb{R} \times M \to M, \frac{d}{dt}\varphi_t(x) = f(x)$$



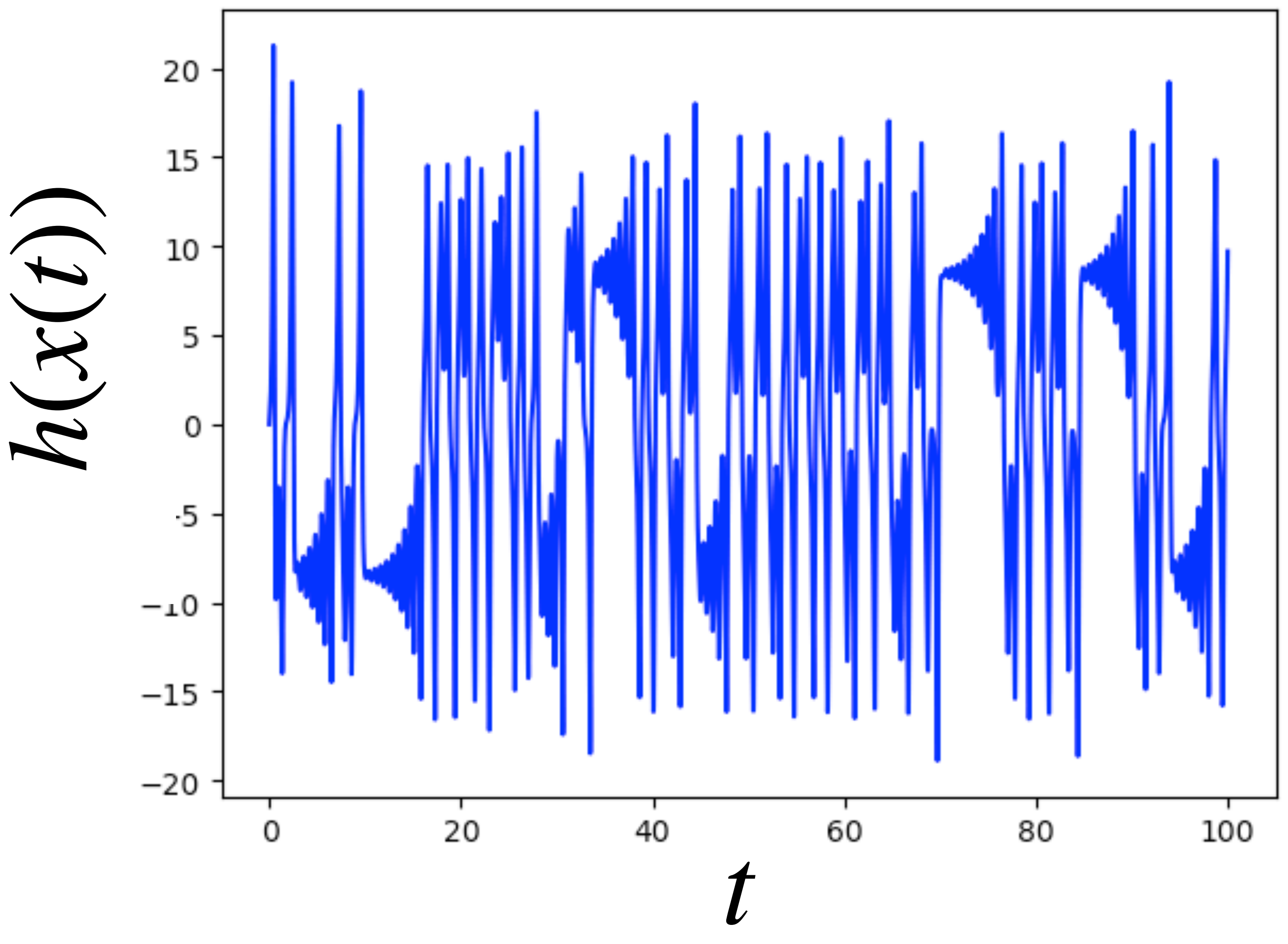
Measurement

$$\hat{h} : M \to \mathbb{R}$$

$$\hat{h} \in L^\infty(M)$$

Dynamics of observables:

$$K_t : L^\infty(M) \to L^\infty(M), K_t\hat{h} = \hat{h} \circ \varphi_t$$



Introducing randomness

$(\Omega, \mathscr{F}, p)$

$\theta_t : \Omega \to \Omega, \theta_t \# p = p$

Source of randomness

$\longrightarrow$

M

$\Phi_t(\omega, \cdot\,)$

State

Random dynamical systems: $\Phi : \mathbb{R} \times \Omega \times M \to M$ with cocycle propriety

$$\Phi_{t+s}(\omega, x) = \Phi_t(\theta_s \omega, x) \circ \Phi_s(\omega, x) \ \forall x, \omega$$

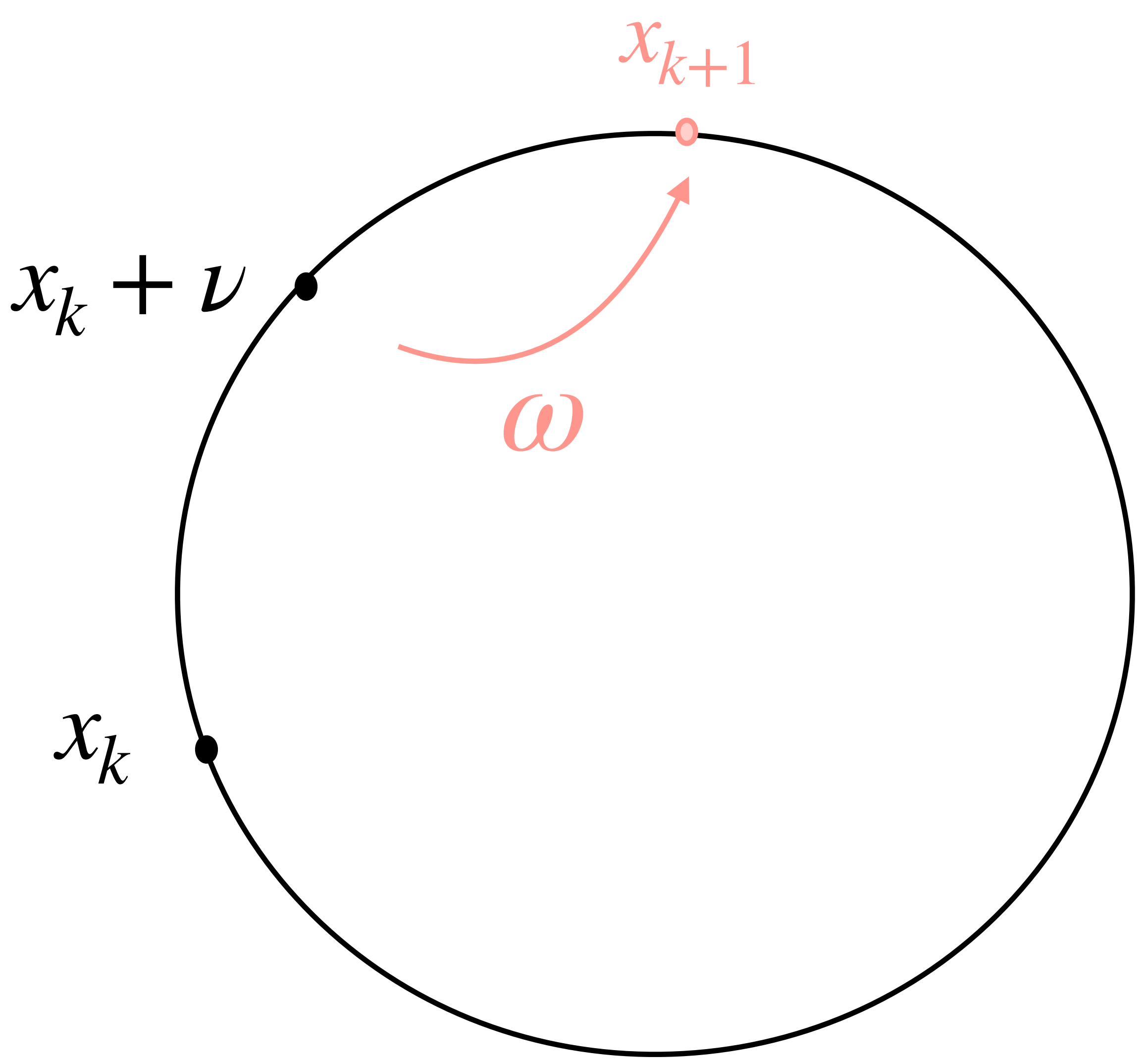
Example: Random rotations on a circle

$\omega = (\omega_0, \omega_1, \ldots)$

p = Bernoulli measure

$\theta(\omega_0, \omega_1, \ldots) = (\omega_1, \ldots)$

$\Phi_1(\omega, x) = x + \nu + \omega_0 \bmod 2\pi$



Introducing randomness

Random dynamical systems: $\Phi : \mathbb{R} \times \Omega \times M \to M$ with cocycle propriety

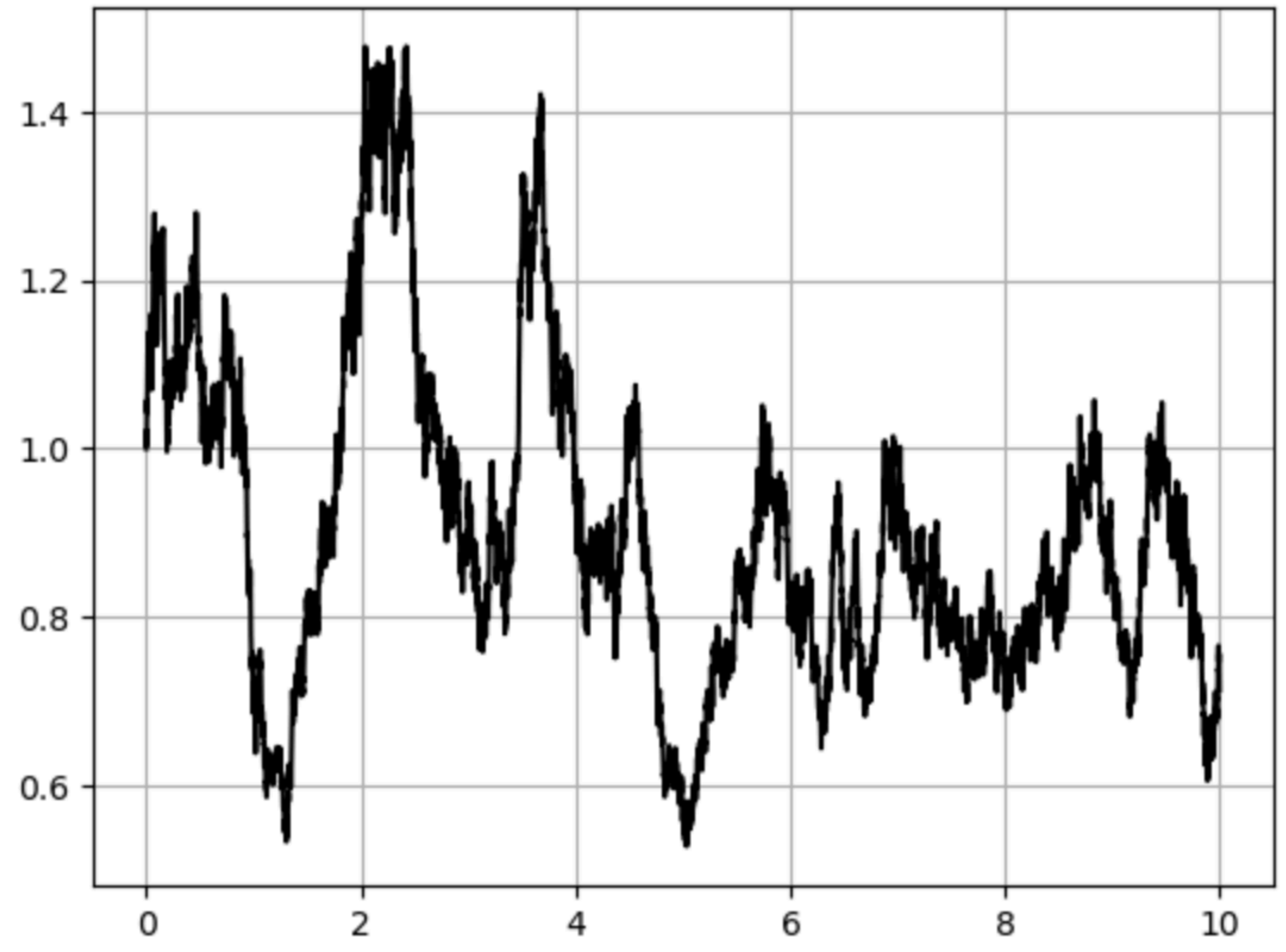
$$\Phi_{t+s}(\omega, x) = \Phi_t(\theta_s \omega, x) \circ \Phi_s(\omega, x) \ \forall x, \omega$$

Example: SDEs $dX_t = a(X_t)dt + b(X_t)d\omega_t$

$\Omega = \mathscr{C}[0,1]$ continuous functions

$\theta_t \omega(s) = \omega(t+s) - \omega(s)$ shift

Wiener measure with σ-algebra generated by cylinder sets[1]

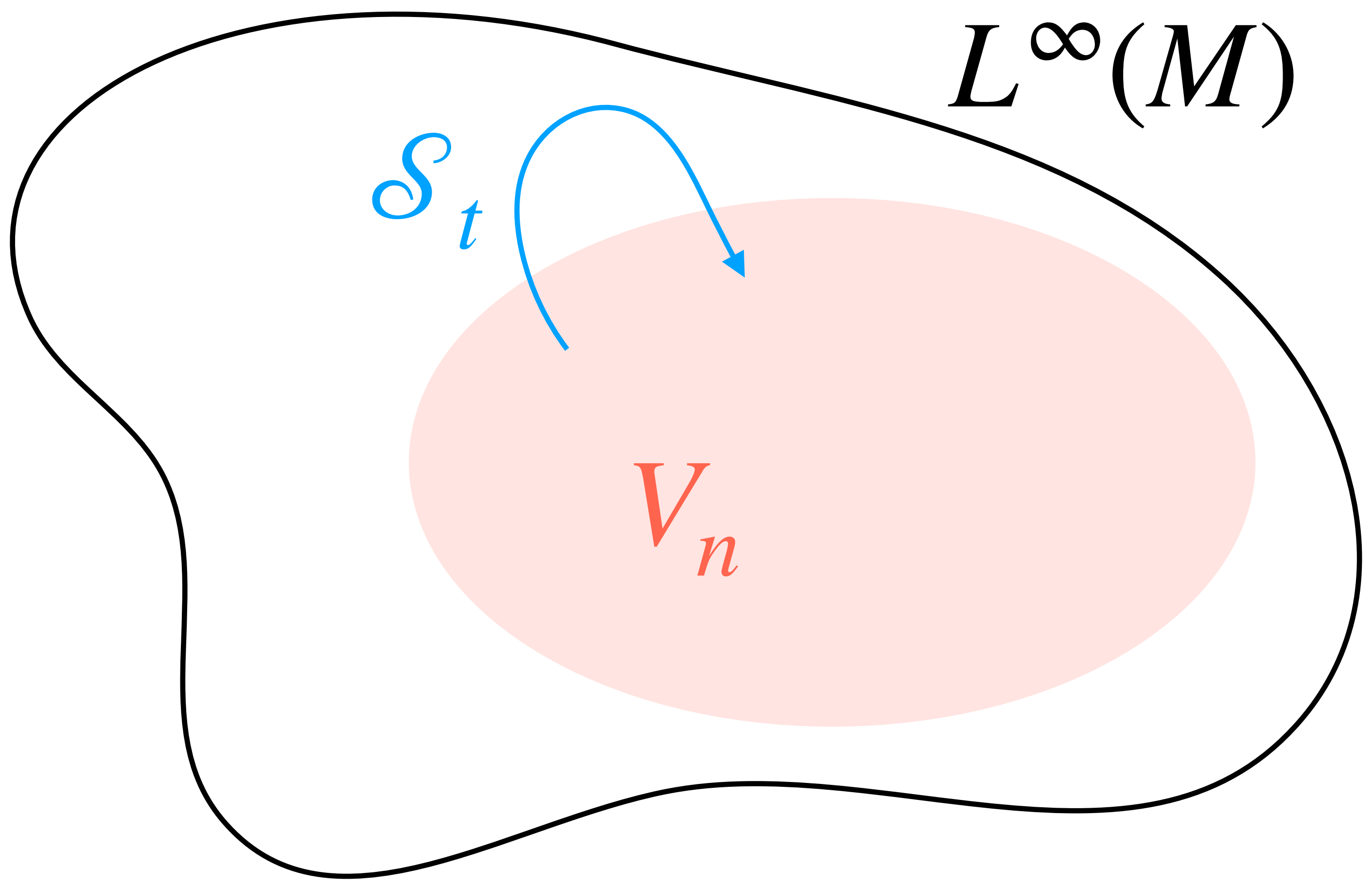


[1] P. Billingsley, Convergence of Probability Measures, A Wiley-Interscience Publication, 1999

The Stochastic Koopman operator (SKO)

Average value of the observable at time t:

$$\mathcal{S}_t : L^\infty(M) \to L^\infty(M), \mathcal{S}_t\hat{h}(x) = \mathbb{E}_{\omega\sim p}[\hat{h}(\Phi_t(\omega, x))]$$



Importance of SKO:

let $V_n = Span\{\hat{h}_1, \ldots, \hat{h}_n\}$ such that $\mathcal{S}_t V_n \subset V_n$

$$\text{Then } \mathcal{S}_t\hat{h}(x) = \sum_{i=1}^{n} \alpha_i \mathcal{S}_t\hat{h}_i(x) = \sum_{i=1}^{n} (\alpha S_m)_i \hat{h}_i(x)$$

matrix approximation

Eigenvalues and eigenvectors:

$$\mathcal{S}_t\hat{h}_i = \lambda_i\hat{h}_i \implies S_m \text{ diagonal}$$

1. M. Wanner and I. Mezic, Robust approximation of the stochastic Koopman operator, SIAM Journal on Applied Dynamical Systems, 21 (2022), pp. 1930–1951.
2. Črnjarić-Žic, Nelida, Senka Maćešić, and Igor Mezić. "Koopman operator spectrum for random dynamical systems." *Journal of Nonlinear Science* 30 (2020): 2007-2056.
3. Minghao Han, Jacob Euler-Rolle, Robert K. Katzschmann, DESKO: Stability-Assured Robust Control of Nonlinear Systems with a Deep Stochastic Koopman Operator

Dynamic Mode Decomposition[1]

$\hat{h}_i(x_{j+1})$

Choose V_n, what is the best approximation of $\mathcal{S}_t|_{V_n}$? $\quad argmin_{S_m} \sum_{i,j} |\mathcal{S}_t \hat{h}_i(x_j) - (S_m)_{ik} \hat{h}_k(x_j)|_2$

Given trajectory data: $\{x_j\}_{j=1}^m, m \geq n$

Compute :$(\Psi_m)_{ij} = \hat{h}_i(x_{j-1})$ and $(\Phi_m)_{ij} = \hat{h}_i(x_j)$

Return $S_m = \Phi_m \Psi_m^\dagger$

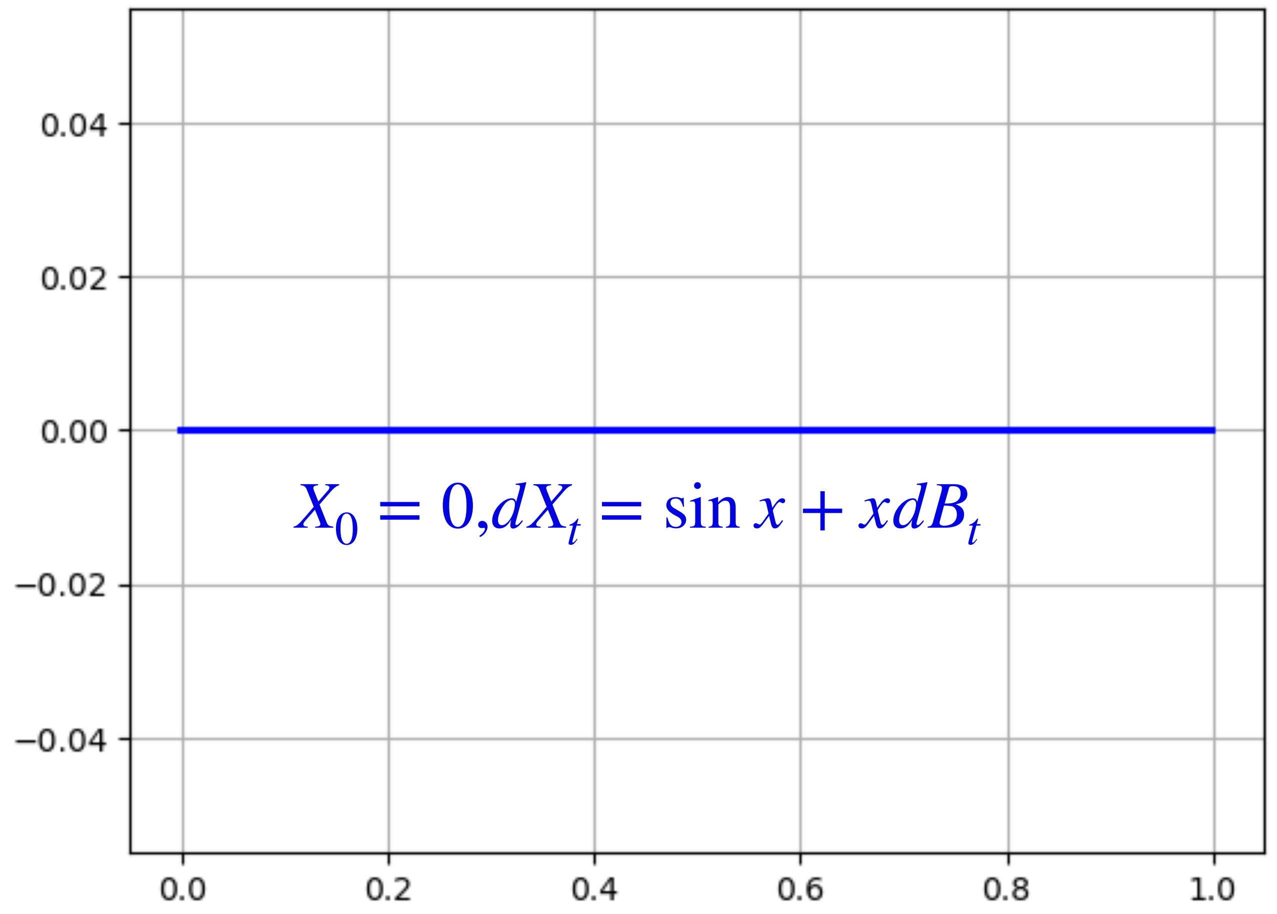
When $m \to \infty, S_m \to argmin \sum_{i=1}^{n} \|\mathcal{S}_t \hat{h}_i - (S_m)_{ik} \hat{h}_k\|^2_{L^2(M,\mu)} := S_\infty$

if $\{x_j\}_{j=1}^m$ ergodic, can interchange space and time averages

Matrix $S_m \in \mathbb{R}^{n \times n}$ vs finite rank operator $\mathcal{S}_m^n : V_n \to V_n$

[1]M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, A data–driven approximation of the Koopman operator: Extending dynamic mode decomposition, Journal of Nonlinear Science, 25 (2015), pp. 1307–1346

Shortcomings of SKO



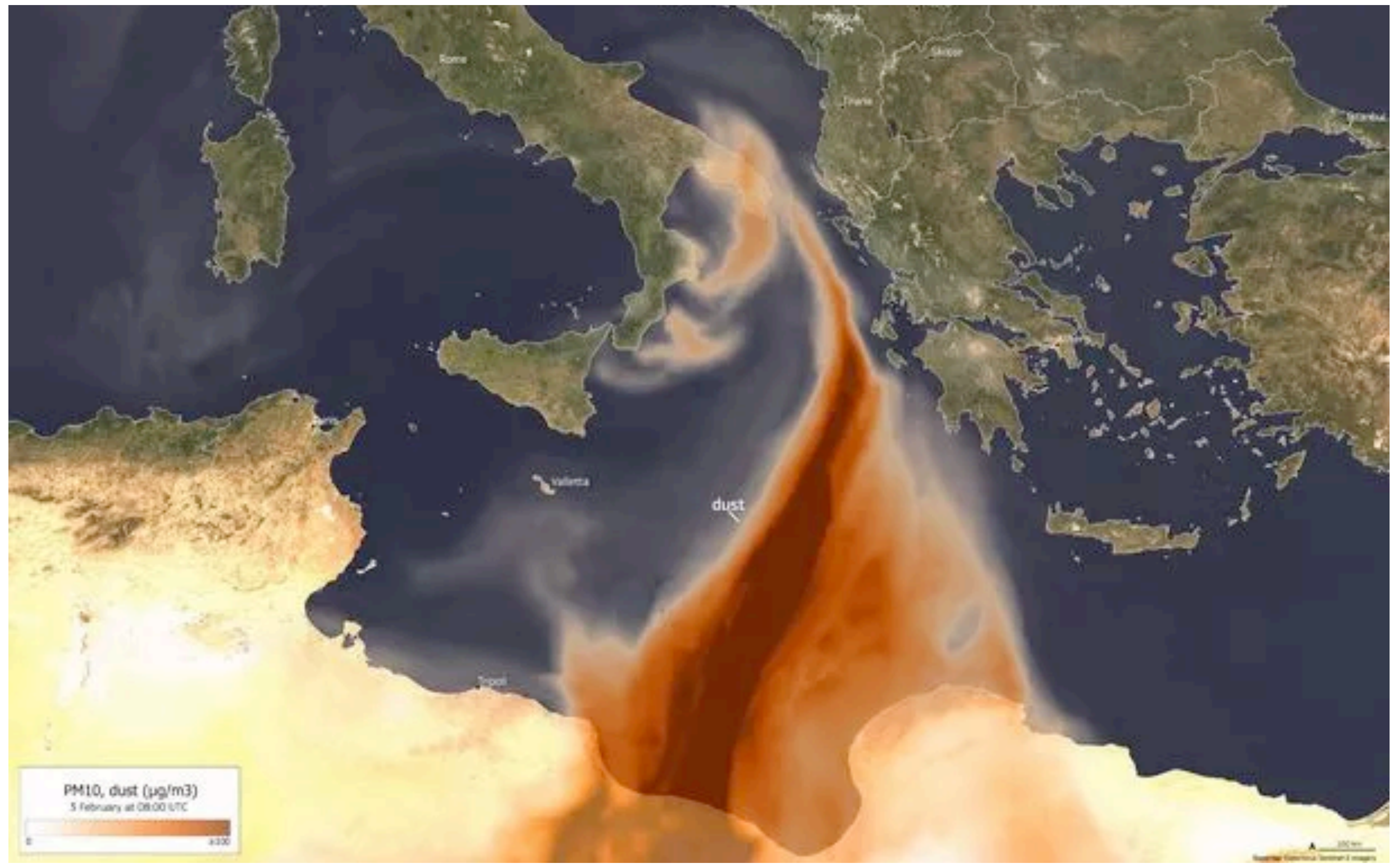
No higher order moment information

Dynamics might not be ergodic $\implies \nexists \mu$

Trajectory not representative

DMD cannot be applied when only aggregate data available

Unnatural $L^2(M, \mu)$ framework



Dust plume data off the coast of Libia

Credit: https://www.bristolpost.co.uk/news/uk-world-news/map-shows-blood-rain-dust-9920288

Lagrangian vs. Eulerian perspective

$x \in M$

$\hat{h} \in L^\infty(M)$

RDS: $\Phi : \mathbb{R} \times \Omega \times M \to M$

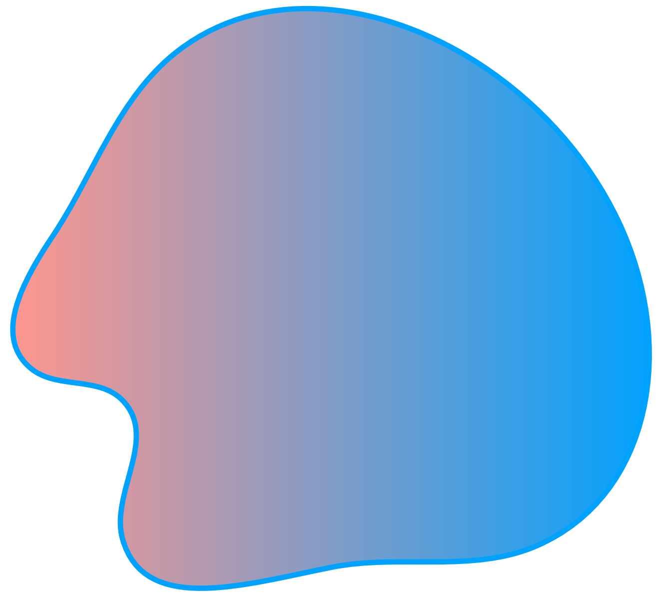
$\pi \in \mathscr{P}(M)$

$h : \mathscr{P}(M) \to \mathbb{R}$ continuous and bounded

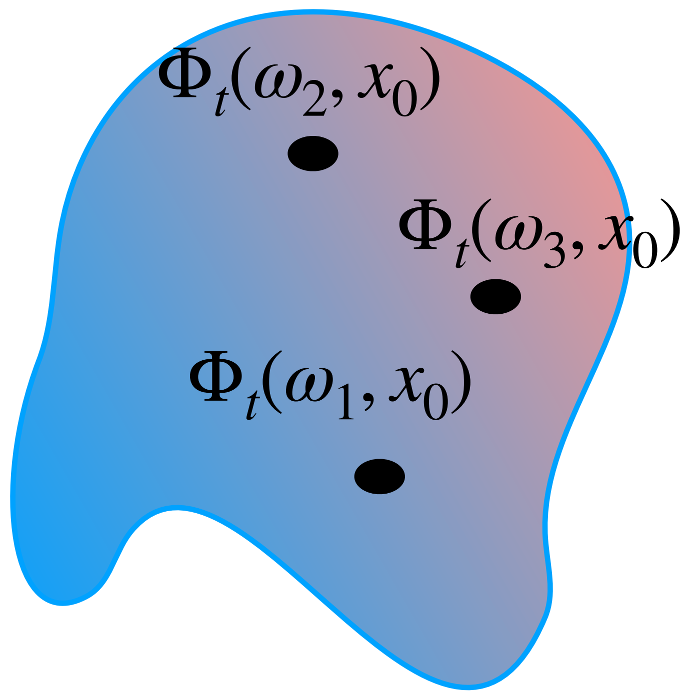
Transfer operator $T_t : \mathscr{P}(M) \to \mathscr{P}(M)$

π_0

$\frac{1}{m}\sum_j \delta_{x_j}$



$\xrightarrow{T_t}$



$\pi_t = T_t(\pi_0)$

$\frac{1}{m}\sum_j \delta_{x_{j+1}}$

The definition of DKO

$$\mathscr{D}_t h(\pi) = h \circ T_t(\pi)$$

Linearity ✓

Semi-group propriety $\mathscr{D}_{t+s} = \mathscr{D}_t \circ \mathscr{D}_s$ ✓

Invariant subspace $H_1 = \{h : \mathscr{P}(M) \to \mathbb{R}$ linear and bounded $\}$

Generalizes SKO when restricted to H_1

- Evaluation at delta: $\mathscr{D}_t h(\delta_x) = \mathscr{S}_t \hat{h}(x)$
- Integrate the $x-$uncertainty: $\mathscr{D}_t h(\pi) = \mathbb{E}_{X\sim\pi}\big[[\mathscr{S}_t \hat{h}](X)\big]$
- Same eigenvalues: $\mathscr{S}_t \hat{h} = \lambda \hat{h} \implies \mathscr{D}_t h = \lambda h$

Observables on $\mathscr{P}(M)$

Linear: $h(\pi) = \int \hat{h}(x)d\pi(x), \hat{h} \in L^\infty(M)$

Centered moments: $h(\pi) = \mathbb{E}_{X\sim\pi}\left[\left(\hat{h}(X) - \mathbb{E}_{X\sim\pi}[\hat{h}(X)]\right)^n\right]$

Inner product: $\Lambda : (\Theta, \mathbb{P}) \to \mathscr{P}(M)$ ⟶ Random measure

$$\Lambda\#\mathbb{P} := P \in \mathscr{P}(\mathscr{P}(M))$$

$$\langle h_1, h_2 \rangle := \int h_1 \circ \Lambda(\theta) h_2 \circ \Lambda(\theta) d\mathbb{P} = \int h_1(\pi)h_2(\pi)dP$$

Hilbert space: $(V_n, \langle \cdot , \cdot \rangle)$, where $V_n = Span\{h_1, \ldots, h_n\}$ ⟶ Finite dimensional

DMD Algorithm

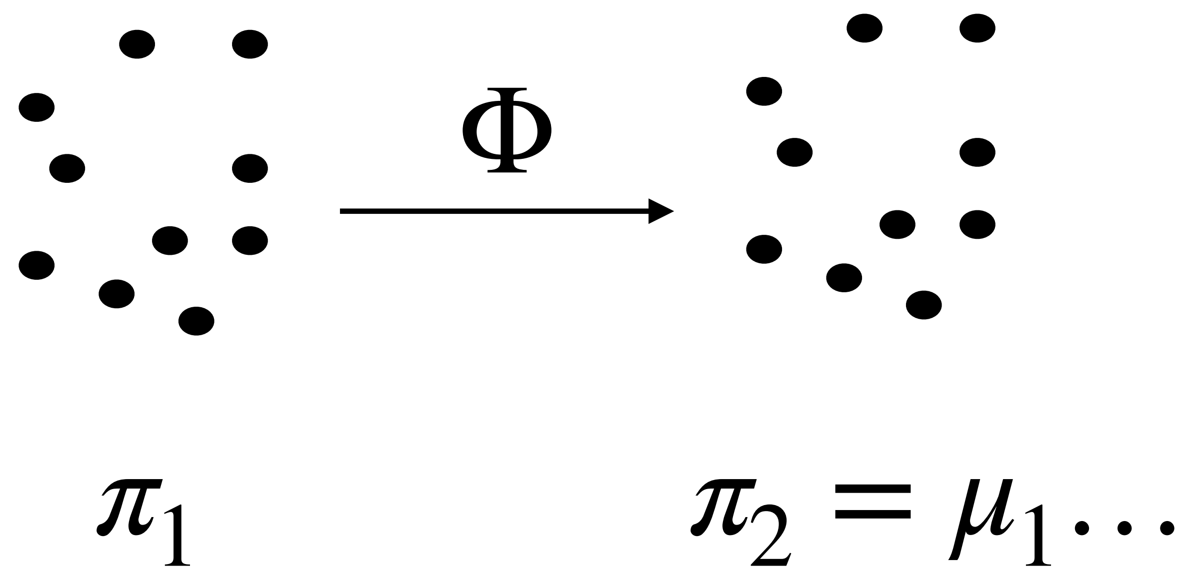
Choose V_n, what is the best approximation of $\mathscr{D}_t|_{V_n}$? $argmin_{D_m} \sum_{i,j} |\mathscr{D}_t h_i(\pi_j) - (D_m)_{ik} h_k(\pi_j)|_2^2$

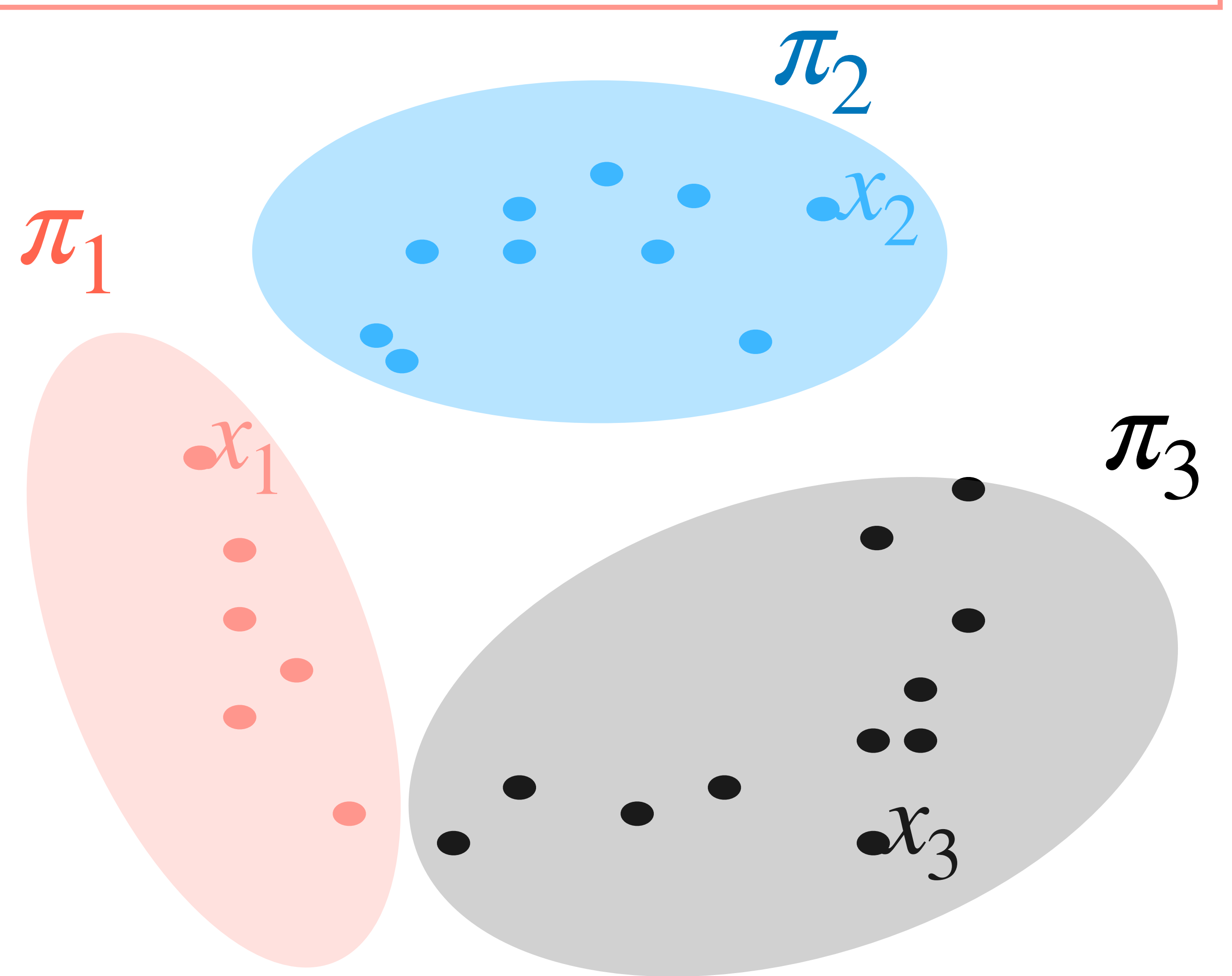
Given distributional data: $\{\pi_j\}_{j=1}^m, \{\mu_j = T_t \pi_j\}_{j=1}^m$

Compute :$(\Psi_m)_{ij} = h(\pi_j)$ and $(\Phi_m)_{ij} = h_i(\mu_j)$

Return $D_m = \Phi_m \Psi_m^\dagger$

Ways to obtain densities





Convergence guarantees

Define a Hilbert Schmidt norm for $Z_n = \{\mathscr{E} : V_n \to H, \mathscr{E} \text{ linear}\}$:

$$\|\mathscr{E}\|_{HS} = \|G^{-1}E\|_F, \text{ where } G_{ik} = \langle h_i, h_k \rangle, E_{ik} = \langle \mathscr{E}h_i, \mathscr{E}h_k \rangle$$

∞ -data

Minimize the HS norm of the residual operator: $\mathscr{D}_t^n = arg \min_{\mathscr{C}:V_n \to V_n} \|\mathscr{D}_t - \mathscr{C}\|_{HS}$

First order optimality: $D_\infty = YG^{-1}$, where $Y_{ik} = \langle h_i \circ T_t, h_k \rangle = \int h_i(T_t(\pi))h_k(\pi)dP$

m -data

$$D_m = \Phi_m \Psi_m^\dagger = \frac{1}{m}\Phi_m \Psi_m^\top \left(\frac{1}{m}\Psi_m \Psi_m^\top \right)^{-1}$$

$$\frac{1}{m}\sum_j h_i(T_t(\pi_j))h_k(\pi_j) \qquad \frac{1}{m}\sum_j h_i(\pi_j)h_k(\pi_j)$$

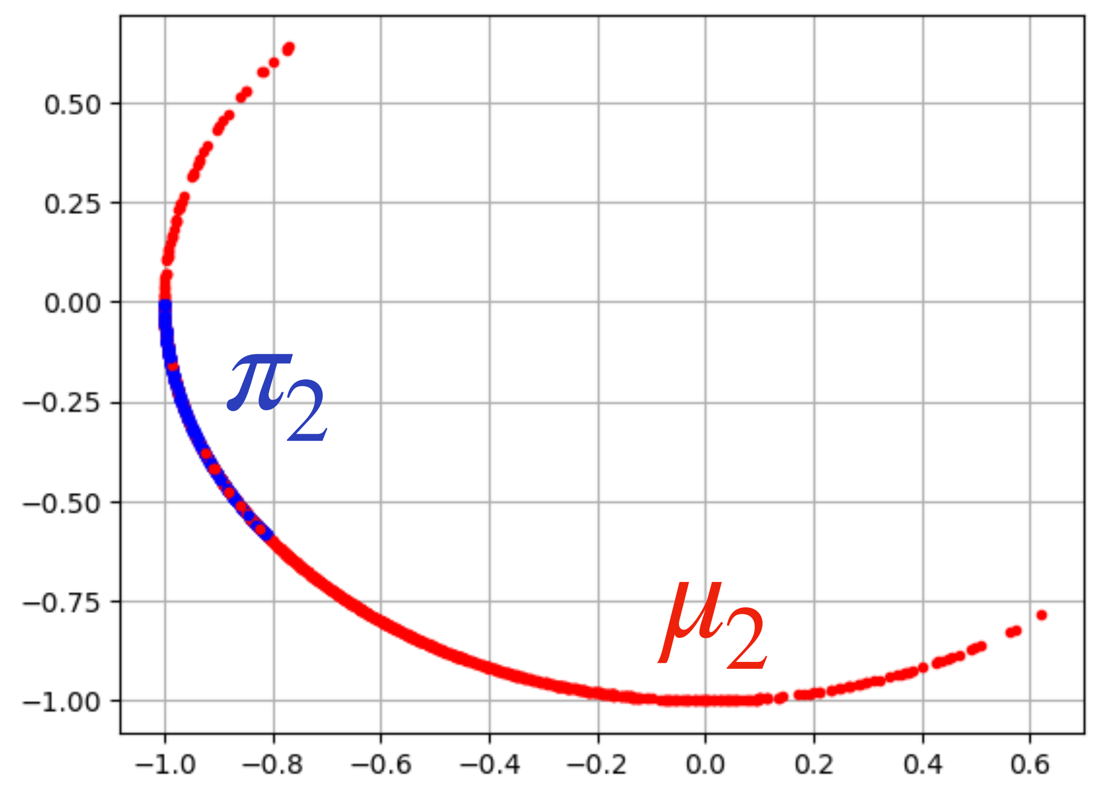
Convergence guarantees

Theorem (O., Townsend, Yang, 2025) Let $h_1, \ldots, h_n$ be linearly independent and assume the condition number of $G_{ik} = \langle h_i, h_k \rangle$ is bounded. If for any fixed m, $\pi_1, \ldots, \pi_m \sim P$ I.i.d then D_m converges to **the best matrix approximation** D_∞ of the DKO on $V_n = Span\{h_i\}_{i=1}^n$

$$\|D_m - D_\infty\|_F \to 0 \iff \|\mathcal{D}_t^n - \mathcal{D}_m^n\|_{HS} \to 0$$

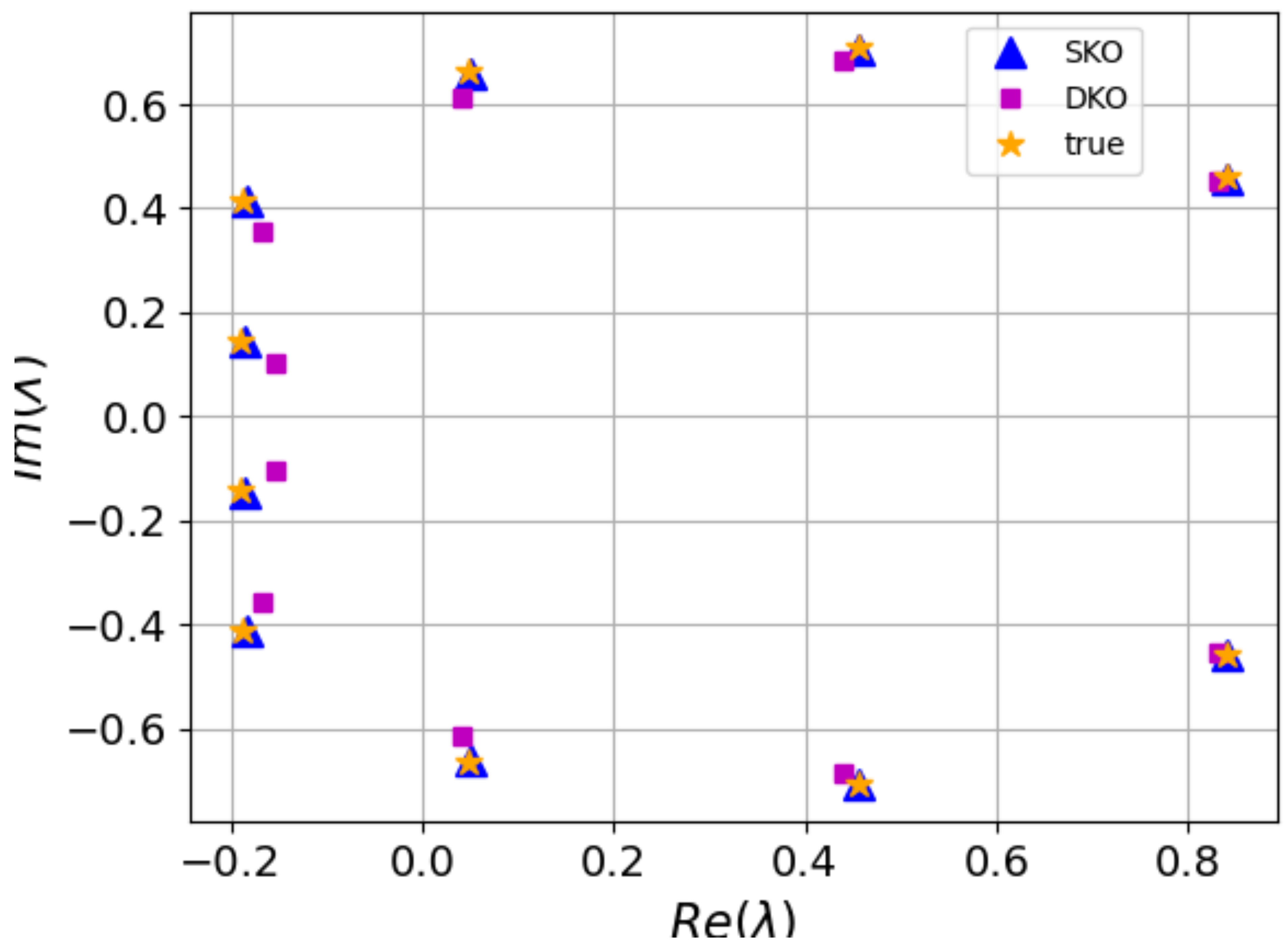
Numerical results

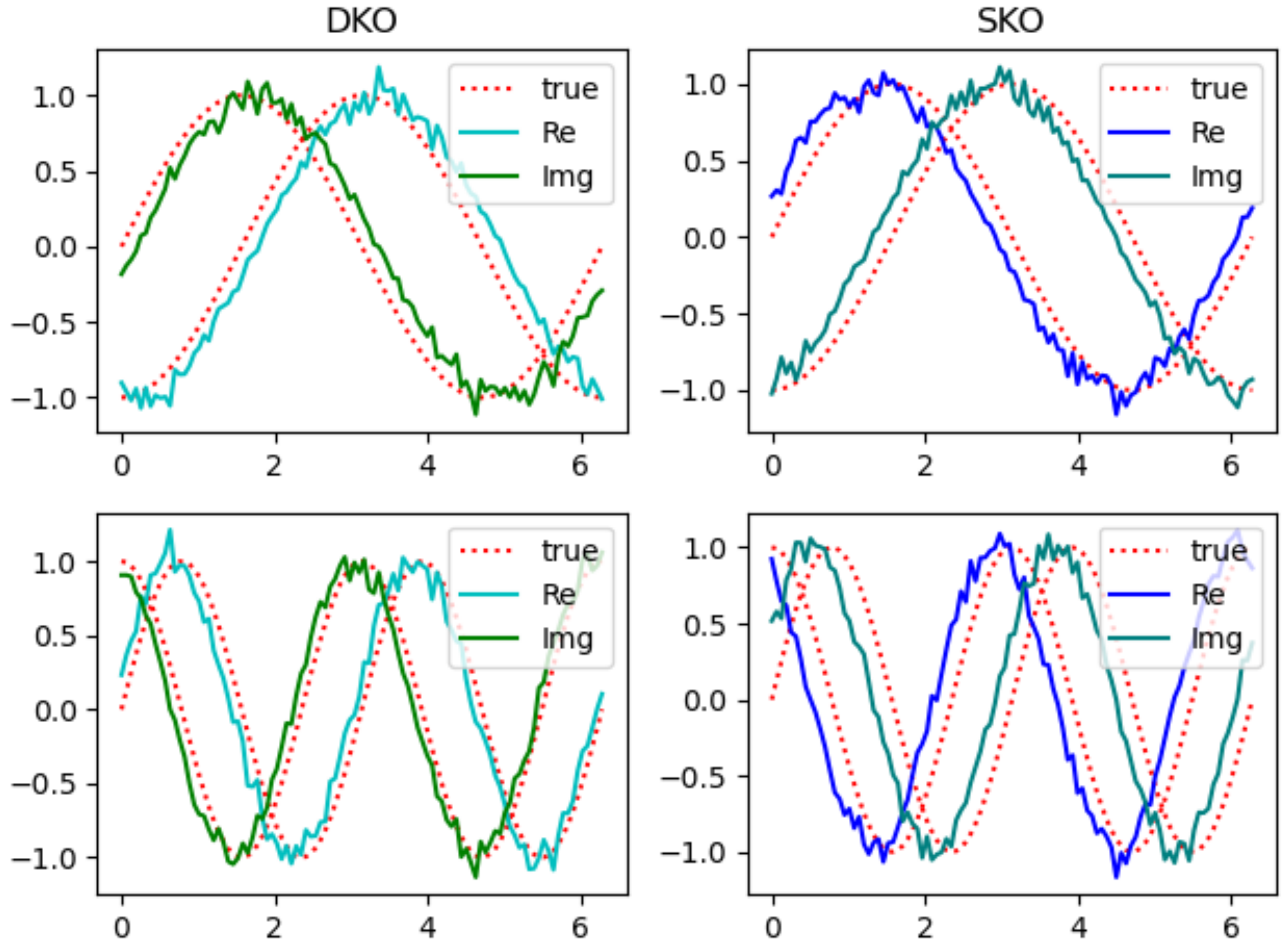
Random rotations on a circle $\theta_{t+1} = \theta_t + \nu + \omega_t,\ \nu = 0.5$



Observables $\hat{h}_i = 1_{\left[i\frac{2\pi}{N},(i+1)\frac{2\pi}{N}\right]}$

Measures $\pi_i = Unif\left[j\frac{2\pi}{N},(j+1)\frac{2\pi}{N}\right]$





Numerical results

Variance for an SDE

$$dX_t = -\sin X_t dt + e^{-0.5(x-1)^2} d\omega_t$$

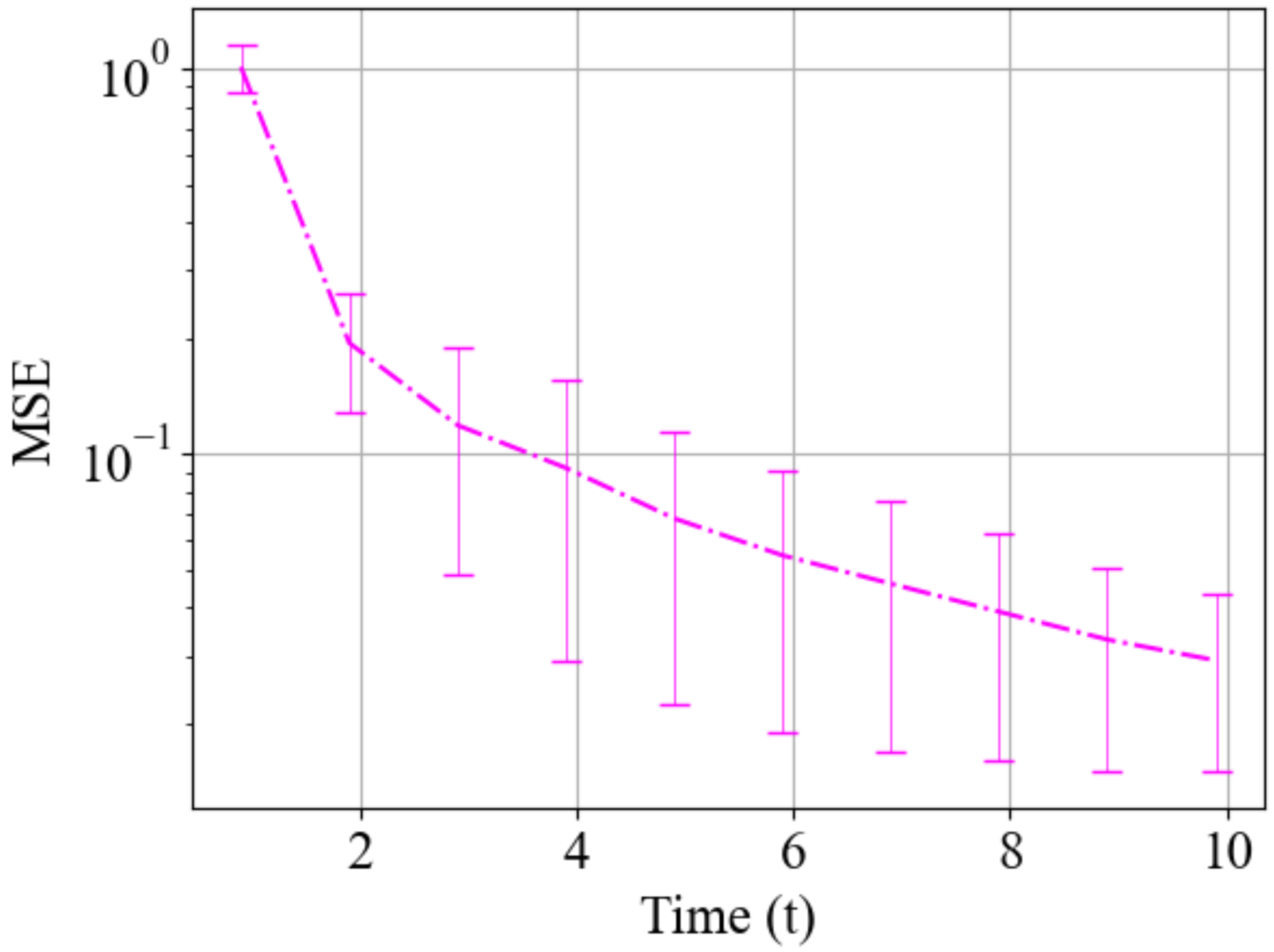
Start with radial basis functions with $n = 9$

$$\hat{h}_i = e^{-\epsilon(x-c_i)^2}$$

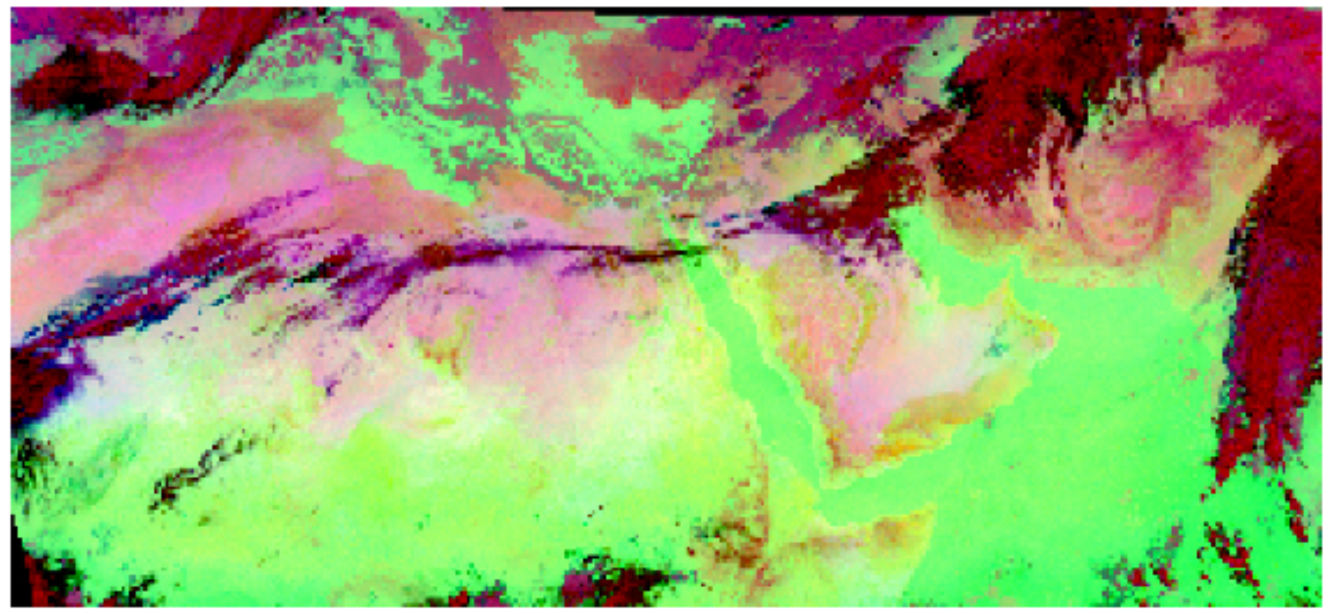
Observables $\mathbb{E}_\pi[\hat{h}_i \hat{h}_k], \mathbb{E}_\pi[\hat{h}_i]\mathbb{E}_\pi[\hat{h}_k]$

Measures $\{\pi_j\}_{j=1}^{20}, \pi_0 = \mathcal{N}(0,1)$

Observe for time T = 2 and predict for time T = 10



Numerical results



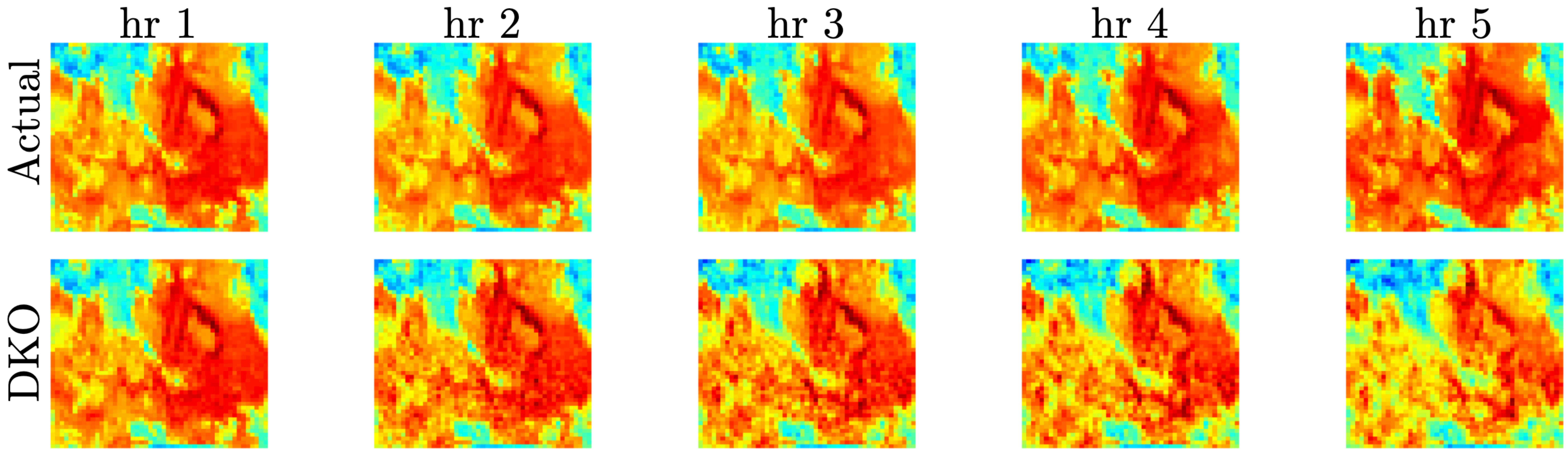
Dust plume DUSTScan2022 data

Dataset = hourly dust observations from SEVIRI on Meteosat-8

Data = pictures of dust density as a function of deviation from magenta

Observables = average over patches of PDI index over 50x50 pixels

Use 28 days worth of data and predict the following 5 hours



Thank you for your attention!

M.O., Alex Townsend, Yunan Yang, The Distributional Koopman Operator for Random Dynamical Systems, arXiv preprint, 2025

Transfer operator

Satisfies the Fokker Plank equation

$$\partial T_t(\pi) = -\nabla \cdot \Big(f(x)\pi_t - \nabla \cdot (g(x)\pi_t\Big)$$

Definition

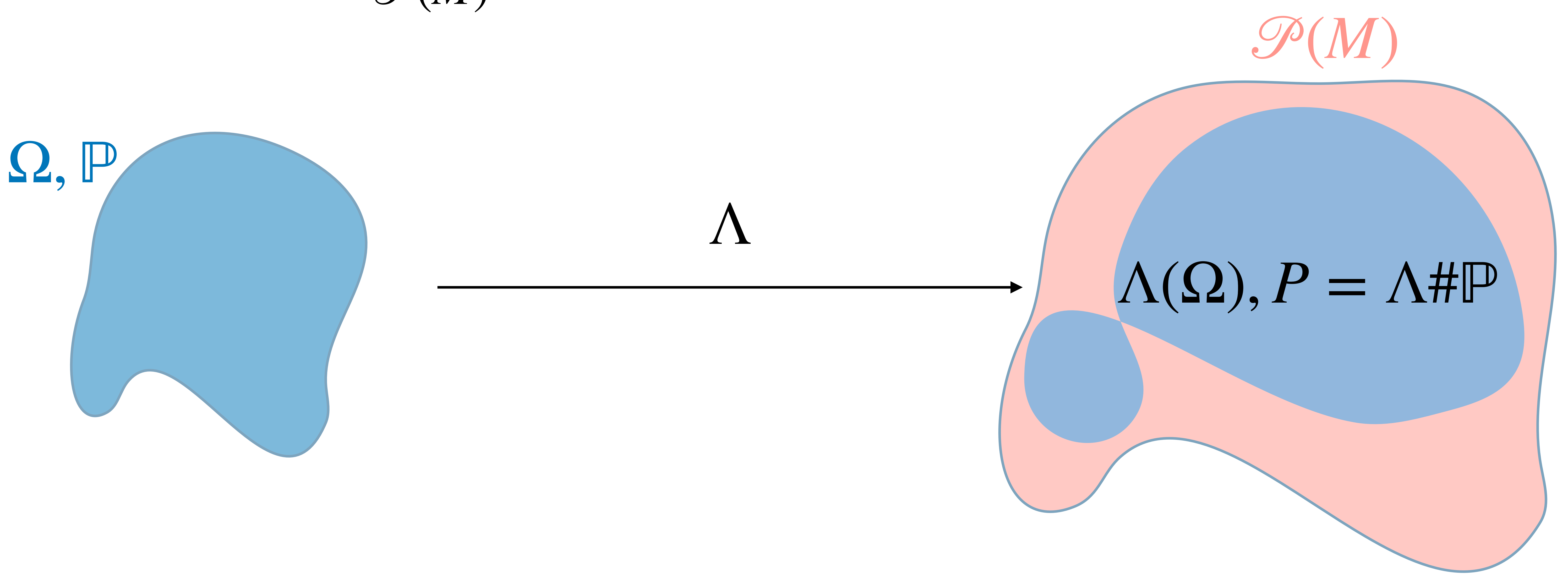
$$\int_M \hat{h}(x)\, d\Big(T_t\pi\Big)(x) = \int_\Omega \int_M \hat{h}(x)\, d\Bigg(\Phi_t(\omega, \cdot\,)\#\pi\Bigg)(x)\, dp(\omega),$$

Adjoint of SKO

$$\langle \mathcal{S}_t \hat{h}, \pi \rangle = \langle \hat{h}, T_t\pi \rangle$$

An inner product structure on H

Holy grail: $\langle h_1, h_2 \rangle = \int_{\mathscr{P}(M)} h_1(\pi) h_2(\pi) dP$ ⟵ Does not exist



Enough to define on a dense subset of $\mathscr{P}(M)$ and extend by continuity

|{ space of all empirical measures} | = |[0,1]|

Sample empirical measures $\pi_j \sim \Lambda \# \mathbb{P}$ and let $\langle h_1, h_2 \rangle = \frac{1}{m} \sum_{j=1}^{m} h_1(\pi_j) h_2(\pi_j)$ ✓

Details for the numerical examples

Random rotations on a circle $\omega \sim Unif[-0.5, 0.5]$

used 20 measures, each with 1000 samples