Transport- and Measure-Theoretic Approaches for Modeling, Identifying, and Forecasting Dynamical Systems

Yunan Yang, Cornell University

October 9, 2025

IPAM Workshop: Bridging Scales from Atomistic to Continuum in Electrochemical Systems, Oct 6-10, 2025

List of works:
— Optimal transport for parameter identification of chaotic dynamics via invariant measures. 2023. *SIADS*.
— Learning dynamics on invariant measures using PDE-constrained optimization. 2023. *Chaos*.
— Invariant Measures in Time-Delay Coordinates for Unique Dynamical System Identification. 2025. *PRL*.
— The distributional Koopman operator for random dynamical systems. 2025. *MCSS*.
— Measure-Theoretic Time-Delay Embedding. arXiv:2409.08768.

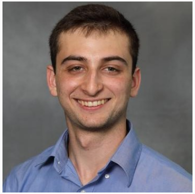
Collaborators

Levon Nurbekyan (Emory)

Robert Martin (ARL)

Elisa Negrini (UCLA)

Mirjeta Pasha (Virginia Tech)



Jonah Botvinick-Greenhouse (Cornell)

Maria Oprea (Cornell)

Romit Maulik (PSU)

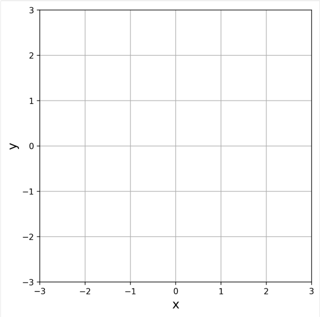
Alex Townsend (Cornell)

Data-Driven Modeling of Dynamical Systems

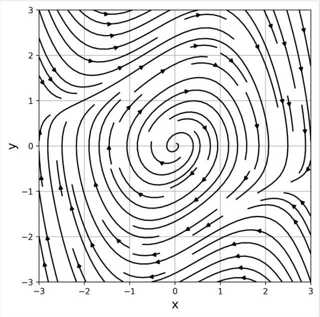
Data-Driven Modeling for Dynamical System

X | $\dot{x} = v(x)$ | data-driven modeling (←) | $\{x(t_k)\}_{k=0}^{N-1}$

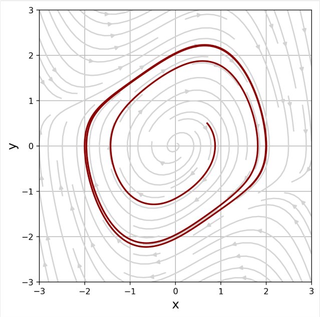
State space



Evolution rule



Trajectory samples



Parameter Identification

A general parameterized dynamical system may take the form

$$\dot{\mathbf{x}} = \begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} = v(x, y, z; \underbrace{\sigma, \rho, \beta}_{\theta}) \approx v(\mathbf{x}, \theta)$$

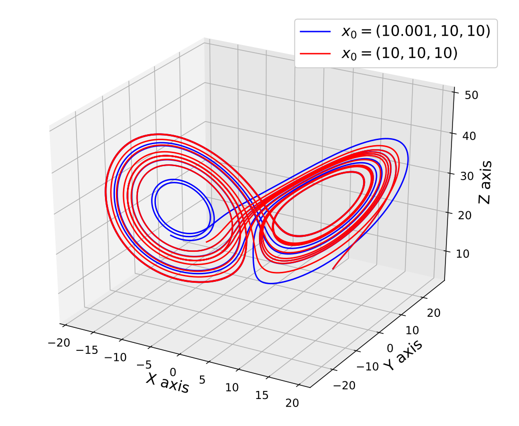
where the mathematical approximation $v \approx v(\cdot, \theta)$ is given by

- polynomials, e.g., SINDy
- other basis functions, e.g., piecewise polynomials, RBFs, Fourier, etc.
- neural networks, and so on,
- (many many references)

where θ corresponds to **expansion coefficients, neural network weights**, etc.

Unique Challenges for Chaotic Systems: Chaos

Challenge One: The initial condition of the system is unknown.



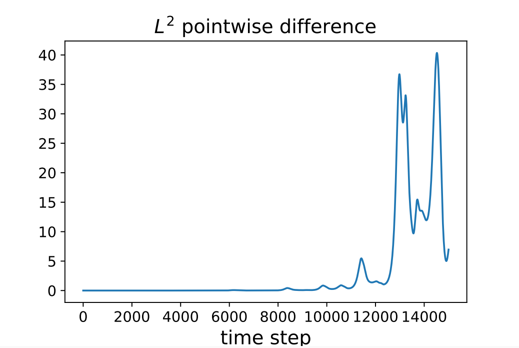


Figure: The comparison between $\mathbf{x}_0 = [10.001, 10, 10]$ and $\mathbf{x}_0 = [10, 10, 10]$.

Unique Challenges for Chaotic Systems: Noises

Challenge Two: The time trajectories contain noise.

No noise

$$\dot{\mathbf{x}} = f(\mathbf{x}).$$

Extrinsic noise

$$\mathbf{x}_\gamma = \mathbf{x} + \gamma, \ \dot{\mathbf{x}} = f(\mathbf{x}).$$

Intrinsic noise

$$\dot{\mathbf{x}} = f(\mathbf{x}) + \omega.$$

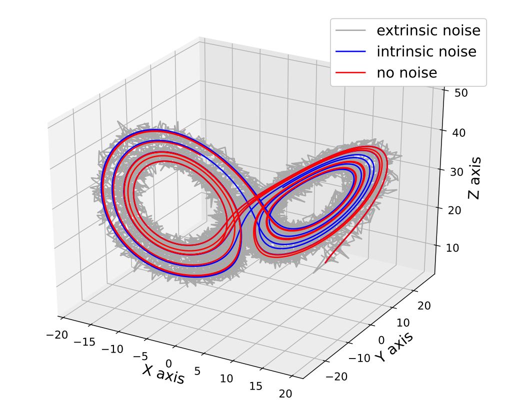


Figure: The comparison among the three cases.

Unique Challenges for Chaotic Systems: Poor Data Quality

Challenge Three: Cannot measure the Lagrangian particle velocity flow

Measurements $\{\mathbf{x_i}\}$ are not good enough to estimate the particle velocity $\dot{\mathbf{x}}$ evaluated at $\{\mathbf{x_i}\}$

$$\hat{v} \approx \frac{\mathbf{x}_{i+1} - \mathbf{x}_i}{t_{i+1} - t_i}$$

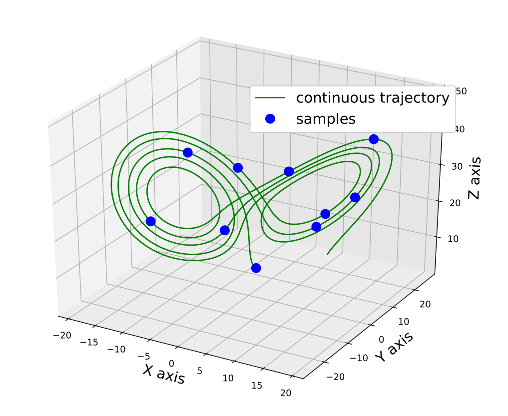


Figure: The continuous trajectory vs the samples

From Lagrangian to Eulerian

Often, chaotic systems admit well-defined statistical properties:

$$\mu_{x,T}(B) = \frac{1}{T}\int_0^T \mathbb{1}_B(\mathbf{x}(s))ds = \frac{\int_0^T \mathbb{1}_B(\mathbf{x}(s))ds}{\int_0^T \mathbb{1}_{\mathbb{R}^d}(\mathbf{x}(s))ds},$$

where $\mathbf{x}(t)$ is a trajectory starting with $\mathbf{x}(0) = x$, and $\mu_{x,T}$ is called the *occupation measure*. We call μ^* a physical measure if $\lim_{T\to\infty} \mu_{x,T} = \mu^*$ for $x \in U$, $\text{Leb}(U) > 0$.

Data Change: take μ^* as observation data instead of the trajectory $\mathbf{x}(t)$.
Model Change: μ^* is the **steady**-state solution to the continuity equation:

$$\frac{\partial \rho(\mathbf{x}, t)}{\partial t} + \nabla \cdot (v(\mathbf{x}, \theta)\rho(\mathbf{x}, t)) = 0\,.$$

Road Map: From Lagrangian to Eulerian

ODE model $\dot{\mathbf{x}} = v(\mathbf{x})$, observe $\{\mathbf{x}(t_i)\}_i$

$\Downarrow$

Occupation measure

$$\mu_{x,T}(B) = \frac{1}{T}\int_0^T \mathbb{1}_B(\mathbf{x}(s))ds = \frac{\int_0^T \mathbb{1}_B(\mathbf{x}(s))ds}{\int_0^T \mathbb{1}_{\mathbb{R}^d}(\mathbf{x}(s))ds}$$

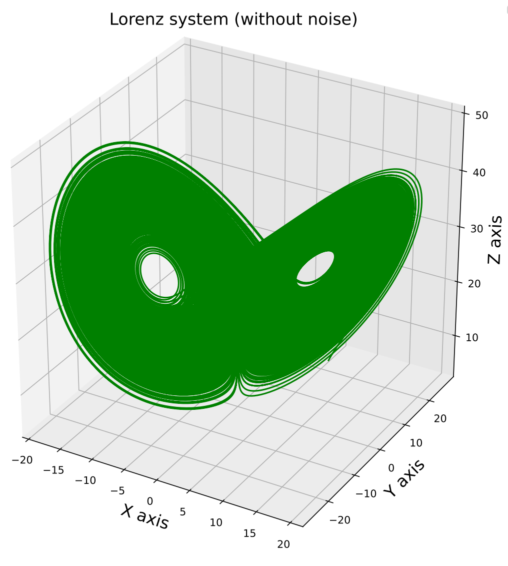
$\Downarrow$

physical measure μ^*

$\Downarrow$

Stationary distributional solutions of

$$\frac{\partial \rho(\mathbf{x}, t)}{\partial t} + \nabla \cdot (v(\mathbf{x}, \theta)\rho(\mathbf{x}, t)) = 0.$$



The New Method — A PDE-Constrained Optimization Problem

We treat the parameter identification problem for the dynamical system as a PDE-constrained optimization problem:

$$\theta = \underset{\theta}{\operatorname{argmin}}\, J(\rho(\theta), \rho^*),$$

$$\text{s.t.} \quad \frac{\partial \rho}{\partial t} = -\nabla \cdot \big(v(\mathbf{x}, \theta)\rho(\mathbf{x}, t)\big) \boxed{+\frac{1}{2}\frac{\partial^2 D_{ij}\rho}{\partial x_i \partial x_j}} = 0.$$

ρ^* : the observed occupation measure converted from time trajectories

$\rho(\theta)$: the distributional steady-state solution of the PDE

J : an appropriate metric/divergence comparing probability measures, e.g., W_2

[Y.-Nurbekyan-Negrini-Martin-Pasha, 2023, SIADS] [Botvinick-Martin-Y., 2023, Chaos]

The New Method — A PDE-Constrained Optimization Problem

We treat the parameter identification problem for the dynamical system as a PDE-constrained optimization problem:

$$\theta = \underset{\theta}{\operatorname{argmin}} J(\rho(\theta), \rho^*),$$

$$\text{s.t.} \quad \frac{\partial \rho}{\partial t} = -\nabla \cdot \big(v(\mathbf{x}, \theta)\rho(\mathbf{x}, t)\big) + \frac{1}{2}\frac{\partial^2 D_{ij}\rho}{\partial x_i \partial x_j} = 0.$$

ρ^* : the observed occupation measure converted from time trajectories

$\rho(\theta)$: the distributional steady-state solution of the PDE

J : an appropriate metric/divergence comparing probability measures, e.g., W_2

Data and forward problem are changed, but parameters remain the same.

The gain is to work with a much **More Stable** inverse problem!

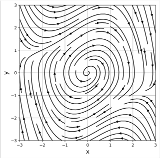
[Y.-Nurbekyan-Negrini-Martin-Pasha, 2023, SIADS] [Botvinick-Martin-Y., 2023, Chaos]

Summary: Our Approach From Lagrangian View to Eulerian Perspective

$\{x(t_k)\}_{k=0}^{N-1}$

?

$v(x)$



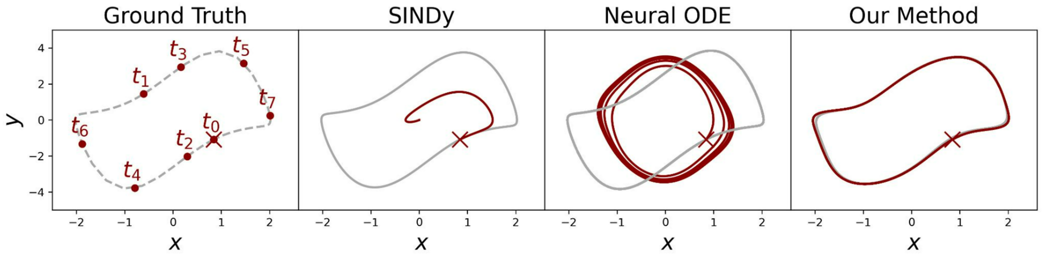
SINDy[1]	Shooting methods[2]	Neural ODEs[3]

- Noise blows up divided difference
- Slow sampling makes divided difference inaccurate
- Unable to distinguish small modeling errors from chaos

Data	*Forward Model*	*Objective Function*
$\rho^* := \frac{1}{N}\sum_{k=0}^{N-1}\delta_{x(t_k)}$	$\theta \mapsto \rho(\theta)$	$\min_{\theta\in\Theta} \mathcal{J}(\rho(\theta), \rho^*)$

[1]Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, *113*(15), 3932-3937.
[2]Michalik, C., Hannemann, R., & Marquardt, W. (2009). Incremental single shooting—a robust method for the estimation of parameters in dynamical systems. *Computers & Chemical Engineering*, *33*(7), 1298-1305.
[3]Chen, R. T., Rubanova, Y., Bettencourt, J., & Duvenaud, D. K. (2018). Neural ordinary differential equations. *Advances in neural information processing systems*, *31*.

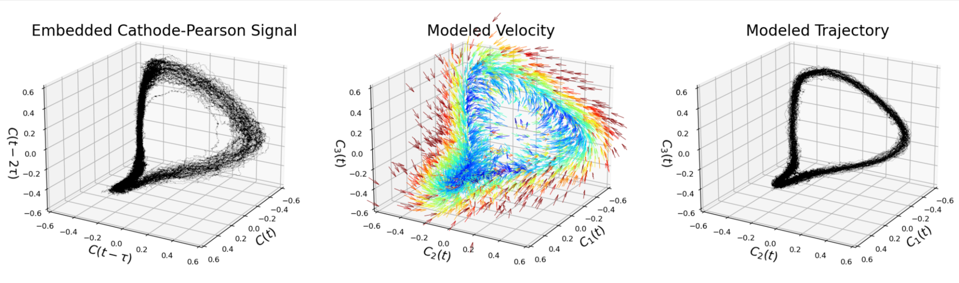
Comparison with Other Methods

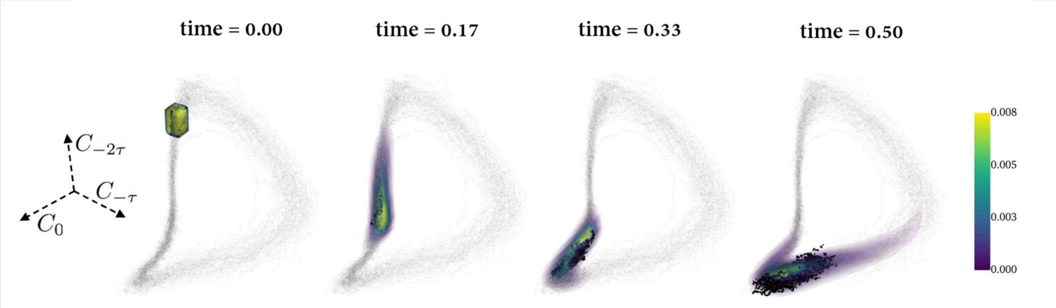


Method	Sampling Freq.	Wall-Clock Time (s)	Error
SINDy	10.00	$2 \cdot 10^{-2}$	$5.6 \cdot 10^{-3}$
Neural ODE	10.00	$5 \cdot 10^{2}$	$5.32 \cdot 10^{-3}$
Ours	10.00	$5 \cdot 10^{2}$	$1.14 \cdot 10^{-1}$

Method	Sampling Freq.	Wall-Clock Time (s)	Error
SINDy	0.25	10^{-2}	3.52
Neural ODE	0.25	$5 \cdot 10^{2}$	1.81
Ours	0.25	$5 \cdot 10^{2}$	$6.79 \cdot 10^{-2}$

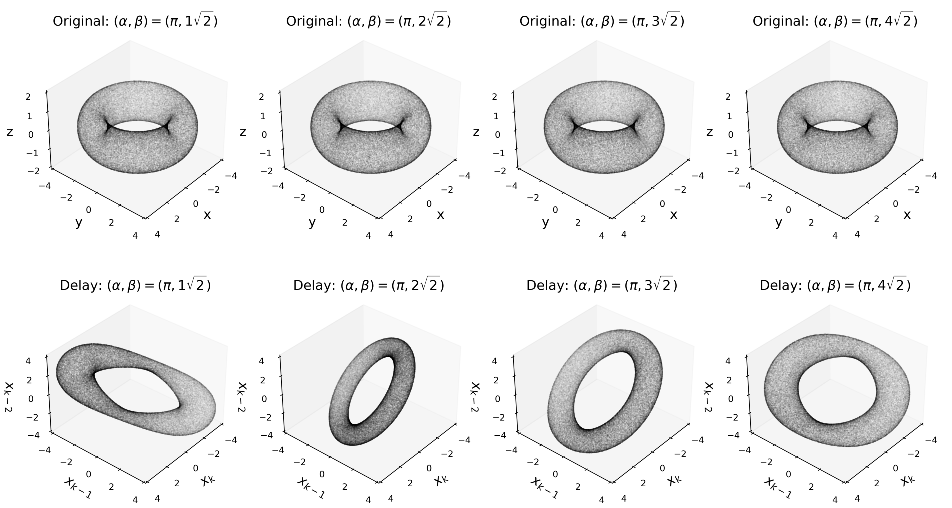
Application to Real-World Data: Hall-Effect Thruster





Limitation: Nonuniqueness

$$T_{\#}\mu = \mu \ \& \ S_{\#}\mu = \mu \ \not\Rightarrow \ T = S$$



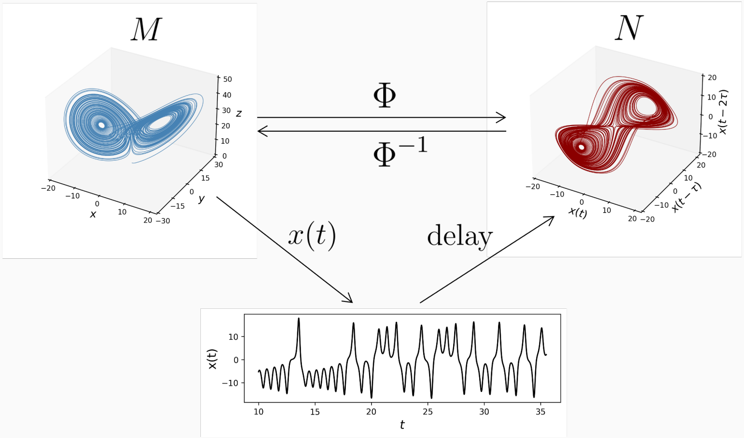
Invariant Measures in *Time-Delay Coordinates* for *Unique* Dynamical System Identification

Takens' Embedding Theorem

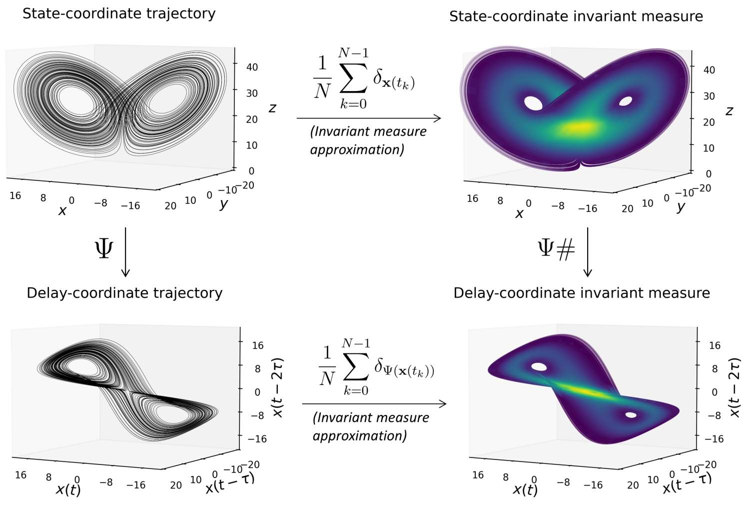
Theorem (Takens, 1981)
Let M be a compact manifold of dimension m. For pairs (y, T), where $T \in C^2(M, M)$ and $y \in C^2(M, \mathbb{R})$, it is a generic property that the mapping $\Phi_{(y,T)} : M \to N \subseteq \mathbb{R}^{2d+1}$ given by $\Phi_{(y,T)}(\mathbf{x}) := (y(\mathbf{x}), y(T(\mathbf{x})), \ldots, y(T^{2m}(\mathbf{x})))$ is an embedding of M in $\mathbb{R}^{2d+1}$.

Takens' Embedding Theorem

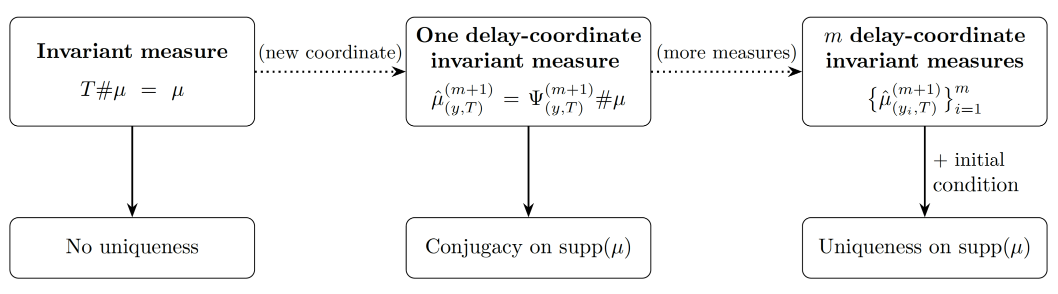
Theorem (Takens, 1981)
Let M be a compact manifold of dimension m. For pairs (y, T), where $T \in C^2(M, M)$ and $y \in C^2(M, \mathbb{R})$, it is a generic property that the mapping $\Phi_{(y,T)} : M \to N \subseteq \mathbb{R}^{2d+1}$ given by $\Phi_{(y,T)}(\mathbf{x}) := (y(\mathbf{x}), y(T(\mathbf{x})), \ldots, y(T^{2m}(\mathbf{x})))$ is an embedding of M in $\mathbb{R}^{2d+1}$.



Invariant Measures in Time-Delay Coordinates

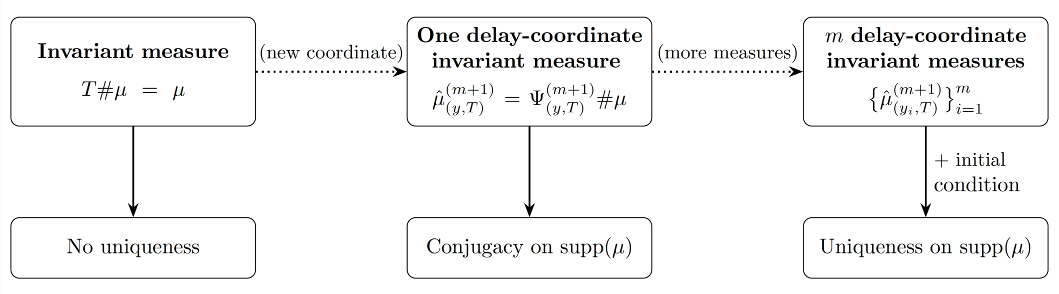


Invariant Measures in Time-Delay Coordinates for Uniqueness



[Botvinick-Greenhouse-Martin-Y., 2025, PRL]

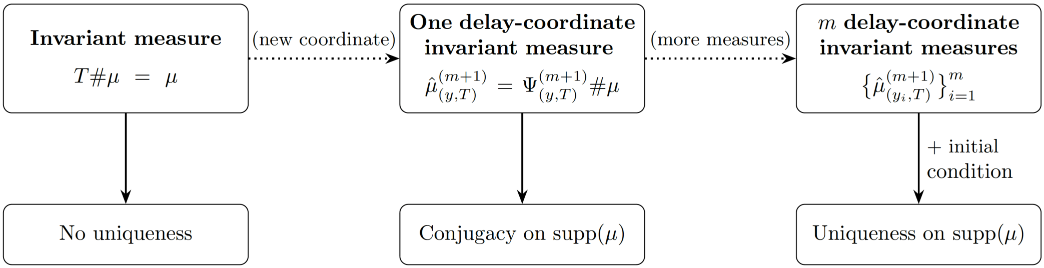
Invariant Measures in Time-Delay Coordinates for Uniqueness



Theorem 1. The equality $\hat{\mu}^{(m+1)}_{(y,T)} = \hat{\nu}^{(m+1)}_{(y,S)}$ implies $T|_{\text{supp}(\mu)}$ and $S|_{\text{supp}(\nu)}$ are topologically conjugate, for almost every $y \in C^1(U, \mathbb{R})$.

[Botvinick-Greenhouse-Martin-Y., 2025, PRL]

Invariant Measures in Time-Delay Coordinates for Uniqueness



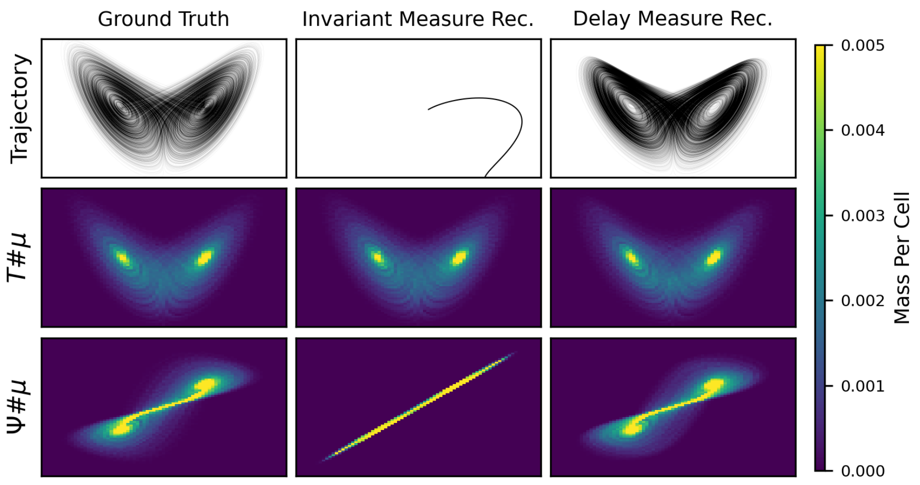
Theorem 1. The equality $\hat{\mu}^{(m+1)}_{(y,T)} = \hat{\nu}^{(m+1)}_{(y,S)}$ implies $T|_{\text{supp}(\mu)}$ and $S|_{\text{supp}(\nu)}$ are topologically conjugate, for almost every $y \in C^1(U, \mathbb{R})$.

Theorem 2. The conditions below imply that $T = S$ on supp(μ), for a.e. $Y \in C^1(U, \mathbb{R}^m)$:

1. there exists $x^* \in B_{\mu,T} \cap \text{supp}(\mu)$, such that $T^k(x^*) = S^k(x^*)$ for $1 \le k \le m-1$, and
2. $\hat{\mu}^{(m+1)}_{(y_j,T)} = \hat{\mu}^{(m+1)}_{(y_j,S)}$ for $1 \le j \le m$, where $Y := (y_1, \ldots, y_m)$ is a vector-valued observable.

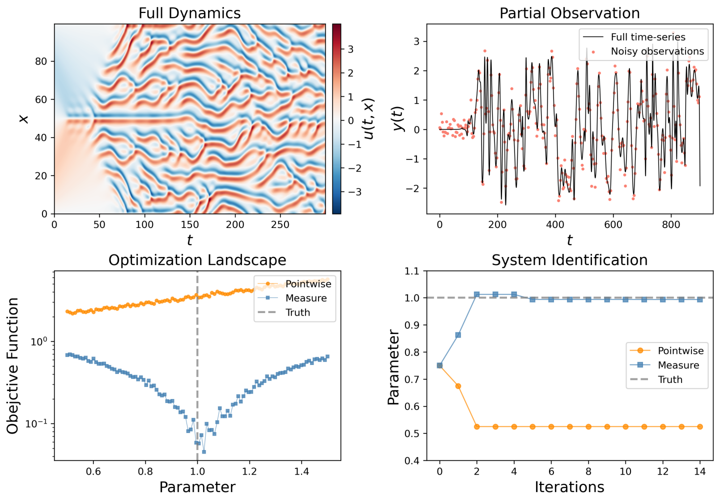
[Botvinick-Greenhouse-Martin-Y., 2025, PRL]

Numerical Example: State- vs. Delay-Coordinate Invariant Measure



[Botvinick-Greenhouse-Martin-Y., 2025, PRL]

Numerical Example: State- vs. Delay-Coordinate Invariant Measure



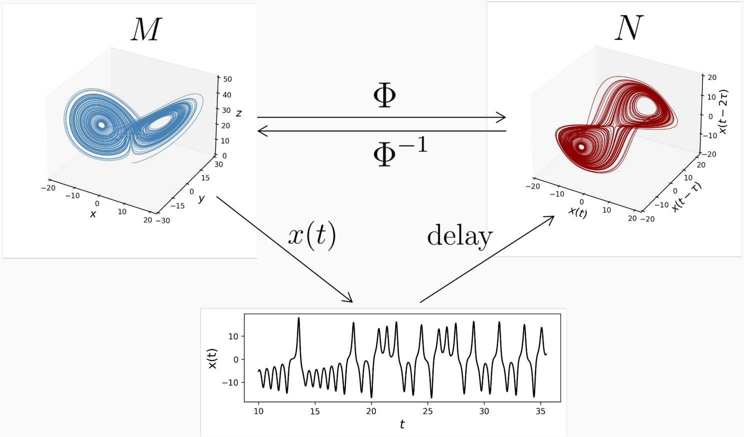
[Botvinick-Greenhouse-Martin-Y., 2025, PRL]

Embedding Over the Probability Space $\mathcal{P}_2(M)$

Takens' Embedding Theorem (Again)

Theorem (Takens, 1981)

Let M be a compact manifold of dimension m. For pairs (y, T), where $T \in C^2(M, M)$ and $y \in C^2(M, \mathbb{R})$, it is a generic property that the mapping $\Phi_{(y,T)} : M \to N \subseteq \mathbb{R}^{2d+1}$ given by $\Phi_{(y,T)}(\mathbf{x}) := (y(\mathbf{x}), y(T(\mathbf{x})), \ldots, y(T^{2m}(\mathbf{x})))$ is an embedding of M in $\mathbb{R}^{2d+1}$.



Measure-Theoretic Embedding

- Challenges: Takens' Theorem no longer applies when dynamics have noise.
- Can we **lift** the statement to the **space of probability measures**?

[Botvinick-Oprea-Maulik-Y., 2024, arXiv:2409.08768]

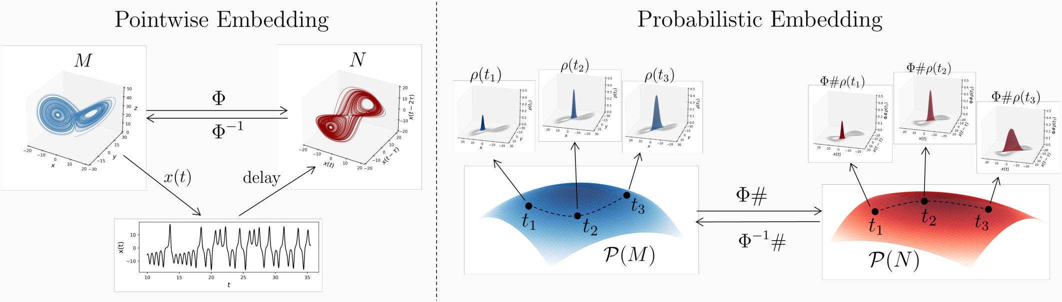
Measure-Theoretic Embedding

- Challenges: Takens' Theorem no longer applies when dynamics have noise.
- Can we **lift** the statement to the **space of probability measures**?
- If $\Phi : M \to N$ is an embedding, is $\Phi\# : \mathcal{P}(M) \to \mathcal{P}(N)$ also an embedding?

[Botvinick-Oprea-Maulik-Y., 2024, arXiv:2409.08768]

Measure-Theoretic Embedding

- Challenges: Takens' Theorem no longer applies when dynamics have noise.
- Can we **lift** the statement to the **space of probability measures**?
- If $\Phi : M \to N$ is an embedding, is $\Phi\# : \mathcal{P}(M) \to \mathcal{P}(N)$ also an embedding?



[Botvinick-Oprea-Maulik-Y., 2024, arXiv:2409.08768]

Measure-Theoretic Embedding

Pointwise embedding (Φ)

1. Φ is injective
2. Φ is smooth
3. $D\Phi$ is injective

[Botvinick-Oprea-Maulik-Y., 2024, arXiv:2409.08768]

Measure-Theoretic Embedding

Pointwise embedding (Φ)

1. Φ is injective
2. Φ is smooth
3. $D\Phi$ is injective

Measure-theoretic embedding ($\Phi\#$)

1. $\Phi\#$ is injective
2. $\Phi\#$ is smooth
3. $D(\Phi\#)$ is injective

[Botvinick-Oprea-Maulik-Y., 2024, arXiv:2409.08768]

Measure-Theoretic Embedding

Pointwise embedding (Φ)

1. Φ is injective
2. Φ is smooth
3. $D\Phi$ is injective

Measure-theoretic embedding ($\Phi\#$)

1. $\Phi\#$ is injective
2. $\Phi\#$ is smooth
3. $D(\Phi\#)$ is injective

Theorem (Our Main Result)
If $\Phi : M \to N$ is an embedding between differentiable manifolds, then the map $\Phi\# : \mathcal{P}_2(M) \to \mathcal{P}_2(N)$ is also an embedding.

[Botvinick-Oprea-Maulik-Y., 2024, arXiv:2409.08768]

Measure-Theoretic Embedding

Pointwise embedding (Φ)

1. Φ is injective
2. Φ is smooth
3. $D\Phi$ is injective

Measure-theoretic embedding ($\Phi\#$)

1. $\Phi\#$ is injective
2. $\Phi\#$ is smooth
3. $D(\Phi\#)$ is injective

Theorem (Our Main Result)
If $\Phi : M \to N$ is an embedding between differentiable manifolds, then the map $\Phi\# : \mathcal{P}_2(M) \to \mathcal{P}_2(N)$ is also an embedding.

Take-away message:

1. Takens' Thm. can be **generalized/lifted** to the space of probability measures.

[Botvinick-Oprea-Maulik-Y., 2024, arXiv:2409.08768]

Measure-Theoretic Embedding

Pointwise embedding (Φ)

1. Φ is injective
2. Φ is smooth
3. $D\Phi$ is injective

Measure-theoretic embedding ($\Phi\#$)

1. $\Phi\#$ is injective
2. $\Phi\#$ is smooth
3. $D(\Phi\#)$ is injective

Theorem (Our Main Result)
If $\Phi : M \to N$ is an embedding between differentiable manifolds, then the map $\Phi\# : \mathcal{P}_2(M) \to \mathcal{P}_2(N)$ is also an embedding.

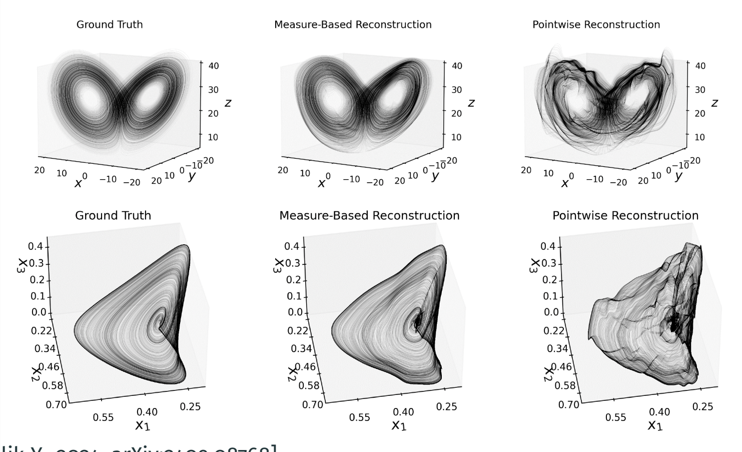
Take-away message:

1. Takens' Thm. can be **generalized/lifted** to the space of probability measures.
2. For the same embedding map Φ, we can learn it either based on **pointwise** data pairs or **measure** data pairs.

[Botvinick-Oprea-Maulik-Y., 2024, arXiv:2409.08768]

Numerical Example

$$\underbrace{\mathcal{L}_{\mathrm{p}}(\theta)}_{\text{pointwise loss}} = \frac{1}{N}\sum_{i=1}^{N}\|x_i - \mathcal{R}_\theta(\Phi(x_i))\|_2^2\,, \qquad \underbrace{\mathcal{L}_{\mathrm{m}}(\theta)}_{\text{measure-theoretic loss}} = \frac{1}{K}\sum_{i=1}^{K}\mathcal{D}(\mu_i, \mathcal{R}_\theta\#(\Phi\#\mu_i))\,.$$



[Botvinick-Oprea-Maulik-Y., 2024, arXiv:2409.08768]

The Distributional Koopman Operator

Koopman Operator K_t

Dynamics of the state on M:

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

Dynamics of observables $\hat{h} : M \to \mathbb{R}$ where $\hat{h} \in L^\infty(M)$:

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

Koopman Operator K_t

Dynamics of the state on M:

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

Dynamics of observables $\hat{h} : M \to \mathbb{R}$ where $\hat{h} \in L^\infty(M)$:

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

Introducing randomness $\omega \Longrightarrow$ *Random Dynamical System*: $\Phi_t(\omega, \cdot)$ where

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

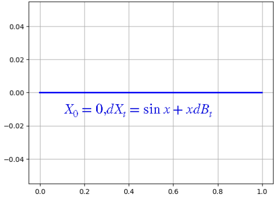
(For example, the solution X_t to a stochastic differential equation (SDE))

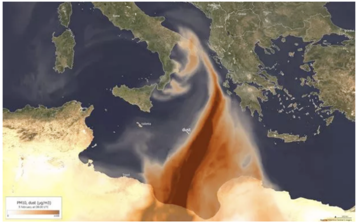
The Stochastic Koopman operator (SKO) average value of the observable at time:

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

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



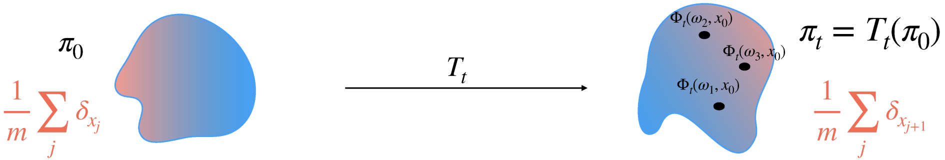


Dust plume data off the coast of Libia

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

Lagrangian vs Eulerian Views of the Dynamics

Lagrangian	Eulerian
$x \in M$	$\pi \in \mathscr{P}(M)$
$\hat{h} \in L^\infty(M)$	$h : \mathscr{P}(M) \to \mathbb{R}$ continuous and bounded
RDS: $\Phi : \mathbb{R} \times \Omega \times M \to M$	Transfer operator $T_t : \mathscr{P}(M) \to \mathscr{P}(M)$



[Oprea-Townsend-Y., 2025, MCSS]

The Distributional Koopman Operator (DKO)

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

Linearity ✓

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

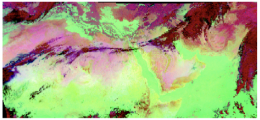
Invariant subspace $H_1 = \{h : \mathscr{P}(M) \to \mathbb{R} \text{ linear and bounded }\}$

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: $\mathscr{D}_t h(\pi) = \mathbb{E}_{X\sim\pi}\left[[\mathscr{S}_t \hat{h}](X)\right]$
- Same eigenvalues: $\mathscr{S}_t \hat{h} = \lambda \hat{h} \implies \mathscr{D}_t h = \lambda h$

[Oprea-Townsend-Y., 2025, MCSS]

Dust plume DUSTScan2022 data

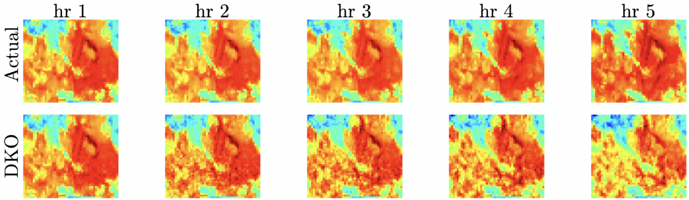


Dataset = hourly dust observations from SEVIRI on Meteosat-8

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

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

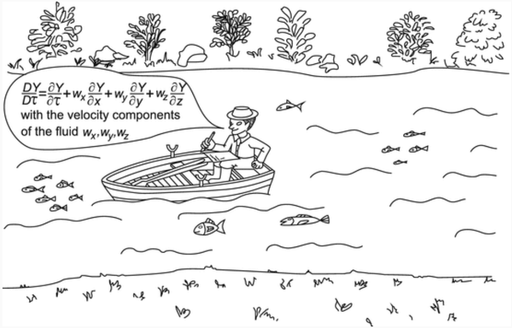
Use 28 days worth of data and predict the following 5 hours



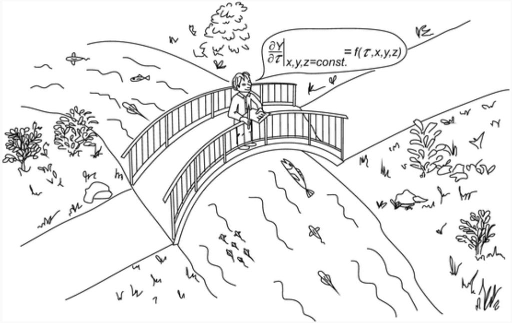
[Oprea-Townsend-Y., 2025, MCSS]

Conclusion

Conclusions



(a) Lagrangian view



(b) Eulerian view

[Bird-Stewart-Lightfoot, Transport Phenomena, 2002]

Conclusions

Summaries

- From **Lagrangian** to **Eulerian** to tackle chaos (ODE $\Longrightarrow$ PDE problem)
- Using **optimal transport** to study dynamical system
  1. Invariant measure matching
  2. Invariant measure in time-delay coordinate matching (for uniqueness)
  3. Generalize “pointwise” embedding to “measure-based” embedding
  4. Generalize “pointwise” Koopman to “measure-based” Koopman

Conclusions

Summaries

- From **Lagrangian** to **Eulerian** to tackle chaos (ODE $\Longrightarrow$ PDE problem)
- Using **optimal transport** to study dynamical system
  1. Invariant measure matching
  2. Invariant measure in time-delay coordinate matching (for uniqueness)
  3. Generalize "pointwise" embedding to "measure-based" embedding
  4. Generalize "pointwise" Koopman to "measure-based" Koopman

Outlook

There is great potential for using **optimal transport and other math tools studying probability measures** in *data-driven modeling of dynamical systems.*

Acknowledgments

Research support from

Thank you for the attention!