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

Yunan Yang, Cornell University

January 23, 2025

Kantorovich Initiative Seminar Series. Online.

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*.
— Measure-Theoretic Time-Delay Embedding. arXiv:2409.08768.
— Invariant Measures in Time-Delay Coordinates for Unique Dynamical System Identification. arXiv:2412.00589.

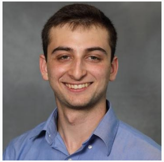
Collaborators

(a) Levon Nurbekyan (Emory)

(b) Robert Martin (ARL)

(c) Elisa Negrini (UCLA)

(d) Mirjeta Pasha (Virginia Tech)



(e) Jonah Botvinick-Greenhouse (Cornell)

(f) Maria Oprea (Cornell)

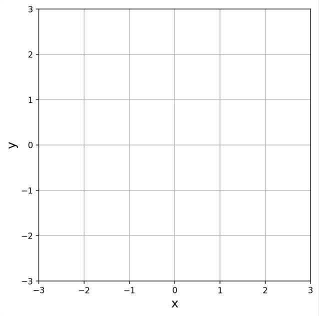
(g) Romit Malik (PSU)

Data-Driven Modeling of Dynamical Systems

Data-Driven Modeling for Dynamical System

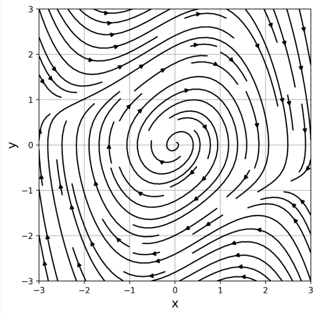
$$X$$

State space



$$\dot{x} = v(x)$$

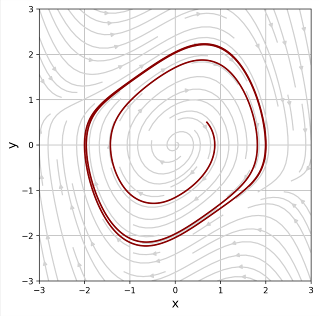
Evolution rule



data-driven modeling

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

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 [Brunton et al., 2016], [Schaeffer-Tran-Ward,2018]
- other basis functions, e.g., piecewise polynomials, RBFs, Fourier, etc.
- neural networks [many references], and so on,

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.

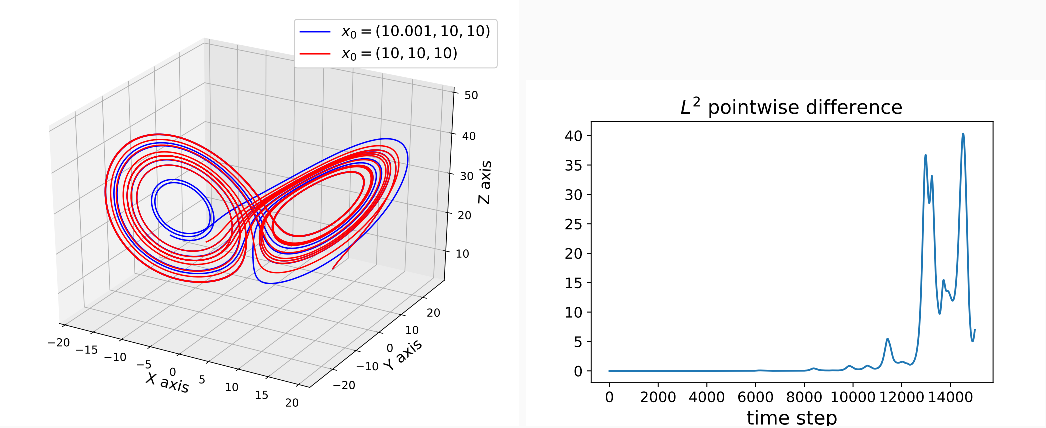


Figure: The comparison between $\mathbf{x}_0 = [10.001, 10, 10]$ and $\mathbf{x}_0 = [10, 10, 10]$.

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

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

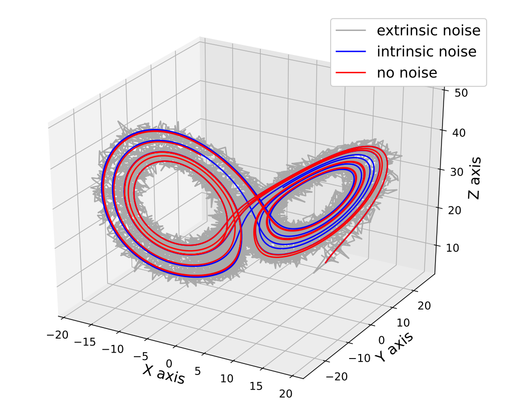


Figure: The comparison among the three cases.

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 $\{\mathbf{x_i}\}$

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

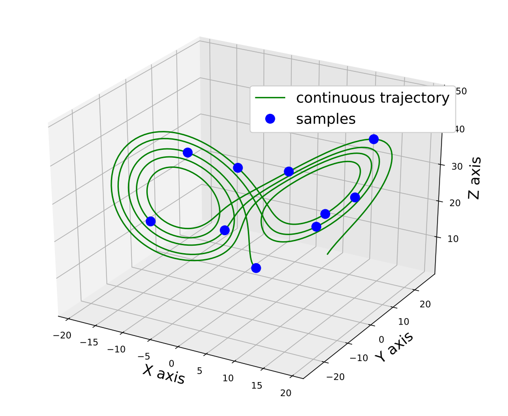


Figure: The continuous trajectory vs the samples

The Eulerian Approach

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

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

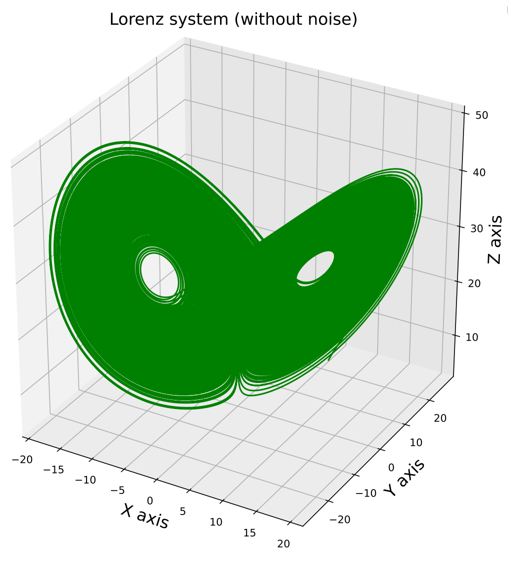
$\Downarrow$

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:

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

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

ρ^* : the observed occupation measure converted from time trajectories
$\rho(\theta)$: the distributional steady-state solution of the PDE
d : an appropriate metric that captures the essential differences, e.g., W_2 metric

The New Method — A PDE-Constrained Optimization Problem

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

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

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

ρ^* : the observed occupation measure converted from time trajectories
$\rho(\theta)$: the distributional steady-state solution of the PDE
d : an appropriate metric that captures the essential differences, e.g., W_2 metric

Data and forward problem are changed, but parameters remain the same.
The gain is to work with a much **More Stable** inverse problem!

The New Method — A PDE-Constrained Optimization Problem

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

$$\theta = \operatorname*{argmin}_{\theta} d(\rho^*, \rho(\theta)),$$

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

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

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

d : an appropriate metric that captures the essential differences, e.g., W_2 metric

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

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

Next: an objective function comparing distributions

Optimal Transport

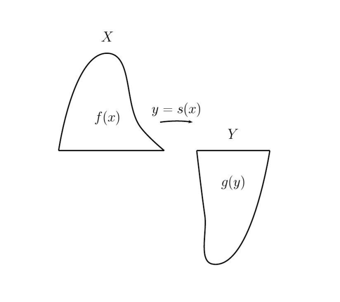


Figure: Proposed by Monge in 1781

- Monge (1781)
- Kantorovich (1975)
- Brenier, Caffarelli, Gangbo, McCann, Benamou, Otto, Villani, Figalli, etc. (1990s - present)

Optimal Transport

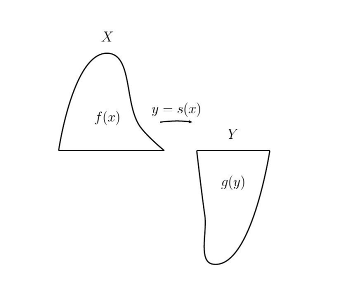


Figure: Proposed by Monge in 1781

- Monge (1781)
- Kantorovich (1975)
- Brenier, Caffarelli, Gangbo, McCann, Benamou, Otto, Villani, Figalli, etc. (1990s - present)

- **Data Assimilation** (Reich, Vidard, Bocquet...)
- **Hyperbolic Model Reduction** (Mula, Peherstorfer,Ravela)
- **Image Processing** (T. Chan, Peyré, C. B. Schönlieb...)
- **Inverse Problems** (Bao, Marzouk, Engquist, Singer, Y.,...)
- **Machine Learning** (Cuturi,Peyré, Solomon, ...)
- **Sampling** (Marzouk, Rigollet, Chewi, ...)
- **And more**

Definition of the Wasserstein Distance

For $f, g \in \mathcal{P}(\Omega)$ ($f, g \geq 0$ and $\int f = \int g = 1$), the Wasserstein distance is formulated as

$$W_p(f, g) = \left(\inf_{T \in \mathcal{M}} \int |x - T(x)|^p f(x) dx \right)^{\frac{1}{p}} \tag{1}$$

$\mathcal{M}$: the set of all maps that rearrange the distribution f into g.

The commonly used cases include $p = 1$ and $p = 2$.

Definition of the Wasserstein Distance

For $f, g \in \mathcal{P}(\Omega)$ ($f, g \geq 0$ and $\int f = \int g = 1$), the Wasserstein distance is formulated as

$$W_p(f, g) = \left(\inf_{T \in \mathcal{M}} \int |x - T(x)|^p f(x) dx \right)^{\frac{1}{p}} \tag{1}$$

$\mathcal{M}$: the set of all maps that rearrange the distribution f into g.

The commonly used cases include $p = 1$ and $p = 2$.

Properties of W_2 as the loss function

(1) Provide better optimization landscape for Nonlinear Inverse Problems:

$$\theta^* = \underset{\theta}{\text{argmin}}\, W_2^2(\rho(\theta), \rho^*)$$

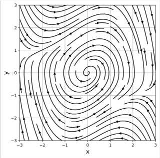
(2) Robust in Inversion with Noisy Data (equivalent to $\dot{H}^{-1}$ norm)

Recap: our approach from Lagrangian view to Eulerian perspective

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

?

$v(x)$



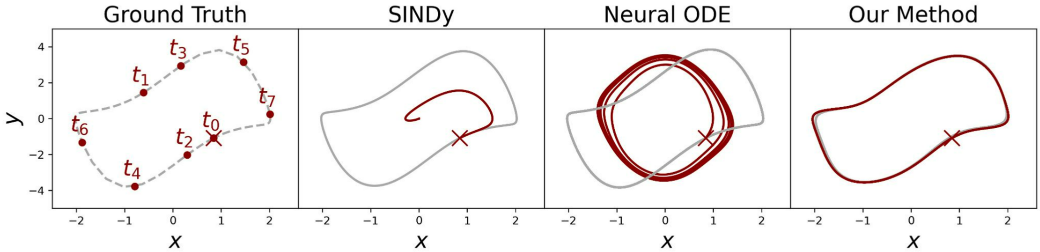
SINDy[1]	Shooting methods[2]	Neural ODEs[3]

- 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*.

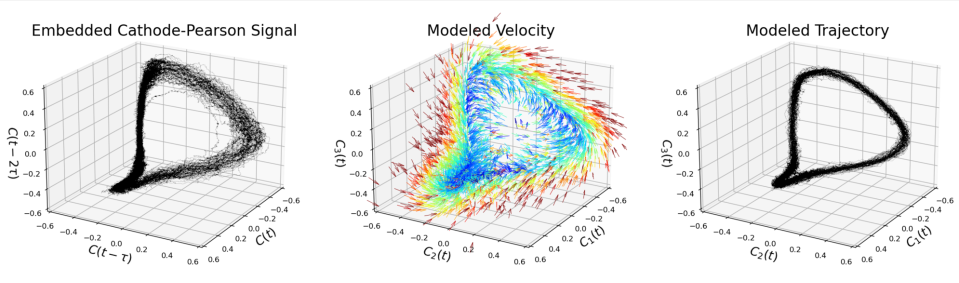
Comparison with Other Methods

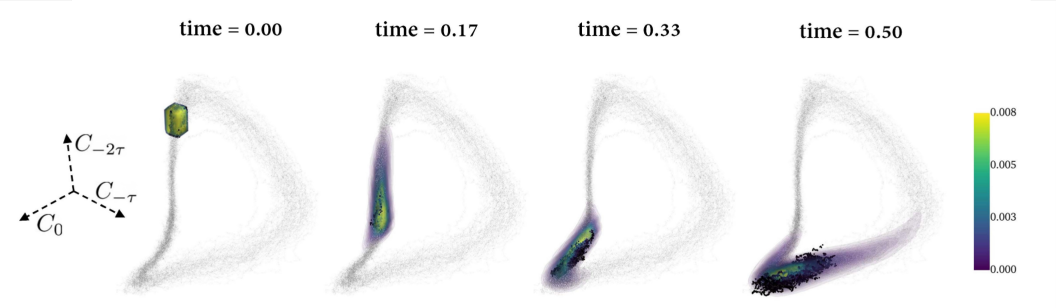


Method	Sampling Freq.	Wall-Clock Time (s)	Error
SINDy	10.00	$2 \cdot 10^{-2}$	$5.6 \cdot 10^{-3}$
Neural ODE	10.00	$5 \cdot 10^{2}$	$5.32 \cdot 10^{-3}$
Ours	10.00	$5 \cdot 10^{2}$	$1.14 \cdot 10^{-1}$

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	$5 \cdot 10^{2}$	$6.79 \cdot 10^{-2}$

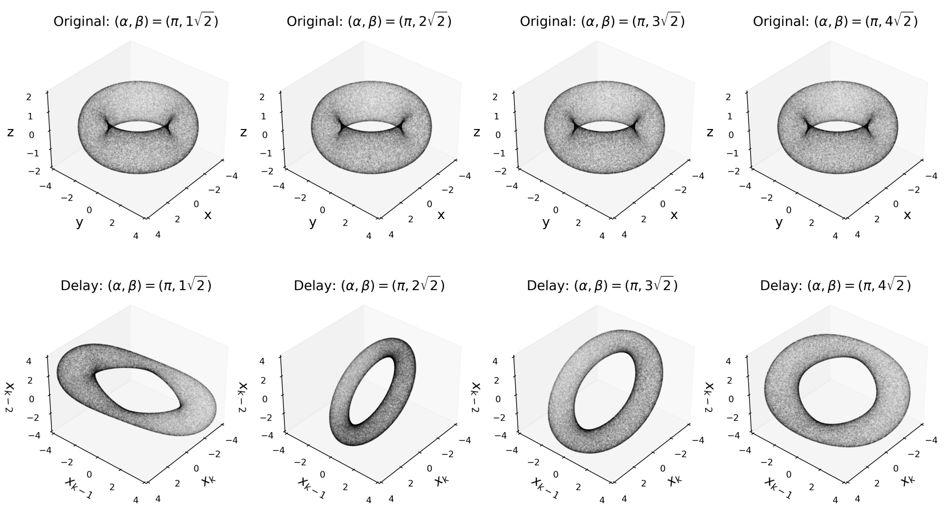
Application to Real-World Data: Hall-Effect Thruster





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

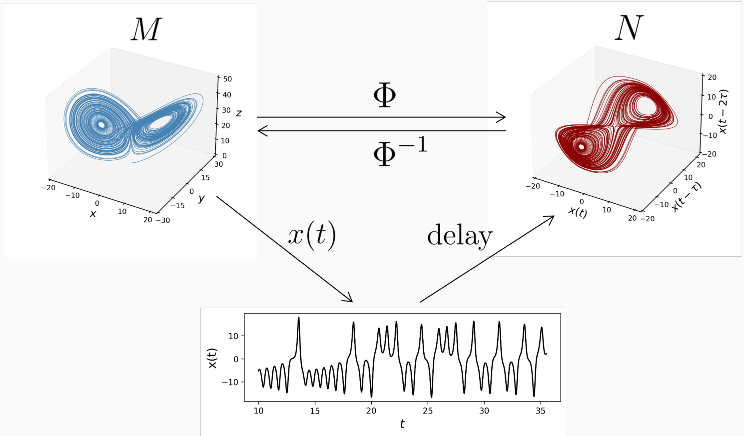
Theorem (Takens, 1981)

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 $\mathbb{R}^{2d+1}$.

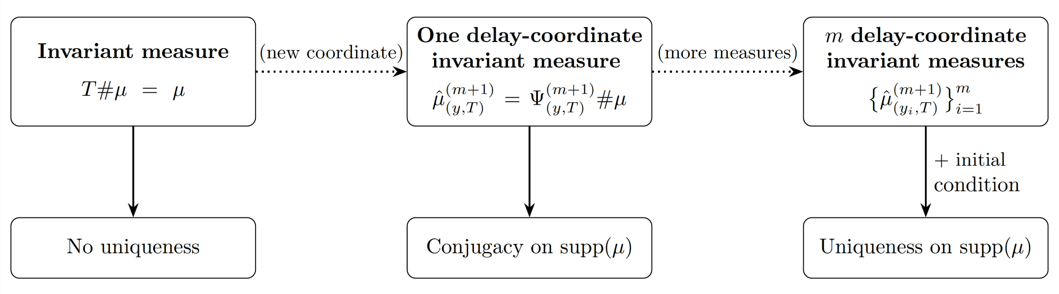
Takens' Embedding Theorem

Theorem (Takens, 1981)

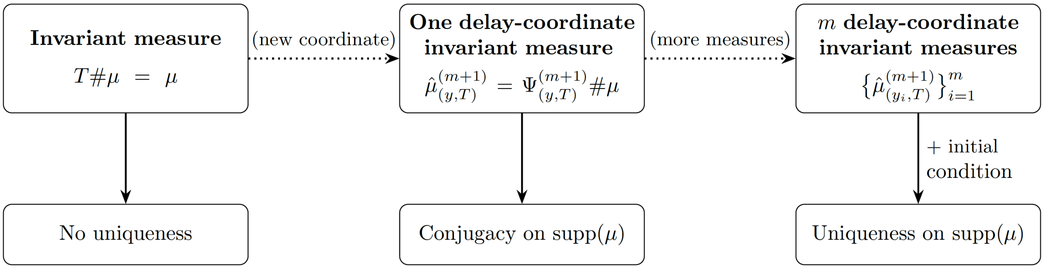
Let M be a compact manifold of dimension m. For pairs (y, T), where $T \in C^2(M, M)$ and $y \in C^2(M, \mathbb{R})$, it is a generic property that the mapping $\Phi_{(y,T)} : M \to N \subseteq \mathbb{R}^{2d+1}$ given by $\Phi_{(y,T)}(\mathbf{x}) := (y(\mathbf{x}), y(T(\mathbf{x})), \ldots, y(T^{2m}(\mathbf{x})))$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



Invariant Measures in Time-Delay Coordinates for Uniqueness

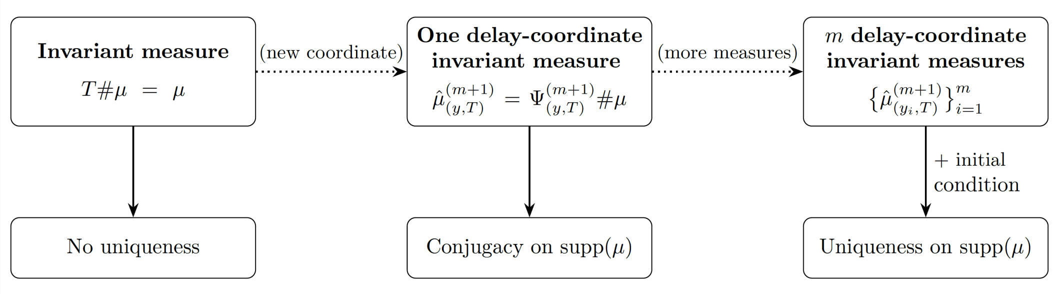


Invariant Measures in Time-Delay Coordinates for Uniqueness



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 $y \in C^1(U, \mathbb{R})$.

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 $y \in C^1(U, \mathbb{R})$.

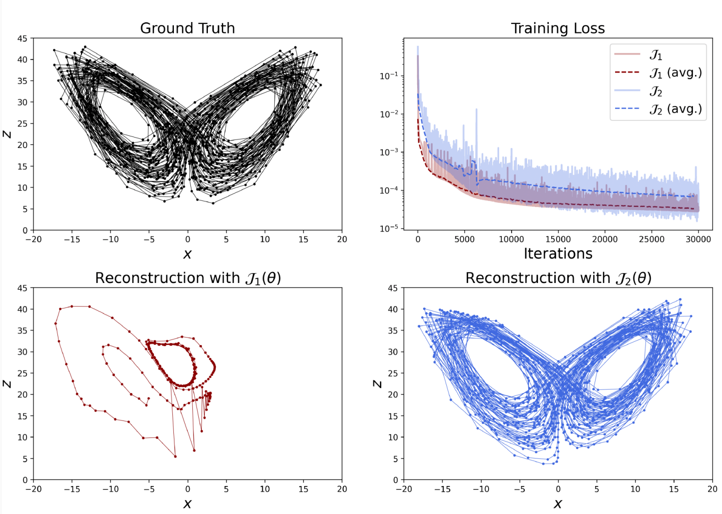
Theorem 2. The conditions below imply that $T = S$ on supp(μ), for a.e. $Y \in C^1(U, \mathbb{R}^m)$:

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.

Numerical Example

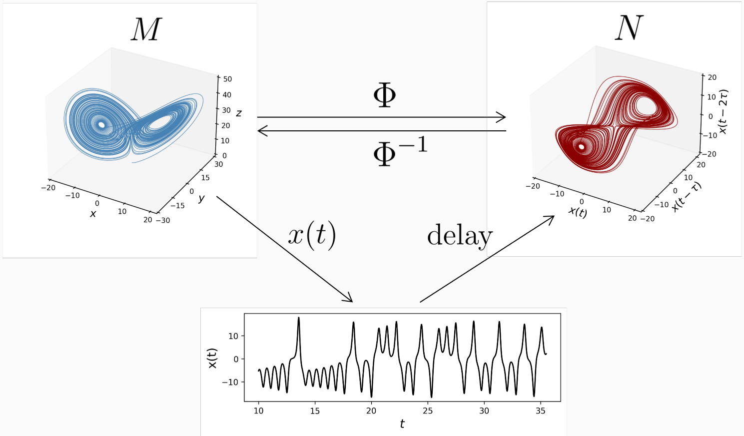
$$\mathcal{J}_1(\theta) := \mathcal{D}(T_\theta \# \mu^*, T^* \# \mu^*), \qquad \mathcal{J}_2(\theta) := \mathcal{D}(T_\theta \# \mu^*, T^* \# \mu^*) + \mathcal{D}(\Psi_\theta \# \mu^*, \Psi^* \# \mu^*).$$

Ψ_θ is the delay map based on T_θ, and Ψ^* is the true delay map.



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

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

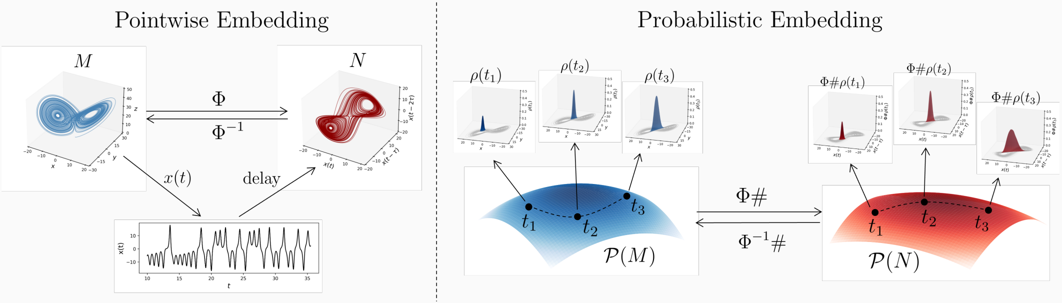
- Challenges: Takens' Theorem no longer applies when dynamics have noise.
- Can we **lift** the statement to the **space of probability measures**?

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?

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?



Measure-Theoretic Embedding

Pointwise embedding (Φ)

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

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

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

Definition (Differentiability of operator $\mathcal{P}_2(M) \to \mathcal{P}_2(N)$)
A map $\Psi : \mathcal{P}_2(M) \to \mathcal{P}_2(N)$ is differentiable if for all $\mu \in \mathcal{P}_2(M)$ there is a bounded linear operator $d\Psi_\mu : T_\mu\mathcal{P}_2(M) \to T_{\Psi(\mu)}\mathcal{P}_2(N)$ s.t. for any differentiable curve $t \mapsto \mu_t$ through μ, the curve $t \mapsto \Psi(\mu_t)$ is differentiable with velocity v_t and $d\Psi_{\mu_t}(v_t) = \frac{d}{dt}\Psi(\mu_t)$.

Measure-Theoretic Embedding

Pointwise embedding (Φ)

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

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

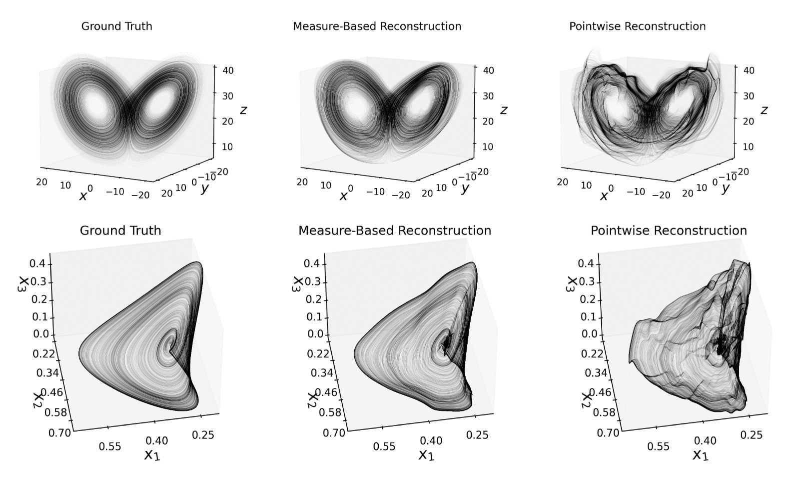
1. $\Phi\#$ is injective
2. $\Phi\#$ is smooth
3. $D(\Phi\#)$ is injective

Definition (Differentiability of operator $\mathcal{P}_2(M) \to \mathcal{P}_2(N)$)
A map $\Psi : \mathcal{P}_2(M) \to \mathcal{P}_2(N)$ is differentiable if for all $\mu \in \mathcal{P}_2(M)$ there is a bounded linear operator $d\Psi_\mu : T_\mu \mathcal{P}_2(M) \to T_{\Psi(\mu)}\mathcal{P}_2(N)$ s.t. for any differentiable curve $t \mapsto \mu_t$ through μ, the curve $t \mapsto \Psi(\mu_t)$ is differentiable with velocity v_t and $d\Psi_{\mu_t}(v_t) = \frac{d}{dt}\Psi(\mu_t)$.

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.

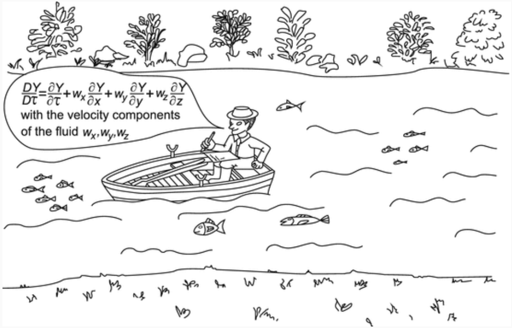
Numerical Example

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

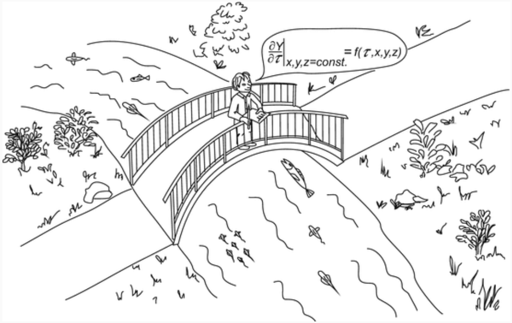


Conclusion

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
  3. Generalize pointwise embedding to measure-theoretic embedding

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
  3. Generalize pointwise embedding to measure-theoretic embedding

Outlook
There is great potential for using optimal transport in data-driven modeling of dynamical systems.

Acknowledgments

Research support from

Thank you for the attention!