Measure-Theoretic Time-Delay Embedding

Jonah Botvinick-Greenhouse[1], Maria Oprea[1], Romit Maulik[2], Yunan Yang[3]

[1]Center for Applied Mathematics, Cornell University
[2]College of Information Sciences and Technology, Pennsylvania State University
[3]Department of Mathematics, Cornell University

SIAM Conference on Mathematics of Data Science (MDS24)
MS53: Advances in Fast and Scalable Bayesian Inference

October 23, 2024

Motivation

- Consider the regression task of learning $\Phi : \mathbb{R}^d \to \mathbb{R}^m$ from paired data $\{(x_i, y_i)\}_{i=1}^N$.
- The data may be **noisy**, i.e., $x_i \sim \mu_{x_i} \in \mathcal{P}(\mathbb{R}^d)$ and $y_i \sim \nu_{y_i} \in \mathcal{P}(\mathbb{R}^m)$.
- We will **lift** the problem to the space of probability measures.

$$
\begin{tikzcd}
\mu_{x_i} \in \mathcal{P}(\mathbb{R}^d) \arrow[r, "\nu_{y_i} = \Phi\#\mu_{x_i}"] & \nu_{y_i} \in \mathcal{P}(\mathbb{R}^m) \\
x_i \in \mathbb{R}^d \arrow[u, "x_i \sim \mu_{x_i}"] \arrow[r, "y_i = \Phi(x_i)"'] & y_i \in \mathbb{R}^m \arrow[u, "y_i \sim \nu_{y_i}"']
\end{tikzcd}
$$

- It holds that $D_f(\Phi\#\mu_{x_i} \mid \Phi\#\tilde{\mu}_{x_i}) = D_f(\mu_{x_i} \mid \tilde{\mu}_{x_i})$ if Φ is invertible.[1]

[1] Li, Q., Oprea, M., Wang, L., Yang, Y. (2024). Stochastic Inverse Problem: stability, regularization and Wasserstein gradient flow. arXiv preprint arXiv:2410.00229.

In this talk...

...we study the problem of learning a dynamical system's full-state reconstruction map as a pushforward map between probability spaces.

(Part 1): Dynamical Systems & Time-Delay Embedding

(Part 2): Measure-Theoretic Embeddings

(Part 3): Computational Approach & Numerical Examples

(Part 4): Conclusions

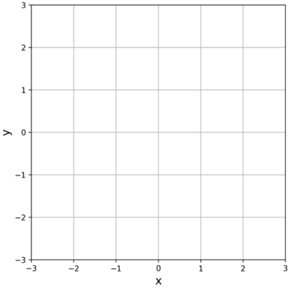
(Part 1): Dynamical Systems & Time-Delay Embedding

Dynamical Systems

$$X$$

$\uparrow$

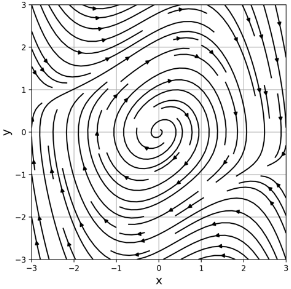
State space



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

$\uparrow$

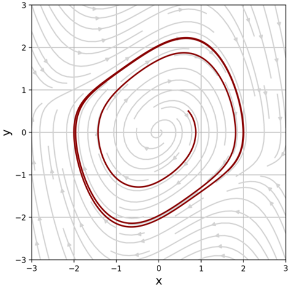
Evolution rule



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

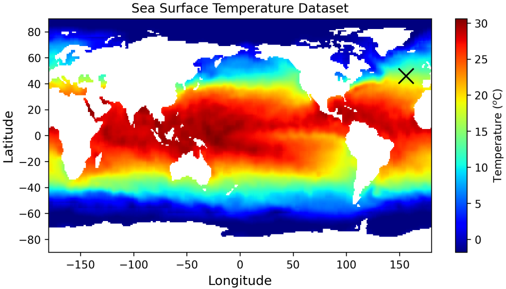
$\uparrow$

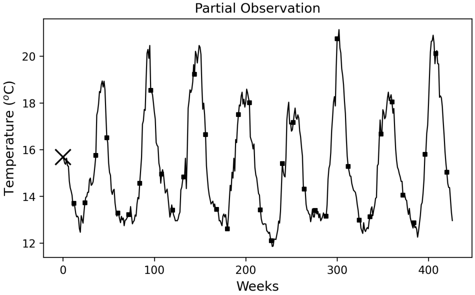
Trajectory samples



Partial Observability

- Rather than observing the **full state** $\{x(t_k)\}_{k=0}^{N-1} \subseteq \mathbb{R}^d$, experimentalists instead observe $\{y(x(t_k))\}_{k=0}^{N-1} \subseteq \mathbb{R}^q$, where $y : \mathbb{R}^d \to \mathbb{R}^q$ is some **measurement function.**



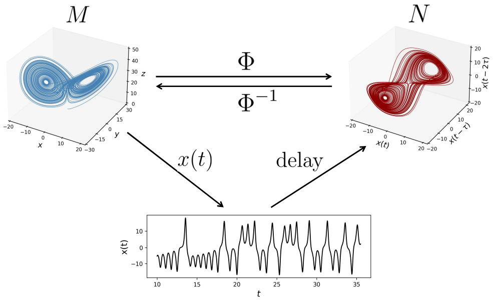


- *How can we reconstruct the full state from these partial observations?*

Takens' Embedding Theorem

Theorem

Let M be a d-dimensional smooth compact manifold, let $m \geq 2d+1$, and choose $\tau > 0$. If $v \in C^2(M, M)$, and $h \in C^2(M, \mathbb{R})$, then $\Phi_{h,\phi_\tau}(x) := (h(x), h(\phi_\tau(x)), \ldots, h(\phi_{(m-1)\tau}(x)))$ is generically an embedding of M, where ϕ_τ is the time-τ flow map of v.

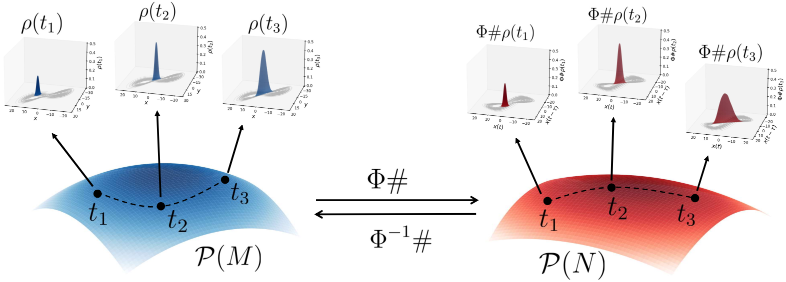


[1] Takens, F. (2006, October). Detecting strange attractors in turbulence. In Dynamical Systems and Turbulence, Warwick 1980: proceedings of a symposium held at the University of Warwick 1979/80 (pp. 366-381). Berlin, Heidelberg: Springer Berlin Heidelberg.

(Part 2): Measure-Theoretic Embeddings

Measure-Theoretic Embeddings

- **Challenge:** Takens' theorem no longer applies when dynamics are subject to noise.
- **Past Work:** Statistical methods for i.i.d. extrinsic noise[2], embedding with stochasticity[3].
- **Motivation:** *Can we lift the statement to the space of probability measures?*



[2] Martin Casdagli, Stephen Eubank, J Doyne Farmer, and John Gibson. State space reconstruction in the presence of noise. Physica D: Nonlinear Phenomena, 51(1-3):52–98, 1991.

[3] Stark, J., Broomhead, D. S., Davies, M. E., Huke, J. (1997). Takens embedding theorems for forced and stochastic systems. Nonlinear Analysis: Theory, Methods Applications, 30(8), 5303-5314.

Measure-Theoretic Embeddings (cont.)

Definition (Pushforward of a Measure)

If $\Phi : M \to N$ is Borel measurable and $\mu \in \mathcal{P}(M)$, then $\Phi\#\mu \in \mathcal{P}(N)$ is the measure defined by $(\Phi\#\mu)(B) := \mu(\Phi^{-1}(B))$, for all Borel sets $B \subseteq N$.

If $\Phi : M \to N$ is an embedding is $\Phi\# : \mathcal{P}(M) \to \mathcal{P}(N)$ as well? In what sense?

Pointwise embedding

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

Measure-theoretic embedding

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

How should we define the derivative of a map between $\mathcal{P}(M)$ and $\mathcal{P}(N)$?

Main Theoretical Result

Definition (Tangent space, Differentiable Curve, & Differentiable Map)

- Given $\mu \in \mathcal{P}_2(M)$, its **tangent space** is
$$T_\mu \mathcal{P}_2(M) := \{v \in L^2_\mu(TM) : \|v + w\|_{L^2_\mu} \geq \|w\|_{L^2_\mu}, \forall w \in L^2_\mu(TM) \text{ s.t. } \nabla \cdot (w\mu) = 0\}.$$
- A **curve** $t \mapsto \mu_t$ is **differentiable** if there exist $v_t \in L^2_{\mu_t}(TM)$, such that $\partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0$ and $\int_0^1 \|v_t\|_{L^2_\mu} dt < \infty$, and $\frac{d}{dt}\mu_t := v_t$.
- 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 $d\Psi_\mu : T_\mu \mathcal{P}_2(M) \to T_{\Psi(\mu)} \mathcal{P}_2(N)$, such that for any differentiable curve $t \mapsto \mu_t$ through μ, the curve $t \mapsto \Psi(\mu_t)$ is differentiable and $d\Psi_{\mu_t} = \frac{d}{dt}\Psi(\mu_t)$.

Theorem (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.

(Part 3): Computational Approach & Numerical Examples

Measure-Theoretic Inversion

- **Setting:** We observe samples $\{(x_i, \Phi(x_i))\}_{i=1}^{N} \subseteq \mathbb{R}^d \times \mathbb{R}^m$
- **Goal:** Learn the inverse embedding map Φ^{-1}.
- **Pointwise Loss:**[4,5] Parameterize $\mathcal{R}_\theta : \mathbb{R}^m \to \mathbb{R}^d$ by $\theta \in \Theta$ and solve

$$\min_{\theta\in\Theta} \mathcal{L}_{\mathrm{p}}(\theta), \qquad \mathcal{L}_{\mathrm{p}}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \|x_i - \mathcal{R}_\theta(\Phi(x_i))\|_2^2.$$

- **Measure-Theoretic Loss**: Construct measure data

$$\{(\mu_i, \Phi\#\mu_i)\}_{i=1}^{K} \subseteq \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^m)$$

  and solve

$$\min_{\theta\in\Theta} \mathcal{L}_{\mathrm{m}}(\theta), \qquad \mathcal{L}_{\mathrm{m}}(\theta) = \frac{1}{K}\sum_{i=1}^{K} \mathfrak{D}(\mu_i, \mathcal{R}_\theta\#(\Phi\#\mu_i)).$$

[4] Bakarji, J., Champion, K., Nathan Kutz, J., Brunton, S. L. (2023). Discovering governing equations from partial measurements with deep delay autoencoders. Proceedings of the Royal Society A, 479(2276), 20230422.

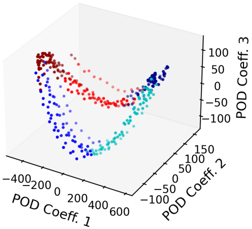
[5] Charles D Young and Michael D Graham. Deep learning delay coordinate dynamics for chaotic attractors from partial observable data. Physical Review E, 107(3):034215, 2023.

Constructing the Measure Data

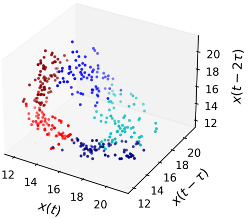
1. Partition $\{\Phi(x_i)\}_{i=1}^{N}$ into Voronoi cells $\{C_i\}_{i=1}^{K}$
2. For each $1 \leq i \leq K$, define

$$\mu_i := \frac{1}{N_i} \sum_{j=1}^{N_i} \delta_{x_j^i} \in \mathcal{P}(\mathbb{R}^d),$$

where N_i is the size of C_i and $\{x_j^i\}_{j=1}^{N_i} \subseteq \mathbb{R}^d$ are the samples satisfying $\Phi(x_j^i) \in C_i$.



(a) Visualization of $\Phi\#\mu_i$



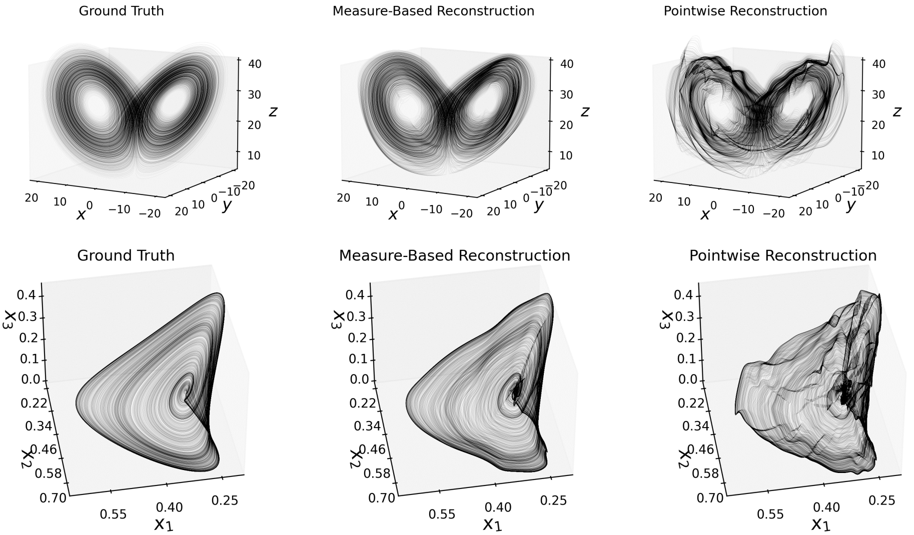
(b) Visualization of μ_i

Discussion: Pointwise ($\mathcal{L}_{\mathrm{p}}$) vs. Measure ($\mathcal{L}_{\mathrm{m}}$) Loss

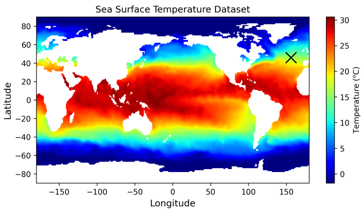
$$\underbrace{\mathcal{L}_{\mathrm{p}}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \|x_i - \mathcal{R}_\theta(\Phi(x_i))\|_2^2}_{\text{pointwise loss}} \qquad \underbrace{\mathcal{L}_{\mathrm{m}}(\theta) = \frac{1}{K}\sum_{i=1}^{K} \mathfrak{D}(\mu_i, \mathcal{R}_\theta \#(\Phi \# \mu_i))}_{\text{measure loss}}$$

1. As the diameter of each cell C_i decreases to zero, $\mathcal{L}_{\mathrm{m}}$ reduces to $\mathcal{L}_{\mathrm{p}}$.
2. A minimizer of $\mathcal{L}_{\mathrm{m}}$ may not minimize $\mathcal{L}_{\mathrm{p}}$.
3. The spectral bias of neural network training will favor smooth minimizers of $\mathcal{L}_{\mathrm{m}}$, while noisy observations cause the minimizers of $\mathcal{L}_{\mathrm{p}}$ to be oscillatory.

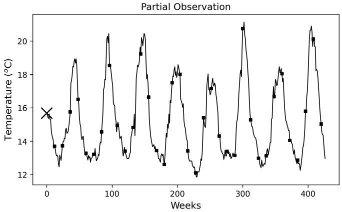
Numerical Example: Noisy Attractor Reconstruction



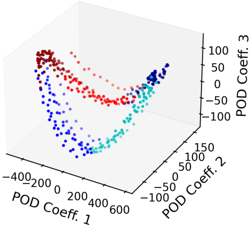
Numerical Example: NOAA SST Reconstruction



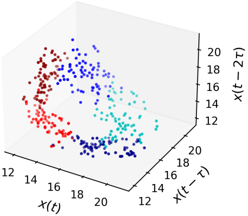
(a) NOAA SST dataset snapshot



(b) SST observation at $(156, 40)$



(c) POD training data

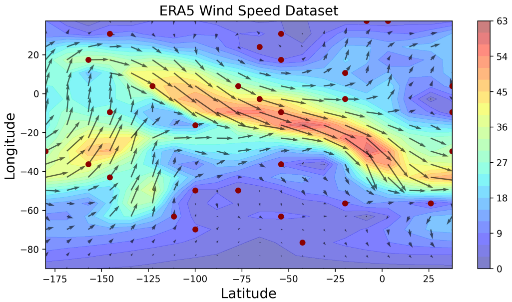


(d) delayed training data

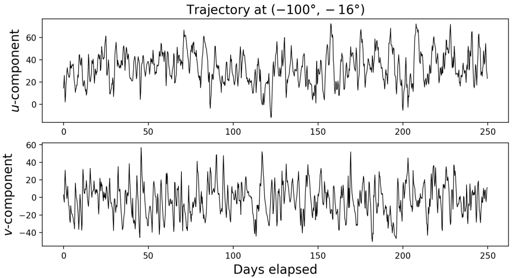
Numerical Example: NOAA SST Reconstruction (Cont.)

Method \ MSE	Initial	10^3 iters	5×10^3 iters	2×10^4 iters
Measure	13.81 ± 0.98	0.72 ± 0.01	0.73 ± 0.02	0.80 ± 0.02
Pointwise	13.95 ± 1.39	0.70 ± 0.01	0.82 ± 0.01	1.31 ± 0.02

ERA5 Wind Field Reconstruction

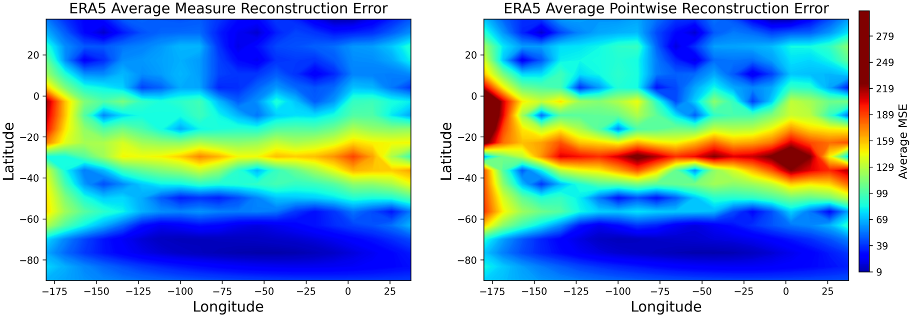


(a) ERA5 wind speed snapshot



(b) Example trajectory from ERA5

ERA5 Wind Field Reconstruction (cont.)



MSE / Method	Initial	10^3 iterations	10^4 iterations	2×10^4 iterations
Gaussian MMD	158.15 ± 0.27	61.04 ± 1.59	61.22 ± 0.54	67.91 ± 0.49
Energy MMD	158.15 ± 0.27	49.61 ± 0.48	67.85 ± 0.41	78.91 ± 0.37
Pointwise	158.15 ± 0.27	51.00 ± 0.19	75.87 ± 0.26	84.85 ± 0.25

(Part 4): Conclusions

Conclusions

1. **Motivation:** The assumptions of Takens' theorem do not hold when data is noisy.

2. **Theory:** We lift the statement of the classic theorem to the space of probability measures, demonstrating that the pushforward of the delay map is an embedding.

3. **Computation:** Based on our theory, we developed a robust computational routine for learning the inverse embedding function from data.

Thank you!

J., Oprea, M., Maulik, R., Yang, Y. (2024). Measure-Theoretic Time-Delay Embedding. arXiv preprint arXiv:2409.08768.

https://github.com/jrbotvinick/Measure-Theoretic-Time-Delay-Embedding