Optimal Transport for Learning Chaotic Dynamics via Invariant Measures

Yunan Yang, Institute for Theoretical Studies, ETH Zürich

September 28, 2022

— Joint work with **Levon Nurbekyan** (UCLA), **Elisa Negrini** (UCLA), **Robert Martin** (formerly, Air Force Research Laboratory, currently, U.S. Army Research Office), **Mirjeta Pasha** (Tufts University)
The paper: `https://arxiv.org/pdf/2104.15138.pdf` (in press for SIADS)

— Joint work with **Jonah Botvinick-Greenhouse** (Cornell CAM)

The Third Symposium on Machine Learning and Dynamical System, The Fields Institute, Sep 26-30, 2022

Learning Dynamical Systems

Parameter Identification

Parameter identification for (chaotic) dynamical systems is important in many applied areas. For example, finding σ, ρ, β for the Lorenz-63 System.

$$\begin{cases} \dot{x} &= \sigma(y - x), \\ \dot{y} &= x(\rho - z) - y, \\ \dot{z} &= xy - \beta z, \end{cases} \tag{1}$$

given time trajectory data $\{\mathbf{x}(t_0), \mathbf{x}(t_1), \mathbf{x}(t_2), \ldots, \mathbf{x}(t_N)\}$.

Similar Extensions

The unknown parameters do not have to be physical. 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.

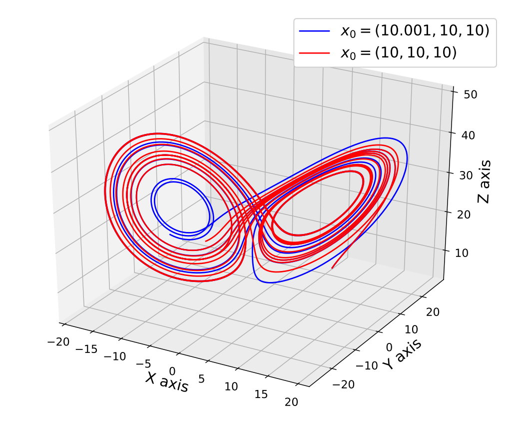
Forward and Inverse Problems

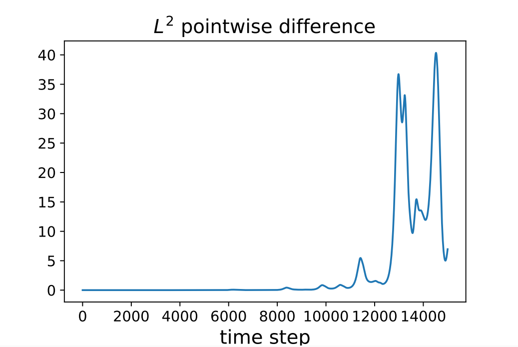
The modeling step: given the parameter ***m*** and the forward dynamical system, compute the time trajectories.

The inversion step: given the time trajectories and the dynamical system, reconstruct the model parameter ***m***.

Unique Challenges for Chaotic Systems

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





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

Unique Challenges for Chaotic Systems

Challenge Two: The time trajectories contain noise.

No noise

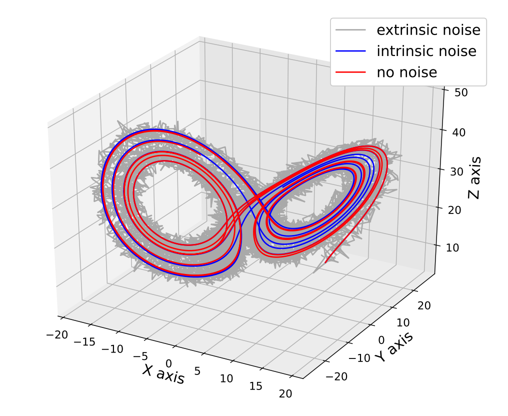
$$\dot{\mathbf{x}} = f(\mathbf{x}).$$

Extrinsic noise

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

Intrinsic noise

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



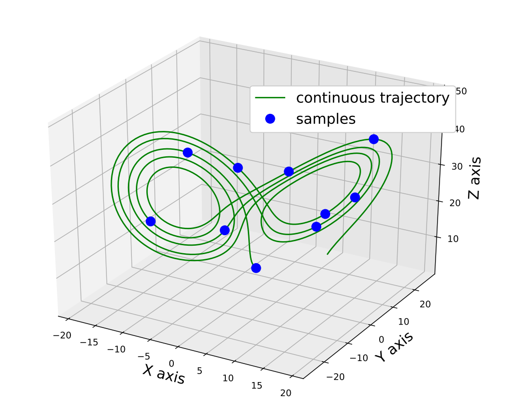
The comparison among the three cases.

Unique Challenges for Chaotic Systems

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 $\{\mathbf{x_i}\}$

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



The continuous trajectory vs the samples

The Eulerian Approach — Data

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 time trajectory with $t \in [0, T]$, starting with $\mathbf{x}(0) = x$, and $\mu_{x,T}$ is called the *occupation measure*. A special observable $f(\mathbf{x}) = \mathbb{1}_B(\mathbf{x})$.

The Eulerian Approach — Data

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 time trajectory with $t \in [0, T]$, starting with $\mathbf{x}(0) = x$, and $\mu_{x,T}$ is called the *occupation measure*. A special observable $f(\mathbf{x}) = \mathbb{1}_B(\mathbf{x})$.

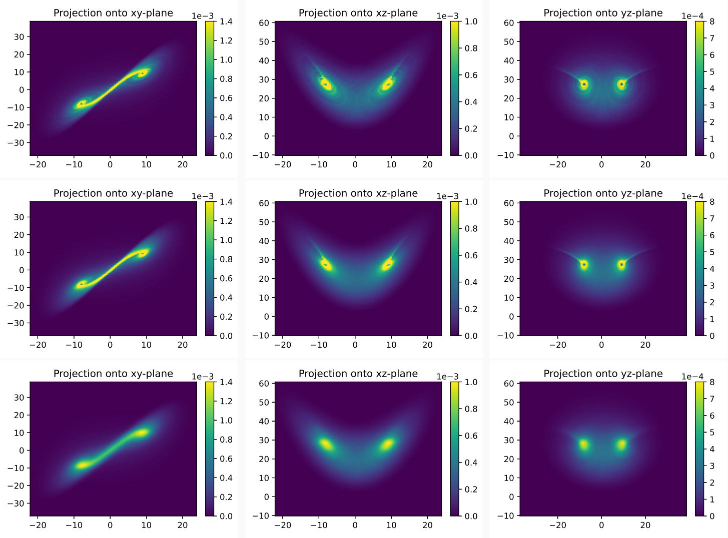
We call μ^* a physical measure, related to Sinai-Ruelle-Bowen (SRB) measures, if

$$\lim_{T\to\infty} \mu_{x,T} = \mu^*$$

for $x \in U$, and $\mathcal{L}^d(U) > 0$ [Young, 2002].

The idea: take μ^* as observation data instead of the trajectory $\mathbf{x}(t)$.

The Stable Invariant Measures



x-y, x-z and y-z projections of 3D invariant measures.

Top: no noise
Middle: w/ intrinsic noise
Bottom: w/ extrinsic noise

The Eulerian Approach — Model

By the definition of the physical measures μ^* [Lai-Sang Young, 2002], we have

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(x(t))dt = \int_{\mathbb{R}^d} f(x)d\mu^*(x), \quad f \in C_c^\infty(\mathbb{R}^d),$$

and $x(0) \in S$ for some S of positive Lebesgue measure.

The Eulerian Approach — Model

By the definition of the physical measures μ^* [Lai-Sang Young, 2002], we have

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(x(t))dt = \int_{\mathbb{R}^d} f(x)d\mu^*(x), \quad f \in C_c^\infty(\mathbb{R}^d),$$

and $x(0) \in S$ for some S of positive Lebesgue measure.

Thus, by choosing $f(x) = \nabla\phi(x) \cdot v(x)$ for some $\phi \in C_c^\infty(\mathbb{R}^d)$ (assuming $v \in C^\infty$), we obtain

$$\int_{\mathbb{R}^d} \nabla\phi(x) \cdot v(x)d\mu^*(x) = \lim_{T\to\infty} \frac{1}{T}\int_0^T \nabla\phi(x(t)) \cdot v(x(t))dt = \lim_{T\to\infty} \frac{1}{T}\int_0^T \nabla\phi(x(t)) \cdot \dot{x}(t)dt$$
$$= \lim_{T\to\infty} \frac{1}{T}\left(\phi(x(T)) - \phi(x(0))\right) = 0.$$

The Eulerian Approach — Model

By the definition of the physical measures μ^* [Lai-Sang Young, 2002], we have

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(x(t))dt = \int_{\mathbb{R}^d} f(x)d\mu^*(x), \quad f \in C_c^\infty(\mathbb{R}^d),$$

and $x(0) \in S$ for some S of positive Lebesgue measure.

Thus, by choosing $f(x) = \nabla\phi(x) \cdot v(x)$ for some $\phi \in C_c^\infty(\mathbb{R}^d)$ (assuming $v \in C^\infty$), we obtain

$$\begin{aligned}\int_{\mathbb{R}^d} \nabla\phi(x) \cdot v(x)d\mu^*(x) &= \lim_{T\to\infty} \frac{1}{T} \int_0^T \nabla\phi(x(t)) \cdot v(x(t))dt = \lim_{T\to\infty} \frac{1}{T} \int_0^T \nabla\phi(x(t)) \cdot \dot{x}(t)dt \\ &= \lim_{T\to\infty} \frac{1}{T} (\phi(x(T)) - \phi(x(0))) = 0.\end{aligned}$$

This shows that μ^* is the stationary distributional solution to

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

From Lagrangian to Eulerian

$$\dot{\mathbf{x}} = f(\mathbf{x}) := v(\mathbf{x}),$$

$$\Downarrow$$

Occupation measure

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

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

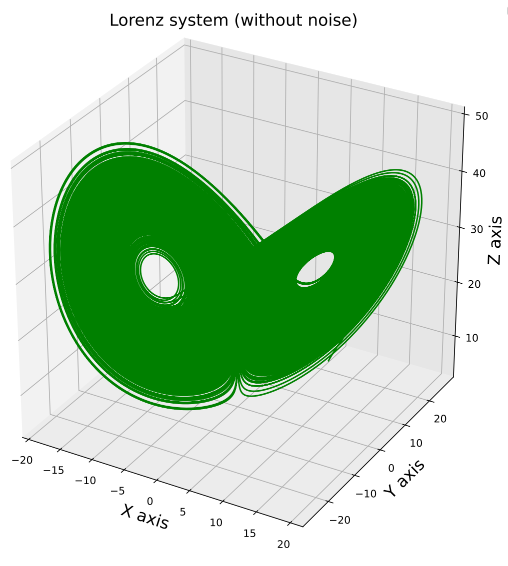
$$\Downarrow$$

physical measure μ^*

$$\Downarrow$$

Stationary distributional solutions of

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



Instead of

The **nonlinear modeling** step: given the parameter and the forward dynamical system (an ODE model), compute the time trajectories.

The **inversion** step: given the time trajectories and the dynamical system, reconstruct the model parameter.

We Now Regard the Problem As

The **nonlinear modeling** step: given the parameter and Continuity Equation (a PDE model), compute steady-state distribution.

The **inversion** step: given the observed occupation measure and the Continuity Equation, reconstruct the parameter.

The 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

The 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!

Highlights in Solving the Forward Problem

- Introduce a new PDE forward model for physical measures

$$\nabla \cdot \big(v(\mathbf{x}, \theta)\rho(\mathbf{x}, t)\big) = 0$$

Highlights in Solving the Forward Problem

- Introduce a new PDE forward model for physical measures

$$\nabla \cdot \big(v(\mathbf{x}, \theta)\rho(\mathbf{x}, t)\big) = 0$$

- Build a CFD discretization (finite-volume method) of the forward model via **first-order upwind scheme** that preserves positivity and mass conservation.

$$M(\theta)\rho = \rho, \quad \rho \cdot \mathbf{1} = 1\,.$$

We thus created a Perron–Frobenius operator M discretized in space.

Highlights in Solving the Forward Problem

- Introduce a new PDE forward model for physical measures

$$\nabla \cdot \big(v(\mathbf{x}, \theta) \rho(\mathbf{x}, t) \big) = 0$$

- Build a CFD discretization (finite-volume method) of the forward model via **first-order upwind scheme** that preserves positivity and mass conservation.

$$M(\theta)\rho = \rho, \quad \rho \cdot \mathbf{1} = 1\,.$$

  We thus created a Perron–Frobenius operator M discretized in space.
- Apply **teleportation** regularization from Google's PageRank [Gleich, 2015]

$$M_\epsilon(\theta) = (1 - \epsilon) M(\theta) + \frac{\epsilon}{n} \mathbf{1}\,\mathbf{1}^\top\,.$$

  Note that M_ϵ is now column-stochastic matrix with **strictly positive** entries, where the Perron–Frobenius theorem applies.

Highlights in Solving the Forward Problem

- Introduce a new PDE forward model for physical measures

$$\nabla \cdot \big(v(\mathbf{x}, \theta)\rho(\mathbf{x}, t)\big) = 0$$

- Build a CFD discretization (finite-volume method) of the forward model via **first-order upwind scheme** that preserves positivity and mass conservation.

$$M(\theta)\rho = \rho, \quad \rho \cdot \mathbf{1} = 1\,.$$

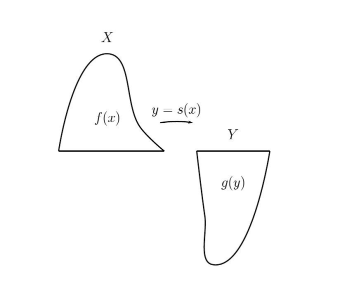
  We thus created a Perron–Frobenius operator M discretized in space.
- Apply **teleportation** regularization from Google's PageRank [Gleich, 2015]

$$M_\epsilon(\theta) = (1-\epsilon)M(\theta) + \frac{\epsilon}{n}\mathbf{1}\,\mathbf{1}^\top\,.$$

  Note that M_ϵ is now column-stochastic matrix with **strictly positive** entries, where the Perron–Frobenius theorem applies.
- Find the solution $\rho(\theta)$ to $M(\theta)\rho = \rho$ based on sparse linear solvers.

Optimal Transport for Data Fitting

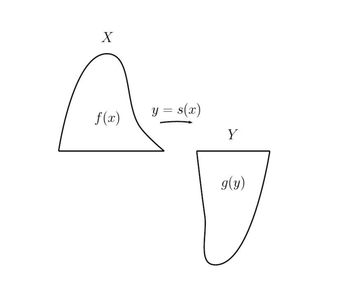
Optimal Transport



Proposed by Monge in 1781

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

Optimal Transport



Proposed by Monge in 1781

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

- Image Processing
- Machine Learning (GAN)
- Inverse Problems
- Model Reduction (Hyperbolic)

Optimal Transport

Monge-Ampère Equation and Differential Geometry

The optimal map for W_2 is $T(x) = \nabla u(x)$, where u solves the Monge-Ampère equation [Brenier, 1991]:

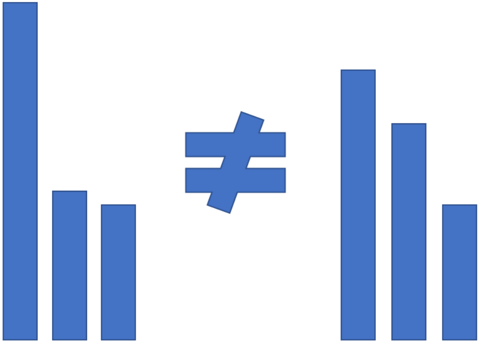
$$\begin{cases} det(D^2 u(x)) = f(x)/g(\nabla u(x)), \quad x \in X, \\ \nabla u : X \to Y, \\ \text{u is convex.} \end{cases}$$

— Studied in the Weyl (1916) and Minkowski (1897) problems in differential geometry of surfaces.

Wasserstein Gradient Flow & Kinetic Descriptions [JKO, 1998]

Energy Functional	Gradient Flow	Well-known PDE
$E(\rho) = \int \rho \log(\rho)$	$\rho_t = \Delta \rho$	Heat Equation
$E(\rho) = \int \rho \log(\rho) + \int \rho V$	$\rho_t = \Delta \rho + \nabla \cdot (\rho \nabla V)$	Linear Fokker-Planck
$E(\rho) = \frac{1}{m-1} \int \rho^m$	$\rho_t = \Delta \rho^m$	Porous Medium Equation
$E(\rho) = \frac{1}{2} \int \rho(x)\rho(y)W(x-y)$	$\rho_t = \nabla \cdot \rho(\nabla(\rho * W))$	McKean-Vlasov Equation

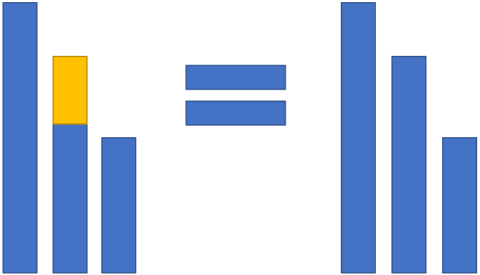
Optimal Transport



[Monge, 1781]

Synthetic data f (left) and observed data g (right)

Optimal Transport



[Monge, 1781]

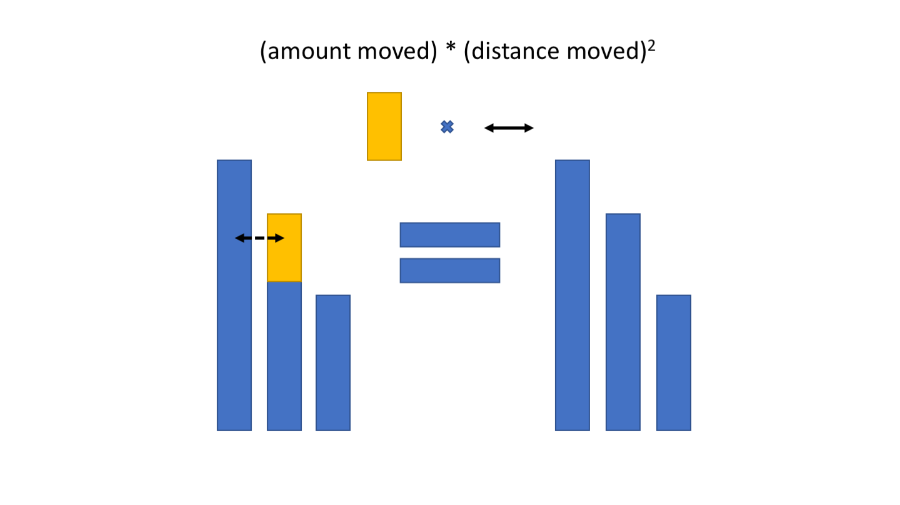
Synthetic data f (left) and observed data g (right)

Optimal Transport

[Monge, 1781]

Synthetic data f (left) and observed data g (right)

Optimal Transport



[Monge, 1781]

Synthetic data f (left) and observed data g (right)

The Wasserstein Distance

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

$$W_p(f, g) = \left(\inf_{T \in \mathcal{M}} \int |x - T(x)|^p f(x) dx \right)^{\frac{1}{p}} \tag{2}$$

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

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

The Wasserstein Distance

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

$$W_p(f,g) = \left(\inf_{T \in \mathcal{M}} \int |x - T(x)|^p f(x) dx \right)^{\frac{1}{p}} \tag{2}$$

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

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

- Provide better optimization landscape for Nonlinear Inverse Problems:

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

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

The Algorithm Based on the Adjoint-State Method

From iteration l to iteration $l+1$, update $\theta \in \mathbb{R}^m$.

$$
\begin{cases}
M_\epsilon(\theta^l)\rho^l = \rho^l, \ \rho^l \cdot \mathbf{1} = 1, \ \textbf{(Solve the forward problem)} \\
(\phi^l, \psi^l) \in \underset{\phi_i + \psi_j \le c(x_i, x_j)}{\operatorname{argmax}} [\phi \cdot \rho^l + \psi \cdot \rho^*], \ \textbf{(Compute the } W_2 \textbf{ distance)} \\
(M_\epsilon(\theta^l)^\top - I)\lambda^l = -\phi^l + \phi^l \cdot \rho^l \, \mathbf{1}, \ \textbf{(Solve the adjoint equation)} \\
\theta_k^{l+1} = \theta_k^l - \tau^l \, \lambda^l \cdot \partial_{\theta_k} M_\epsilon(\theta^l)\rho^l, \ 1 \le k \le m. \ \textbf{(Gradient descent)}
\end{cases}
$$

The Algorithm Based on the Adjoint-State Method

From iteration l to iteration $l+1$, update $\theta \in \mathbb{R}^m$.

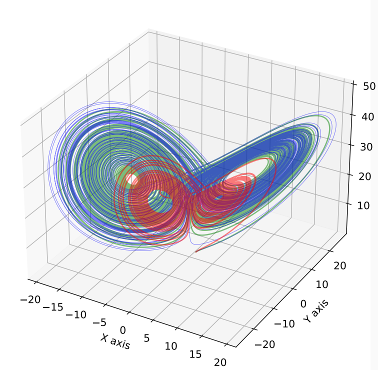
$$
\begin{cases}
M_\epsilon(\theta^l)\rho^l = \rho^l,\ \rho^l \cdot \mathbf{1} = 1, & \textbf{(Solve the forward problem)} \\
(\phi^l, \psi^l) \in \underset{\phi_i + \psi_j \le c(x_i, x_j)}{\operatorname{argmax}} [\phi \cdot \rho^l + \psi \cdot \rho^*], & \textbf{(Compute the } W_2 \textbf{ distance)} \\
(M_\epsilon(\theta^l)^\top - I)\lambda^l = -\phi^l + \phi^l \cdot \rho^l\ \mathbf{1}, & \textbf{(Solve the adjoint equation)} \\
\theta_k^{l+1} = \theta_k^l - \tau^l\ \lambda^l \cdot \partial_{\theta_k} M_\epsilon(\theta^l)\rho^l,\ 1 \le k \le m. & \textbf{(Gradient descent)}
\end{cases}
$$

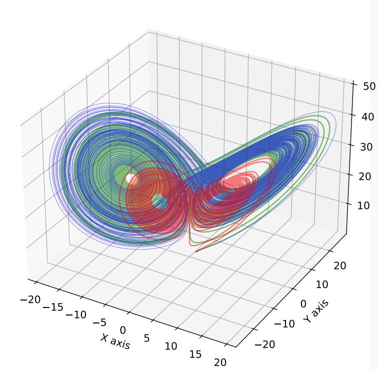
Only two PDE solves are needed, independent of the dimensionality of θ.

Example One: Reconstructing σ, ρ and β in Lorenz-63

The reference PDF is the histogram from a long time trajectory with *intrinsic* noise.

Red: initial; Green: truth; Blue: reconstructed.





Initial: (5, 20, 1); Truth: (10, 28, 8/3). Left: (10.58, 27.83, 2.97), with intrinsic noise; Right: (10.63, 28.82, 3.04), with extrinsic noise.

Variable Coefficients

This is an on-going work with Jonah Botvinick-Greenhouse (Cornell CAM).

Consider an artificial model where the Lorenz-63 model has variable coefficients

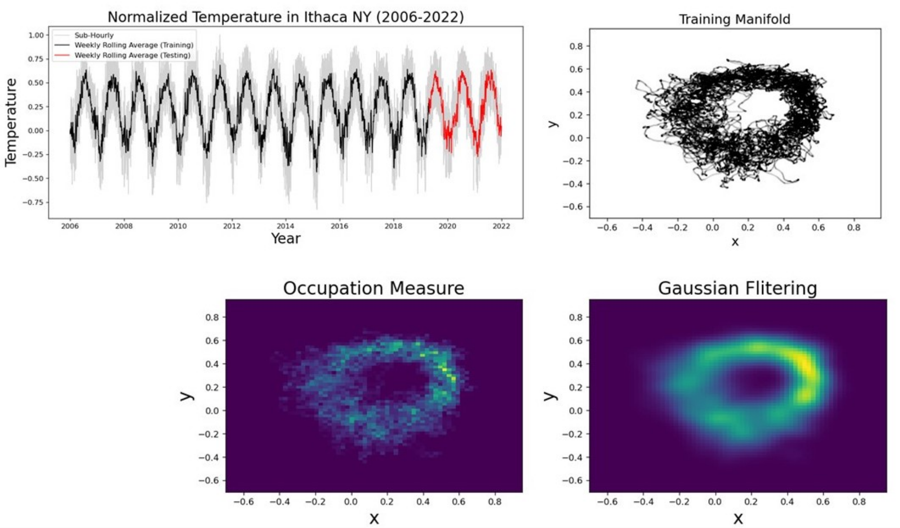
$$\begin{cases} \dot{x} &= \sigma(y-x) \\ \dot{y} &= x(\rho - z) - y \\ \dot{z} &= xy - \beta z \end{cases} \Longrightarrow \begin{cases} \dot{x} &= V_1(x,y,z;\theta) \\ \dot{y} &= V_2(x,y,z;\theta) \\ \dot{z} &= V_3(x,y,z;\theta) \end{cases}$$

How to parametrize $V(x,y,z;\theta)$?

1. Piecewise polynomial based on the mesh of the finite-volume discretization
2. **Neural network parameterization**
3. Many others (RBF, Fourier basis, etc.)

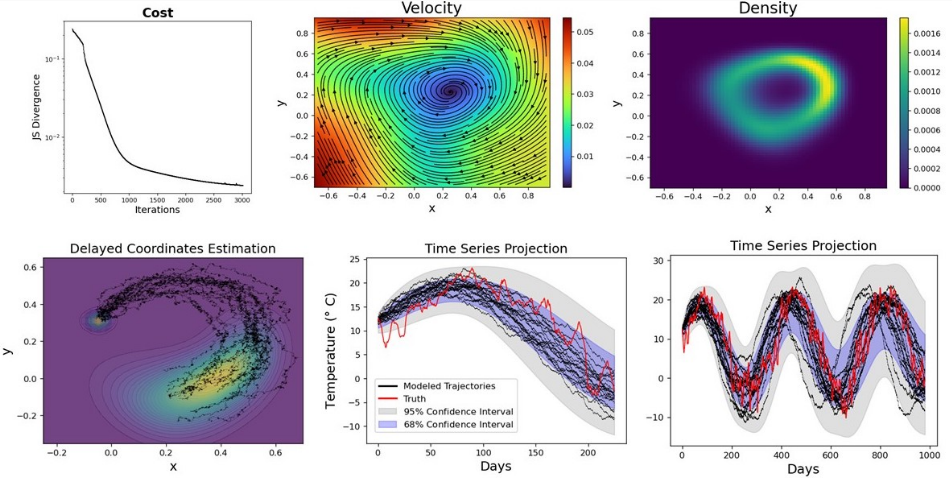
Parameter size increases to tens of thousands.

Example Two: Variable Coefficients—Real Weather Data



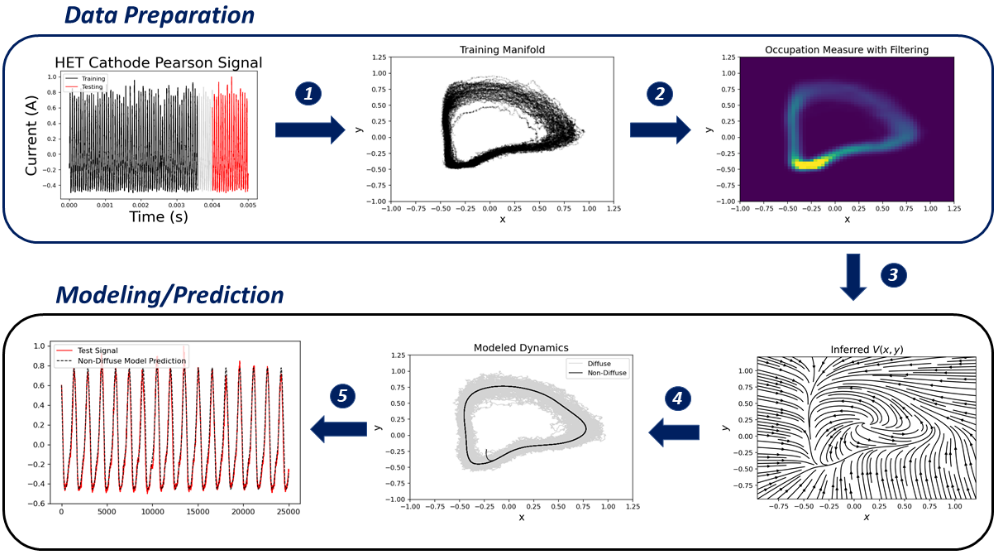
Normalized temperature time series from Ithaca, NY (upper left), the time delay embedding (upper right), occupation measure of the time delay embedding (lower left), and the smoothed occupation measure (lower right).

Example Two: Variable Coefficients—Real Weather Data



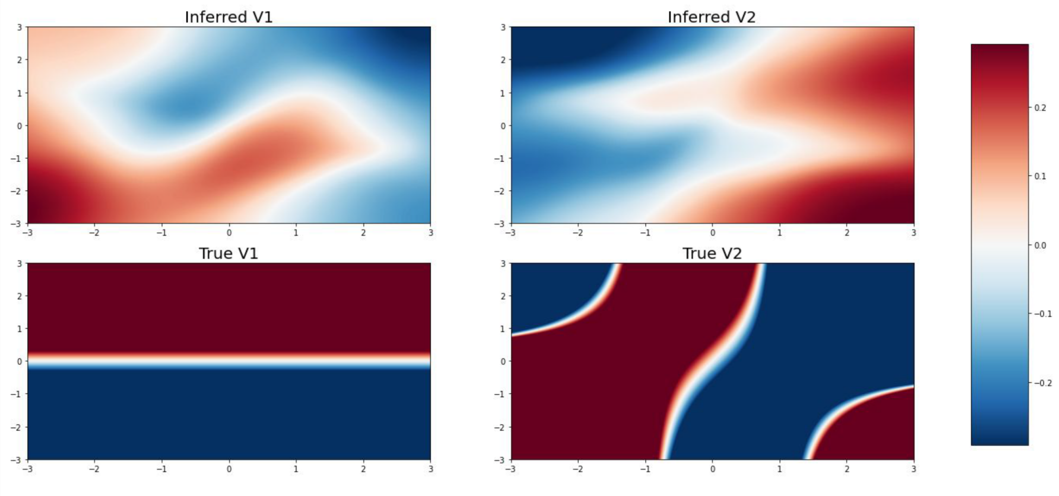
Objective function decay (upper left), velocity learned by the NN (upper middle), the density obtained by solving the forward PDE with this velocity (upper right), evolving both the PDE and twenty SDE sample paths forward in time 225 days in time-delay coordinates (lower left), time-series projections of the probability flow, sample paths, and true data (lower middle), and prediction of the next 981 days vs the ground truth (lower right).

Example Three: Hall-effect thruster (HET)

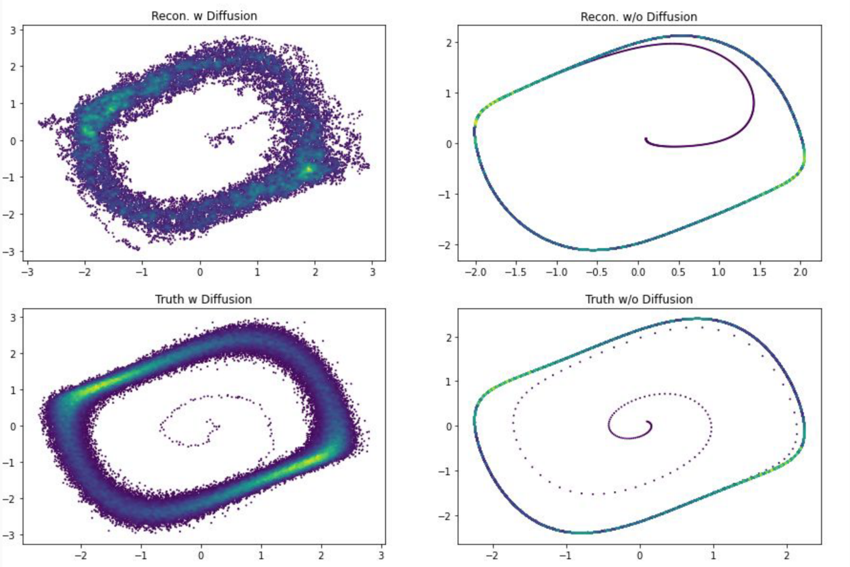


Example Four: Nonuniqueness and ill-posedness — Van der Pol Oscillator

Neural Network Parameterization



Example Four: Nonuniqueness and ill-posedness — Van der Pol Oscillator



Lagrangian vs Eulerian

Summaries

- From Lagrangian to **Eulerian** to tackle chaotic behaviors
- Propose an efficient algorithm based on **CFD discretization** of the forward model.
- The **optimal transport**-based metric as the objective function
- **Adjoint-state** method to calculate the gradient (independent of θ dimension)
- Use **coordinate gradient** descent to tackle multi-parameter inversion

Lagrangian vs Eulerian

Summaries

- From Lagrangian to **Eulerian** to tackle chaotic behaviors
- Propose an efficient algorithm based on **CFD discretization** of the forward model.
- The **optimal transport**-based metric as the objective function
- **Adjoint-state** method to calculate the gradient (independent of θ dimension)
- Use **coordinate gradient** descent to tackle multi-parameter inversion

Future Work

- Improve the finite volume solver for the continuity equation
- Seek other surrogate models for approximating the invariant measure
- Reduce the model discrepancy (noise from both data & solver)
- Dynamical systems with multiple attractors
- **Investigate the connections and differences between these two inverse problems**

Acknowledgments

Research support from

- NSF DMS-1337943
- NSF DMS-1913129
- AFOSR MURI FA 9550 18-1-0502
- AFOSR Grant FA9550-20RQCOR098 (PO: Leve)
- AFOSR Grant FA9550-20RQCOR100 (PO: Fahroo)

We thank Prof. Adam Oberman for initiating our collaboration and Prof. Alex Townsend for constructive suggestions.

This work was in part completed during the long program on High Dimensional Hamilton-Jacobi PDEs held in the Institute for Pure and Applied Mathematics (IPAM) at UCLA, March 9 – June 12, 2020.

Thank you for the attention!