Learning Dynamics on Invariant Measures Using PDE-Constrained Optimization

Jonah Botvinick-Greenhouse[1], Robert Martin[2], Yunan Yang[3]

[1]Cornell University, [2]DEVCOM Army Research Laboratory, [3]Cornell University

Abstract

Common approaches for modeling dynamical trajectories of **complex physical systems** adopt a **Lagrangian** perspective and seek a pointwise matching with either the observed data or its approximate state derivatives. When the observed data is corrupt with noise and has a poor temporal resolution, these approaches may struggle. We instead propose am **Eulerian** alternative which treats the **global statistics** of the observed dynamics as the inference data.

Introduction

Problem Statement:

- Consider the ordinary differential equation (ODE) $\dot{x} = v(x;\theta)$, where $\theta \in \mathbb{R}^p$.
- Suppose we observe samples $\{x(t_i)\}_{i=1}^N \subseteq \mathbb{R}^d$ of the ODE.
- Our goal is to **recover** the **unknown parameters** $\theta \in \mathbb{R}^p$.
- The method should be **robust** to the challenges depicted in Figure 1.

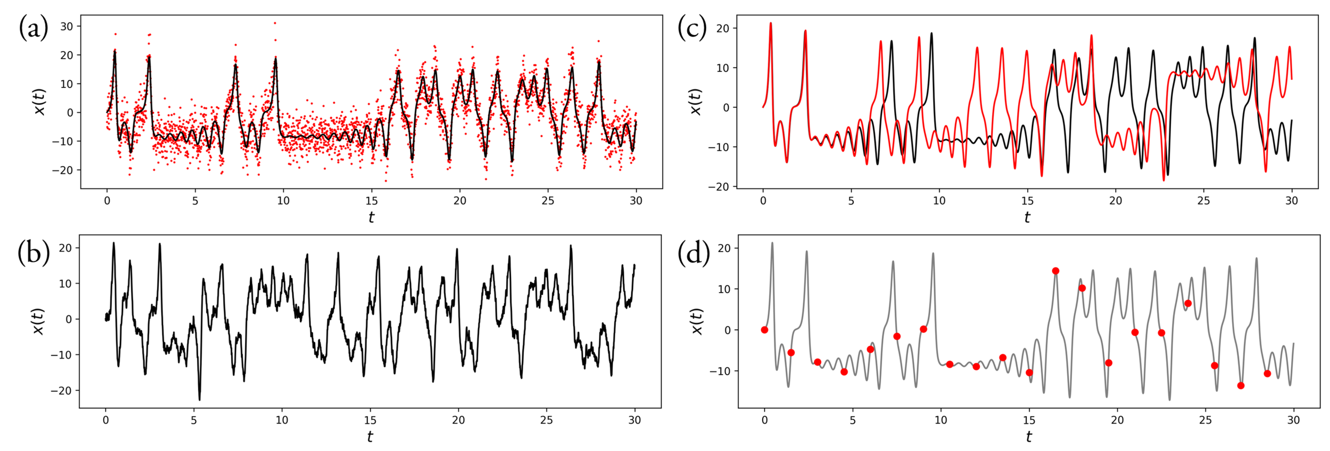

Figure 1. Illustration of the challenges one may encounter during the parameter identification problem: (a) extrinsic noise, (b) intrinsic noise, (c) chaos, and (d) slow sampling.

Our Approach

- We consider the **occupation measure**

$$\rho^* := \frac{1}{N}\sum_{k=0}^{N-1} \delta_{x(t_k)} \in \mathcal{P}(\mathbb{R}^d).$$

- We implement a regularized PDE forward model $\theta \mapsto \rho_\varepsilon(\theta)$.
- We perform the optimization $\min_{\theta\in\mathbb{R}^p} \mathcal{J}(\rho^*, \rho_\varepsilon(\theta))$ via gradient methods.

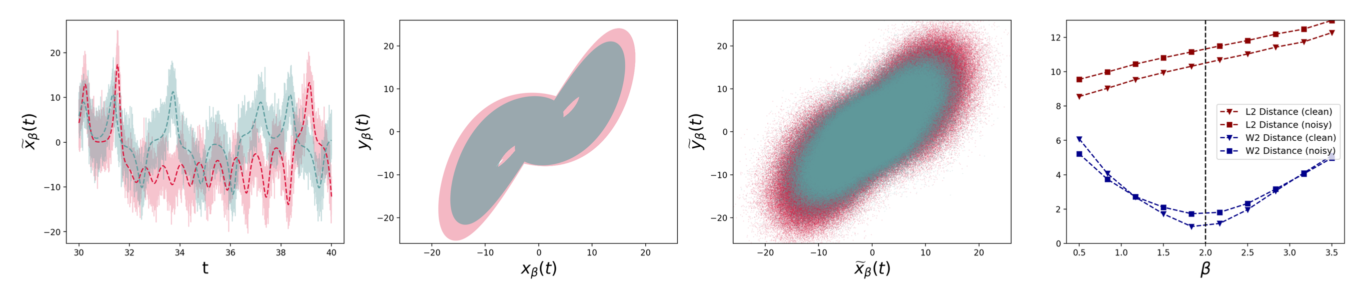

Figure 2. Using the invariant measure to compare nearby dynamical systems results in a meaningful optimization landscape, whereas a trajectory-based loss does not.

The Regularized Forward Model

- The **Eulerian** Fokker–Planck (FP) partial differential equation (PDE) gives the law of the **Lagrangian** stochastic differential equation (SDE):

$$\underbrace{\partial_t \rho + \nabla\cdot(\rho v) = D\Delta\rho}_{\text{Eulerian}} \iff \underbrace{dX_t = v(X_t)dt + \sqrt{2D}dW_t}_{\text{Lagrangian}} \qquad (1)$$

- Our continuous forward model is $\theta \mapsto \rho(\theta)$, satisfying

$$\nabla\cdot(\rho(\theta)v(\theta)) = D\Delta\rho(\theta).$$

- We discretize the FP equation with a first-order **upwind finite volume method**, which yields the discrete forward model $\theta \mapsto \rho(\theta)$, where

$$M_\varepsilon \rho(\theta) = \rho(\theta), \qquad M_\varepsilon := (1-\varepsilon)M + \frac{\varepsilon}{N}\mathbf{1}\mathbf{1}^\top \in \mathbb{R}^{k\times k}.$$

- M is a **Markov matrix** and $\varepsilon > 0$ is a regularization parameter ensuring uniqueness of the recovered $\rho(\theta)$.

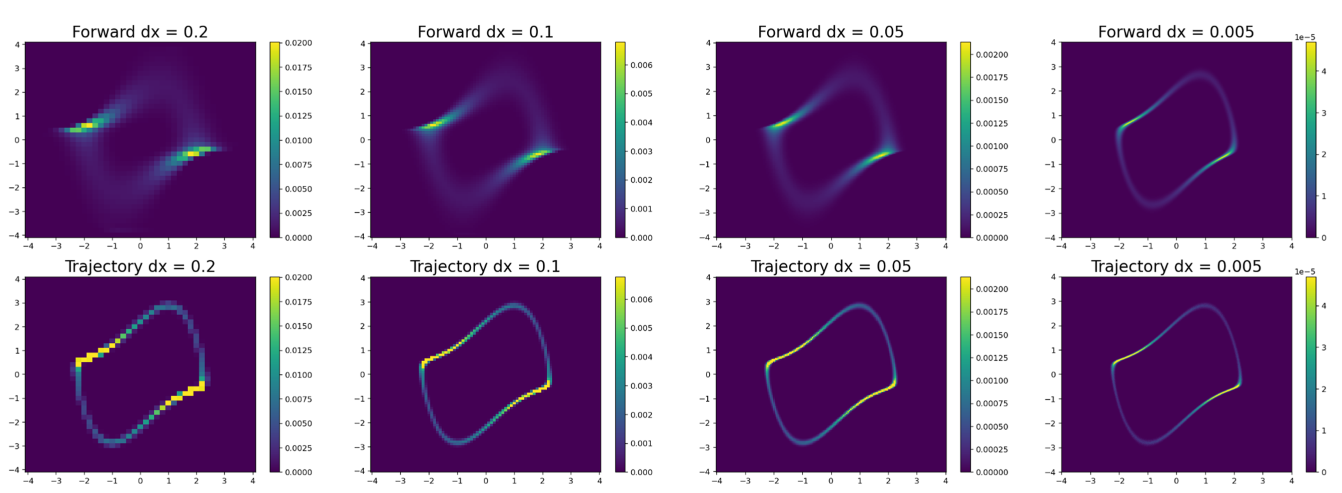

Figure 3. Comparison of the forward model with trajectory histograms.

Choice of Objective Function

- We consider the following objectives for comparing $\rho(\theta)$ and ρ^*.
  1. Squared L^2 norm: $\mathcal{J}(\theta) = \int_\Omega |\rho(x;\theta) - \rho^*(x)|^2 dx.$
  2. Kullback–Liebler: $\mathcal{J}(\theta) = D_{\mathrm{KL}}(\rho(\theta), \rho^*) = \int_\Omega \rho^*(x)\log\left(\frac{\rho^*(x)}{\rho(x;\theta)}\right)dx.$
  3. Jenson–Shanon: $\mathcal{J}(\theta) = \frac{1}{2}D_{\mathrm{KL}}(\rho(\theta), (\rho(\theta)+\rho^*)/2) + \frac{1}{2}D_{\mathrm{KL}}((\rho(\theta)+\rho^*)/2, \rho(\theta))$
  4. Wasserstein distance: $\mathcal{J}(\theta) = \inf_{\phi\in\mathcal{T}_{\rho(\theta),\rho^*}} \int_\Omega |x-\phi(x)|^2 \rho(x;\theta)dx$

Gradient Calculation via the Adjoint-State Method

- We compute $\partial_\theta \mathcal{J}(\theta)$ using the adjoint-state method:
  1. Compute the Fréchet derivative $\partial_\rho \mathcal{J}$.
  2. Solve the adjoint equation
  $$(M_\varepsilon^\top - \mathrm{Id})\lambda = -(\partial_\rho\mathcal{J})^\top + (\partial_\rho\mathcal{J})^\top \rho\mathbf{1}.$$
  3. Compute the gradient
  $$\partial_\theta\mathcal{J} = \lambda^\top(\partial_\theta M_\varepsilon)\rho.$$
- The sparsity of M allows us to analytically derive an efficient formula for $\partial_\theta\mathcal{J}$.

Velocity Parameterization

- We test the following large-scale velocity parameterizations:
  1. Piecewise constant: $v(x;\theta) = \sum_{i=1}^d \sum_{i=j}^N v_j^i \chi_{C_j}(x)\,\mathbf{e}_i$, where $\theta = \{v_j^i\}$.
  2. Global polynomial: $v_i(x;\theta) = \sum_{\ell=1}^N a_\ell^i (x^\top \mathbf{e}_1)^{{}^1k_\ell^i} \dots (x^\top\mathbf{e}_d)^{{}^dk_\ell^i}$, where $\theta = \{a_\ell^i\}$.
  3. Neural network: $v(x;\theta) = \mathrm{NN}(x;\theta)$, where θ comprises the weights/biases.

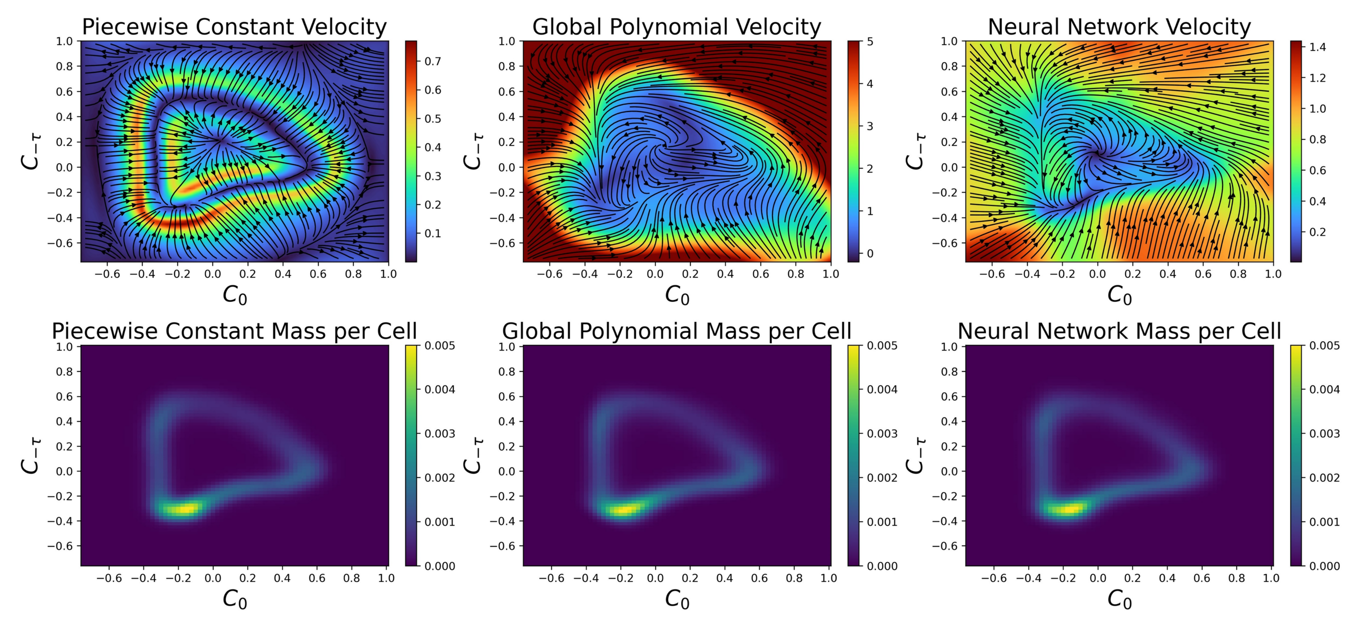

Figure 4. Investigating the choice of parameterization on velocity reconstruction. The top row shows the reconstructed velocity and the bottom row shows the computed invariant measure based on the learned velocity.

Comparison with Existing Methods

- We compare our approach with the SINDy and Neural ODE frameworks for learning slowly sampled dynamics of the Van der Pol oscillator.

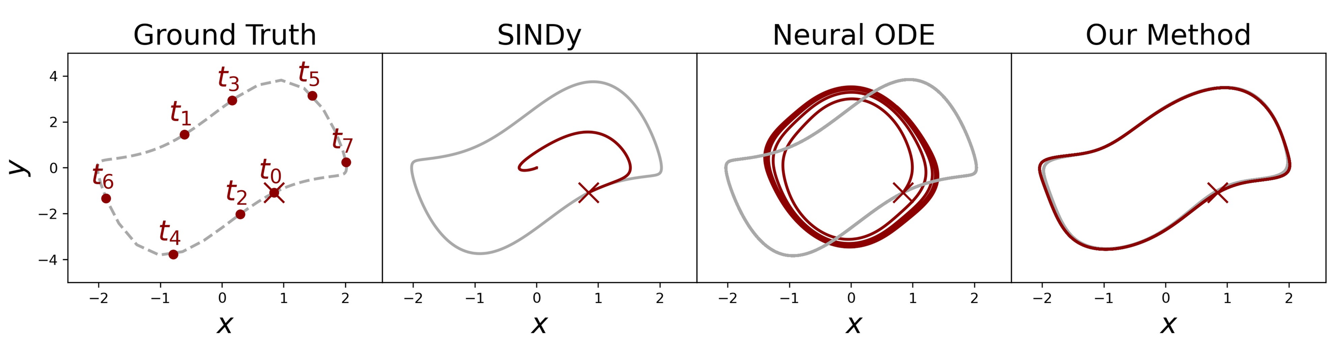

Figure 5. The left panel shows the first eight points of a slowly sampled trajectory (in red). While the SINDy and Neural ODE frameworks can learn the quickly sampled dynamics, both struggle to learn from the slowly sampled trajectory.

Method	Sampling Freq.	Wall-Clock Time (s)	Error
SINDy	10.00	$2\cdot10^{-2}$	$5.6\cdot10^{-3}$
Neural ODE	10.00	$5\cdot10^{2}$	$5.32\cdot10^{-3}$
Ours	10.00	$5\cdot10^{2}$	$1.14\cdot10^{-1}$
SINDy	0.25	10^{-2}	3.52
Neural ODE	0.25	$5\cdot10^{2}$	1.81
Ours	0.25	$5\cdot10^{2}$	$6.79\cdot10^{-2}$

Table 1. Comparison for learning from trajectories sampled at different frequencies (Hz).

Application to Chaotic Dynamics

- Consider the following Arctan Lorenz system:

$$\begin{cases} \dot{x} &= 50\arctan(c_1(y-x)/50) \\ \dot{y} &= 50\arctan(x(c_2-z)/50 - y/50)\,, \\ \dot{z} &= 50\arctan(xy/50 - c_3 z/50) \end{cases} \qquad \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 28 \\ 8/3 \end{bmatrix}$$

- In Figure 6, we learn $\dot{x}$ from the Arctan Lorenz systems's invariant measure.

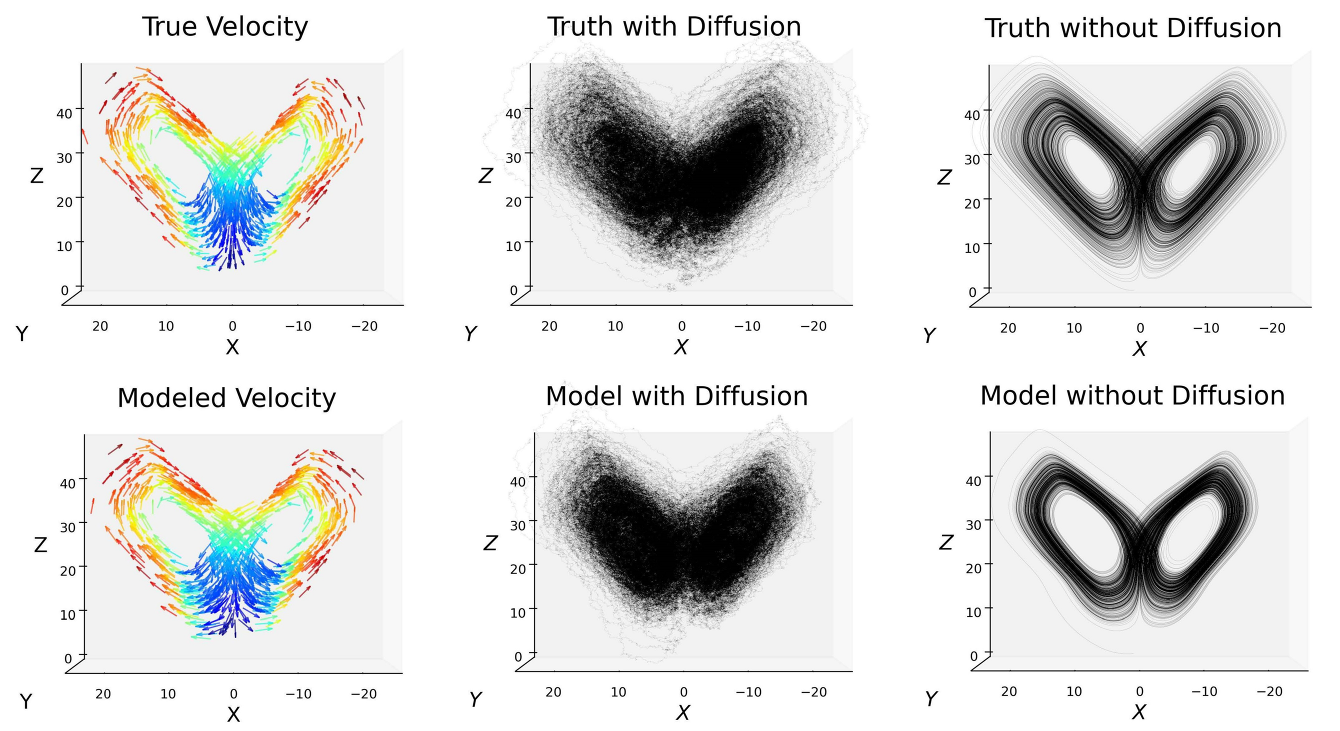

Figure 6. Neural network parameterization of $\dot{x}$ using the Arctan Lorenz system's stochastically perturbed invariant measure with $D = 10$.

Forecasting Dynamics with Uncertainty Quantification

- In Figure 7, we learn the velocity $v(x;\theta)$ of a Hall-effect thruster system based upon real data.

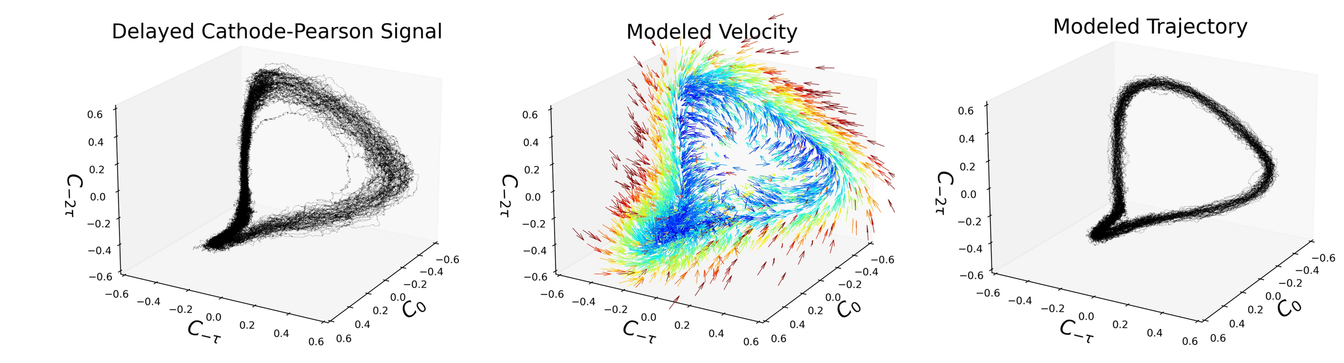

Figure 7. Learning the velocity of the Hall-effect thruster system.

- In Figure 8, we use the learned PDE model to perform uncertainty quantification for evolving trajectories.

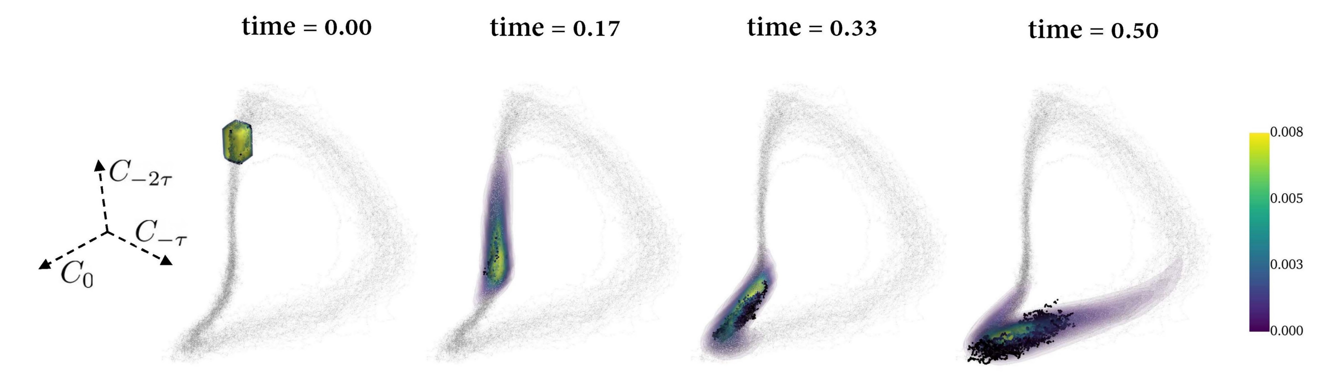

Figure 8. We show the evolution of ground truth samples from the Hall-effect thruster system, plotted in black, and our Fokker–Planck PDE model, represented by the probability contours.

Conclusions

1. We reformulated ODE learning as a PDE-constrained optimization problem by treating **invariant measures** as inference data.
2. We used the **adjoint-state method** to compute the gradient.
3. The proposed approach is **robust** to learning from slowly sampled dynamics.
4. The learned models were used to perform **uncertainty quantification.**

Acknowledgments

This work was supported in part by a fellowship award under contract FA9550-21-F-0003 through the National Defense Science and Engineering Graduate (NDSEG) Fellowship Program, sponsored by the Air Force Research Laboratory (AFRL), the Office of Naval Research (ONR) and the Army Research Office (ARO). R. Martin was partially supported by AFOSR Grants FA9550-20RQCOR098 (PO: Leve) and FA9550-20RQCOR100 (PO: Fahroo). This work was done in part while Y. Yang was visiting the Simons Institute for the Theory of Computing in Fall 2021. Y. Yang acknowledges support from Dr. Max Rössler, the Walter Haefner Foundation and the ETH Zürich Foundation. This material is based upon work supported by the National Science Foundation under Award Number DMS-1913129.

References

[1] Botvinick-Greenhouse, J., Martin, R., Yang, Y. (2023). Learning dynamics on invariant measures using PDE-constrained optimization. Chaos: An Interdisciplinary Journal of Nonlinear Science, 33(6).

[2] 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.

[3] Chen, R. T., Rubanova, Y., Bettencourt, J., Duvenaud, D. K. (2018). Neural ordinary differential equations. Advances in neural information processing systems, 31.

[4] Greve, C. M., Hara, K., Martin, R. S., Eckhardt, D. Q., Koo, J. W. (2019). A data-driven approach to model calibration for nonlinear dynamical systems. Journal of Applied physics, 125(24).

[5] Langville, A. N., Meyer, C. D. (2006). Google's PageRank and beyond: The science of search engine rankings. Princeton university press.

[6] Yang, Y., Nurbekyan, L., Negrini, E., Martin, R., Pasha, M. (2023). Optimal transport for parameter identification of chaotic dynamics via invariant measures. SIAM Journal on Applied Dynamical Systems, 22(1), 269-310.