Generative Modeling of Time-Dependent Densities via Optimal Transport and Projection Pursuit

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

[1]Cornell University, [2]ETH Zürich, [3]Argonne National Laboratory

Abstract

Deep learning algorithms for generative modeling of time-dependent data may face difficulties related to **optimal hyperparameter selection** and can be **computationally expensive**. In this work, we propose a generative model based on numerical linear algebra and **projection pursuit optimal transport**, which requires only a few hyperparameter choices. We demonstrate the effectiveness of our model on synthetic data of stochastic dynamical systems and real-life examples of **fish schooling** and **embryoid growth**.

Introduction

Problem Statement:

- Let $\rho(x,t)$ be a time-dependent density on $\mathbb{R}^d$.
- We observe $\mathbf{X}^{(t_j)} \in \mathbb{R}^{N_j \times d}$ sampled from $\rho(x,t)$ at the times $\{t_j\}_{j=1}^M$.
- Goal: learn to sample from $\rho(x,t)$, for any $t \in [t_1, t_M]$.

Motivation:

- The data $\mathbf{X}^{(t_j)}$ naturally arises in cellular dynamics and animal swarm behavior.

Challenges:

- In high dimensions, classic density estimation techniques are not applicable.
- Deep learning requires a hyperparameter search and nonconvex optimization.

Optimal Transport

If p_X and p_Y are densities on $\mathbb{R}^d$, then ϕ is a transport map whenever

$$p_Y(x) = p_X(\phi^{-1}(x))|\det D\phi^{-1}(x)|.$$

Minimizing over all such transport maps defines the Wasserstein metric

$$W_p(p_X, p_Y) := \left(\inf_{\phi \in \Phi} \int_{\mathbb{R}^d} |x - \phi(x)|^p p_X(x)dx\right)^{1/p}.$$

The minimizer ϕ_* is the optimal transport map (OTM). In the case $d = 1$,

$$\phi_*(x) = (G^{-1} \circ F)(x), \tag{1}$$

where F and G are the CDFs of p_X and p_Y, respectively.

Projection Pursuit Monge-Map (PPMM)

The PPMM algorithm [1] learns a **high-dimensional OTM** by computing one-dimensional OTMs along **dominant projection directions.** Given $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{N \times d}$, set $\mathbf{X}_0 = \mathbf{X}$, $k = 0$, and iterate the following until convergence.

1. Compute a projection direction $p_k \in \mathbb{R}^d$ between $\mathbf{X}_k$ and $\mathbf{Y}$.
2. Compute the one-dimensional OTM η_k from $\mathbf{X}_k p_k$ to $\mathbf{Y} p_k$ using (1).
3. Set $\mathbf{X}_{k+1} = \mathbf{X}_k + \left(\eta_k(\mathbf{X}_k p_k) - \mathbf{X}_k p_k\right) p_k^T$ and $k = k + 1$.

The final OTM $\hat{\phi}_k : \mathbb{R}^d \to \mathbb{R}^d$ is defined recursively by setting $\hat{\phi}_0 := \mathrm{Id}$ and

$$\widehat{\phi}_{\ell+1}(x) = \widehat{\phi}_\ell(x) + \left(\eta_\ell\left(\widehat{\phi}_\ell(x) p_\ell\right) - \widehat{\phi}_\ell(x) p_\ell\right) p_\ell^T, \qquad 0 \le \ell \le k - 1.$$

Our Proposed Approach (D-PPMM)

Training:

1. Sample the rows of $\mathbf{X}^{(t_0)} \in \mathbb{R}^{N_1 \times d}$ i.i.d. from the reference distribution $\mathcal{N}(\mu, \Gamma)$.
2. Use PPMM to learn the OTM $\hat{\phi}_{k_j}^{(t_j)}$ between $\mathbf{X}^{(t_j)}$ and $\mathbf{X}^{(t_{j+1})}$, for $0 \le j \le M - 1$.

Generation at $\{t_j\}_{j=1}^M$:

1. Sample $\mathbf{Y}^{(t_0)} \in \mathbb{R}^{\tilde{N} \times d}$ i.i.d. from the reference distribution $\mathcal{N}(\mu, \Gamma)$.
2. Compute $\mathbf{Y}^{(t_{j+1})} = \hat{\phi}_{k_j}^{(t_j)}(\mathbf{Y}^{(t_j)})$ for $0 \le j \le M - 1$.

Transport spline interpolation [2] between $\{t_j\}_{j=1}^M$:

1. Compute the interpolants

$$p_n := \mathtt{Spline}\left([t_1, t_2, \ldots, t_M], [\mathbf{Y}_n^{(t_1)}, \mathbf{Y}_n^{(t_2)}, \ldots, \mathbf{Y}_n^{(t_M)}]\right).$$

2. Evaluate $\mathbf{Y}_i^{(t)} = p_i(t)$, for $1 \le i \le \tilde{N}$ and $t \in [t_1, t_M]$.

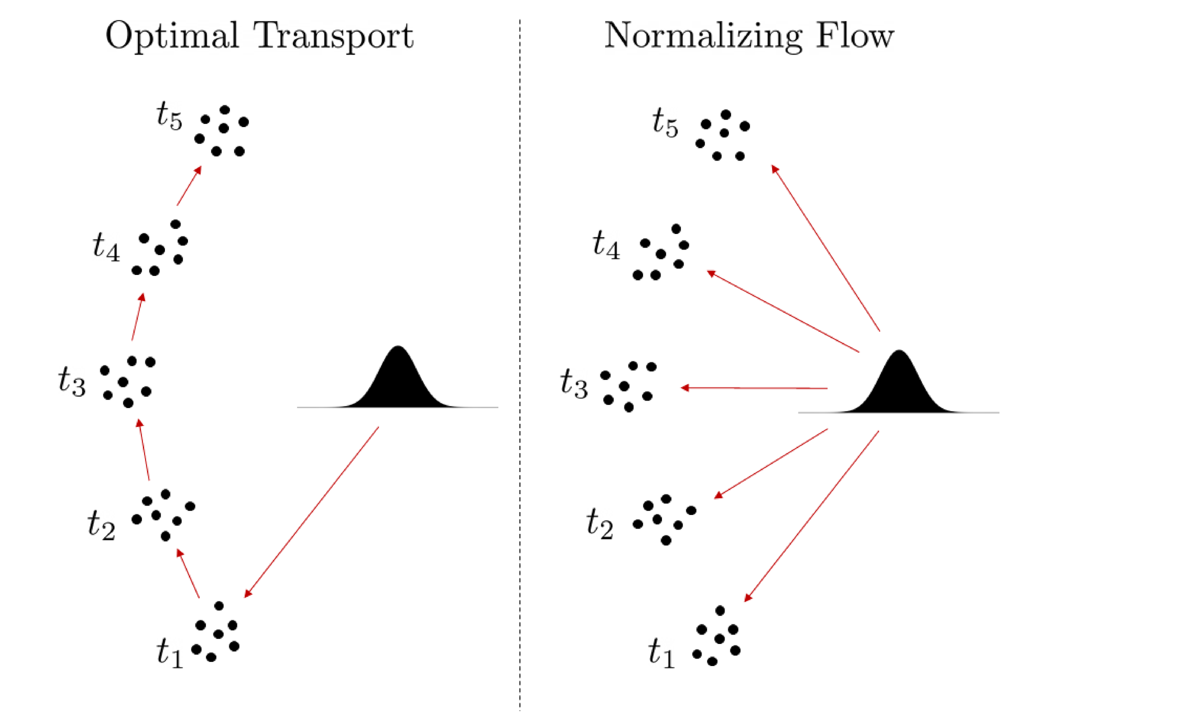


Figure 1. Illustration of models based on optimal transport (left) and deep learning (right).

Numerical Results (Synthetic Data)

We now present numerical comparisons of D-PPMM with a state-of-the-art normalizing flow (the RQ-NSF) [3] conditioned on time.

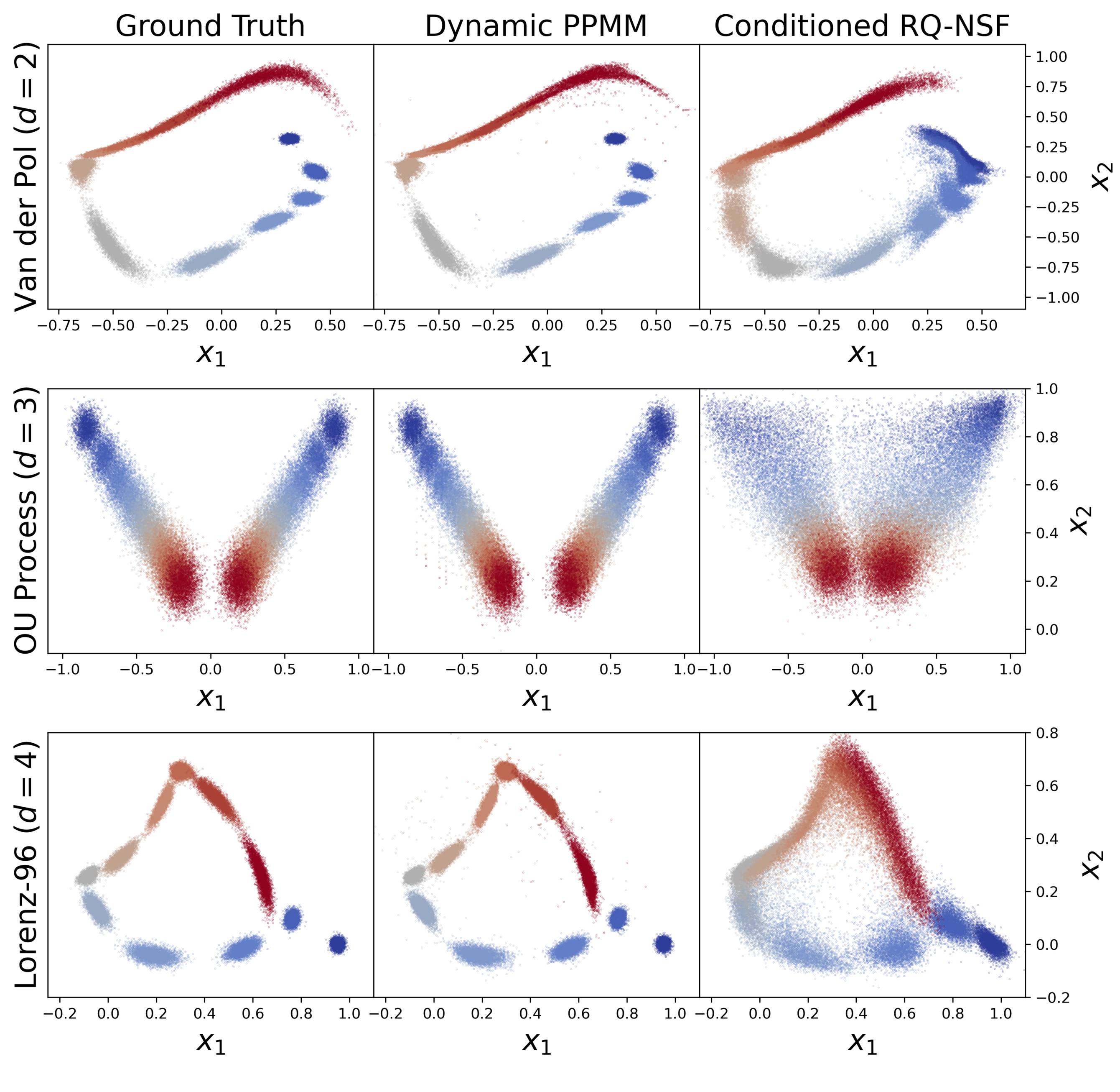


Figure 2. Comparison for low-dimensional systems with the same wall-clock training times.

We quantify error between $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{N \times d}$ using the maximum mean discrepancy:

$$\frac{1}{N(N-1)} \sum_{n=1}^{N} \sum_{m \neq n}^{N} \left(\gamma_\sigma(\mathbf{X}_n, \mathbf{X}_m) + \gamma_\sigma(\mathbf{Y}_n, \mathbf{Y}_m)\right) - \frac{2}{N^2} \sum_{n=1}^{N} \sum_{m=1}^{N} \gamma_\sigma(\mathbf{X}_n, \mathbf{Y}_m), \tag{2}$$

where γ_σ is a Gaussian kernel with bandwidth σ. We report the maximum of (2) over a family $\mathcal{I}$ of bandwidths and average the error across all times.

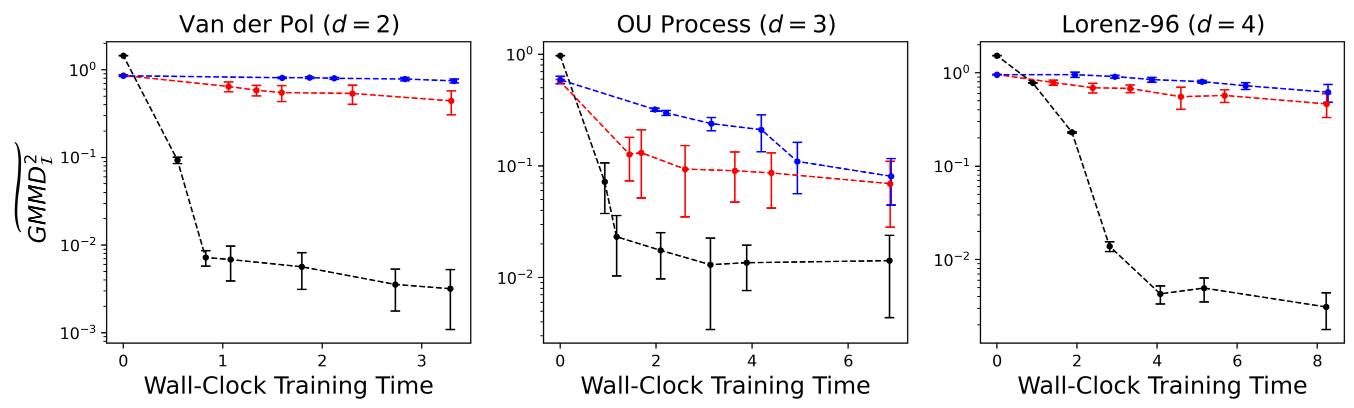


Figure 3. Convergence comparison for the low-dimensional systems shown in Figure 2. The D-PPMM error is plotted in black, whereas the blue and red curves correspond to different RQ-NSF architectures.

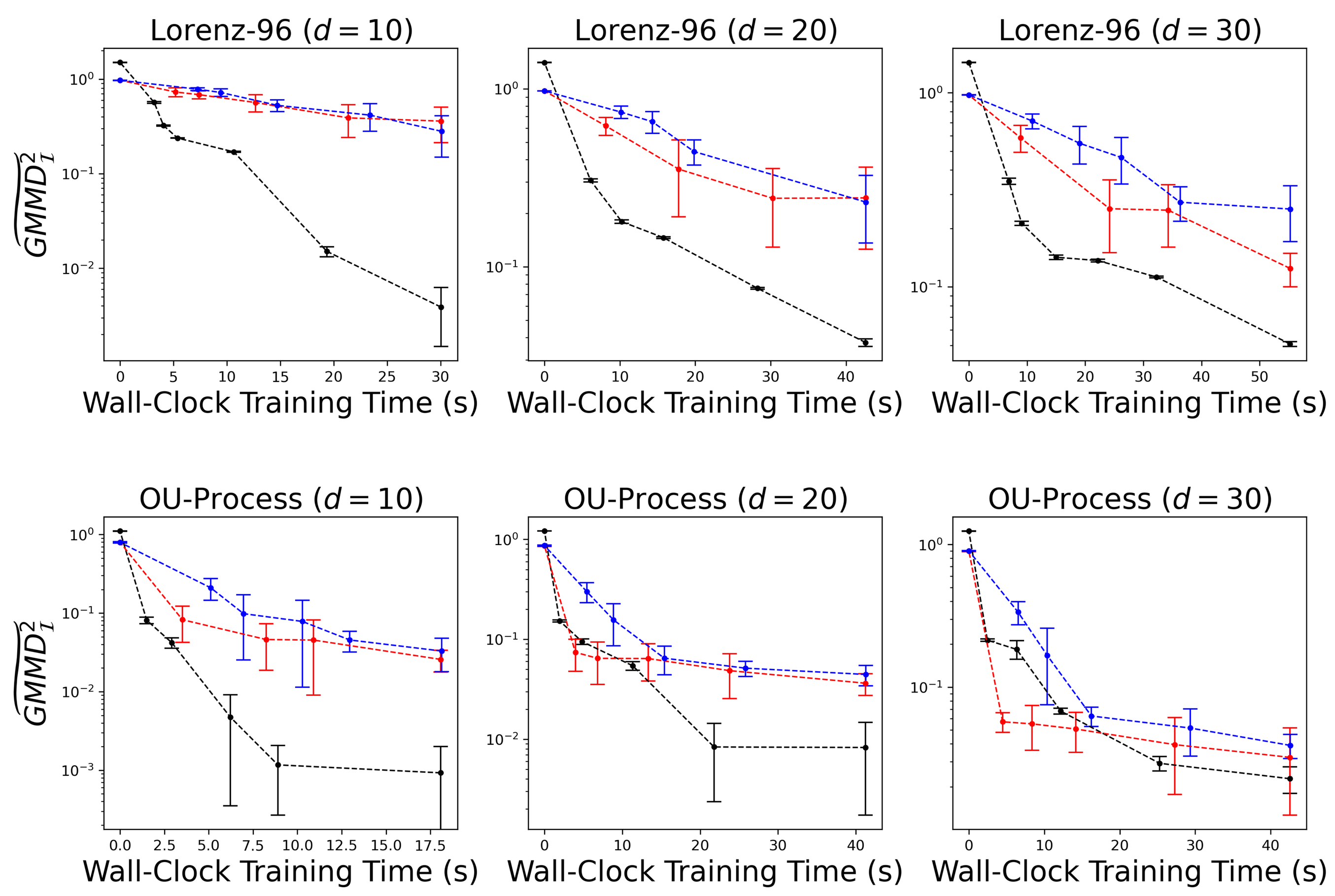


Figure 4. Convergence comparison for high-dimensional Lorenz-96 systems and OU-processes. The D-PPMM error is plotted in black, whereas the blue and red curves correspond to different RQ-NSF architectures.

Numerical Results (Real Data)

We now study the applicability of the proposed approach for modeling the dynamics of fish schooling and embryoid growth.

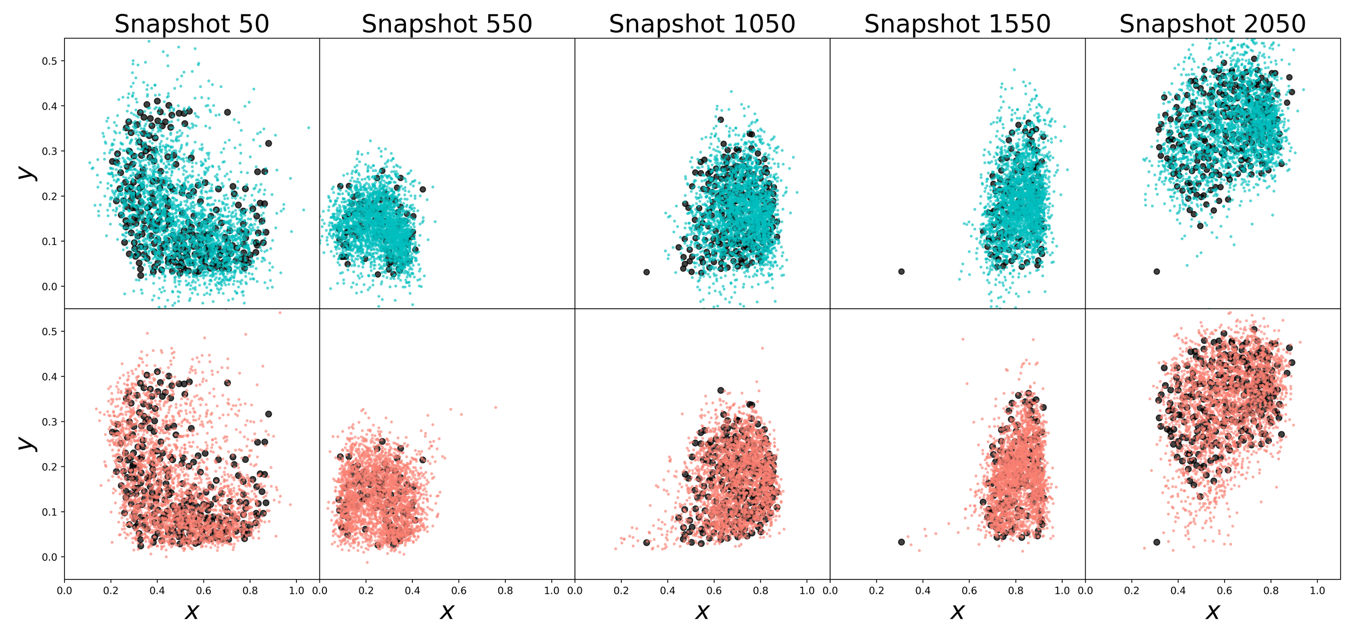


Figure 5. Using D-PPMM to model fish schooling behavior. The generated samples from D-PPMM are shown in cyan, the generated samples from an RQ-NSF are shown in orange, and the unseen ground truth samples are plotted in black.

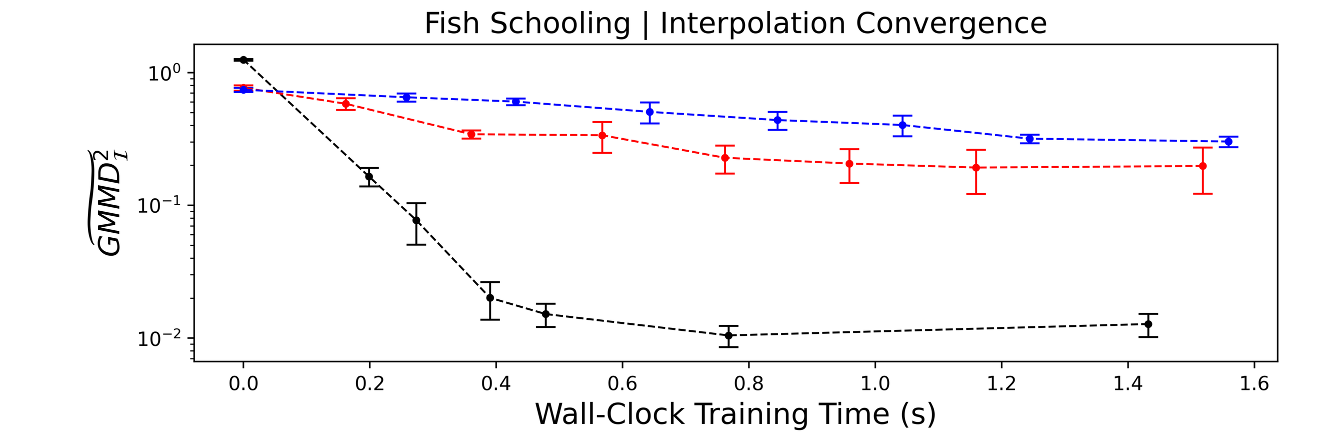
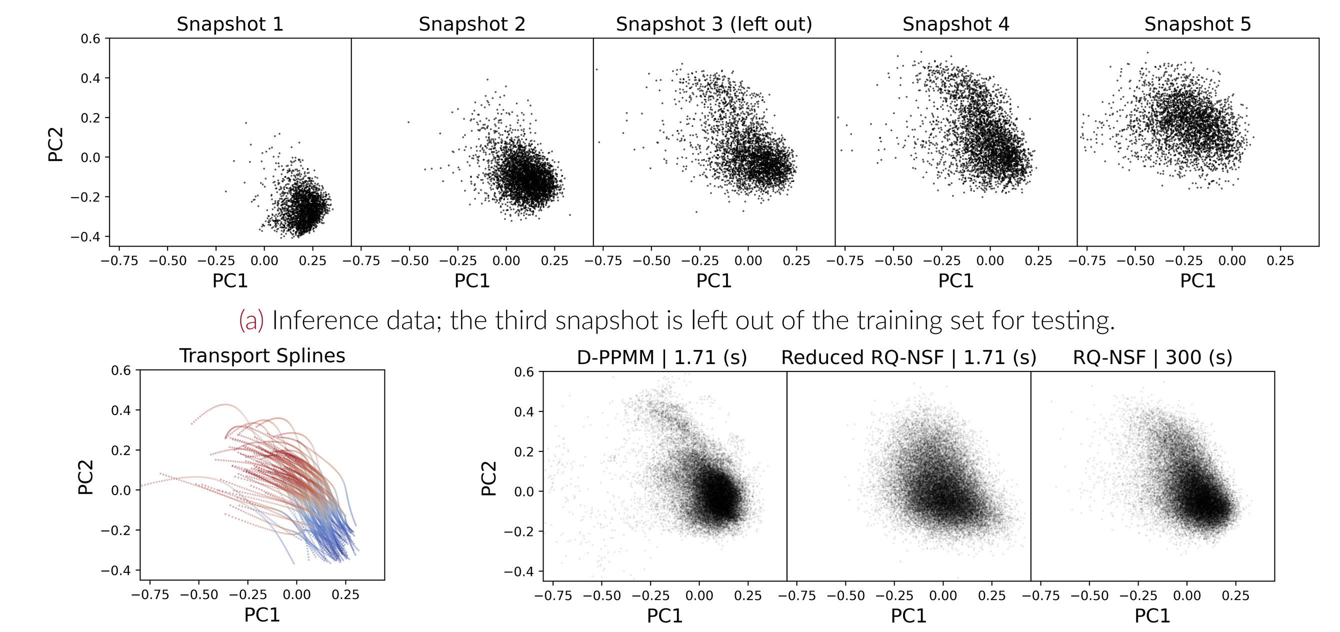


Figure 6. Convergence comparison at the held out interpolation times for the fish schooling dataset. The D-PPMM error is plotted in black, whereas the blue and red curves correspond to different RQ-NSF architectures.



(a) Inference data; the third snapshot is left out of the training set for testing.

(b) D-PPMM transport splines. (c) Interpolation with D-PPMM and RQ-NSFs for various training times.

Figure 7. Top row: embryoid body training and testing datasets. Bottom left: Transport splines that D-PPMM uses to interpolate the left-out snapshot. Bottom right: interpolation comparison of the third snapshot.

Method	Comp. Time (s)	$\mathrm{GMMD}_{\mathcal{I}}^2$ $(\times 10^{-3})$
D-PPMM	0.81	13.51 ± 4.35
Reduced RQ-NSF	0.81	319.25 ± 5.49
RQ-NSF	1200	8.76 ± 0.85

Method	Comp. Time (s)	$\mathrm{GMMD}_{\mathcal{I}}^2$ $(\times 10^{-3})$
D-PPMM	1.70	24.97 ± 1.26
Reduced RQ-NSF	1.70	54.99 ± 2.28
RQ-NSF	1200	24.14 ± 1.04

Table 1. Comparison of the modeling error ($\mathrm{GMMD}_{\mathcal{I}}^2$) for the fish schooling dataset (left) and the embryoid body dataset (right). The wall-clock computation time is also reported.

Conclusion

We have introduced **D-PPMM**, which constructs a **generative model** of an evolving density by joining particle snapshots via **optimal transport maps** and interpolating the evolving density with cubic **transport splines**. Through several numerical experiments, we showed that the approach is **highly competitive** with state-of-the-art normalizing flows conditioned upon time.

Acknowledgements

This work was supported by the U.S. Department of Energy (DOE), Office of Science, Office of Advanced Scientific Computing Research (ASCR), under Contract No. DEAC02–06CH11357, at Argonne National Laboratory. We also acknowledge funding support from ASCR for DOE-FOA-2493, "Data-intensive scientific machine learning". This research was supported in part by an appointment with the National Science Foundation (NSF) Mathematical Sciences Graduate Internship (MSGI) Program sponsored by the NSF Division of Mathematical Sciences. This program is administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and NSF. ORISE is managed for DOE by ORAU. All opinions expressed in this poster are the author's and do not necessarily reflect the policies and views of NSF, ORAU/ORISE, or DOE. J. Botvinick-Greenhouse was supported in part by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate (NDSEG) Fellowship Program. Y. Yang acknowledges support from Dr. Max Rössler, the Walter Haefner Foundation, and the ETH Zürich Foundation.

References

[1] Meng, C., Ke, Y., Zhang, J., Zhang, M., Zhong, W. & Ma, P. Large-scale optimal transport map estimation using projection pursuit. *Advances In Neural Information Processing Systems*. **32** (2019)

[2] Chewi, S., Clancy, J., Le Gouic, T., Rigollet, P., Stepaniants, G. & Stromme, A. Fast and smooth interpolation on Wasserstein space. *International Conference On Artificial Intelligence And Statistics*. pp. 3061-3069 (2021)

[3] Durkan, C., Bekasov, A., Murray, I. & Papamakarios, G. Neural spline flows. *Advances In Neural Information Processing Systems*. **32** (2019) Wesley, Massachusetts, 2nd ed.

jrb482@cornell.edu