NUMERICAL ANALYSIS AND GEOMETRY OF MEASURE-VALUED DATA

NICHOLAS KARRIS

My research centers around the mathematics of data science and machine learning, with a recent bend towards problems involving data which is fundamentally measure-valued. In these situations, many natural questions arise about how one should perform standard data science tasks. My work answers these questions by viewing the measures themselves as single points and leveraging the tools of optimal transport and the geometry of the Wasserstein manifold to develop intuitive analogs of standard data science techniques for measure-valued data. By computing optimal transport maps from a fixed reference, we embed each measure into a common linear space (the tangent space of the Wasserstein manifold at the fixed reference), which allows us to apply out-of-the-box methods on the embeddings. When the data is static, we use these methods to build classifiers, generate new measures from particular classes, and interpolate between measures with geodesics, and when the data is evolving in time, we describe numerical methods for approximating the evolution, finding steady state distributions, and analyzing stability.

Introduction

There are many situations in which one would like to apply standard machine learning or numerical analysis techniques, except the available data are measures, not points in Euclidean space. For example, for a collection of particles evolving over time (e.g., as in a simulation of some biological process), it may be more natural to interpret the collection of particle locations as an *unordered* distribution as opposed to an *ordered* vector [44, 25]. In such cases, we may still be interested in asking questions about the "derivative" of the collection's evolution and what one can glean about the evolution from this derivative. Or perhaps we have 3D scans of different objects, each represented as a finite, unordered collection of points in $\mathbb{R}^3$, and we want to train a classifier on the known data [5, 40, 51]. Each of these tasks is standard when the data is vector-valued, but when the data is fundamentally measure-valued, the problem is less straightforward.

To motivate our methods, we turn to the geometry of the Wasserstein manifold. Formally, we can view $\mathcal{P}_2(\mathbb{R}^d)$ (the space of probability measures on $\mathbb{R}^d$ with finite second moment) as a Riemannian manifold with metric given by the Wasserstein distance W_2 [37, 2]. The existence of a tangent space, and more importantly the ability to access the logarithm map via optimal transport, allows us to use "linearized optimal transport" [50, 38, 36, 15, 33] to extend the intuition of ordinary techniques to this new regime. In particular, one can now view a time-evolving measure as a curve on the Wasserstein manifold, and the correct derivative object is the tangent vector to that curve, which we can compute by lifting the curve to the tangent space and computing a derivative, just as we would on any manifold. With this formalism, many standard methods from numerical analysis now have natural analogs, like Euler's method for building timesteppers [21] or Newton's method

for finding steady states [48]. Similarly, we can use the logarithm map to lift a collection of data points on the Wasserstein manifold to a common Hilbert space, on which we can now perform any standard machine learning method we want [29]. This process of "linearizing" the geometry of measure-valued data allows us to motivate natural extensions of these standard methods to the measure-valued setting.

Current Work and Progress

Numerical Analysis on Temporally Evolving Measures. We can use the formalism of the Wasserstein manifold to interpret a probability measure evolving "continuously" in time as a continuous curve on the manifold. This perspective, together with a notion of geodesics and tangent vectors, allows one to use optimal transport maps to compute an appropriate "derivative" of such a curve. Formally, given a "smooth" curve $\mu : [0, T] \to \mathcal{P}_2(\mathbb{R}^d)$, written $t \mapsto \mu_t$, on the Wasserstein manifold, there exists a (time-dependent) tangent field $\mathbf{v}_t : \mathbb{R}^d \to \mathbb{R}^d$ that represents the derivative of μ via the continuity equation

$$\partial_t \mu_t + \nabla \cdot (\mathbf{v}_t \mu_t) = 0,$$

and if μ_t has a density at all times t, then we can explicitly compute $\mathbf{v}_t$ as a limit of optimal transport maps:

$$\mathbf{v}_t = \lim_{h \to 0} \frac{T_{\mu_t}^{\mu_{t+h}} - \mathrm{id}}{h}$$

(where $T_{\mu_t}^{\mu_{t+h}}$ denotes the optimal transport map from μ_t to μ_{t+h}) [2]. This gives a tangent vector to the curve μ that can be used to construct a geodesic starting at μ_t in the direction of $\mathbf{v}_t$, which gives rise to a measure-valued analog of Euler's method (see Figure 1). We make this intuition precise in [21], and we prove (under hypotheses about the regularity of the curve) that the method gives a first-order accurate approximation of the true curve with error bounds similar to those of ordinary Euler's method. We also use this intuition to build a macro-scale "timestepper" from a micro-scale one [3, 10, 13, 14, 23, 25, 27]. If we have a model that simulates a collection of particles forward in time for a small time step, then we can use the optimal transport map from the original to the updated distribution to compute a finite-difference approximation $\mathbf{v}_t^h$ of the true tangent field $\mathbf{v}_t$. We use $\mathbf{v}_t^h$ as a surrogate for $\mathbf{v}_t$ in Euler's method, which gives an Euler-type method that takes a much larger time step than one step of the micro-scale model. This "macro-scale" timestepper gives an accurate way to predict the evolution of the system a long time in the future without needing to run the micro-scale model for the equivalent duration. The power of our approach is that it requires only *local* data to take a large time step, whereas other similar approaches focus on learning the action of the micro-scale model *globally* [32, 16, 17, 34, 24, 47, 4]. In this way, our method is more similar to that of Wasserstein gradient flows [19, 35], except that we are able to approximate arbitrary curves of measures instead of only an optimal curve induced by some functional. In [21], we experimentally demonstrate (see Figure 1) that our method outperforms other approaches for predicting the evolution of example particle systems (a bacterial chemotaxis model, a fluid flow model, and Langevin dynamics) when we have limited data from the micro-scale model. Our method is also the first to easily generalize to the case of multiple dimensions, which we similarly illustrate in the paper. Thus, we have built a novel macro-scale timestepper which can now be used to do a number of system-level tasks like control and identification,

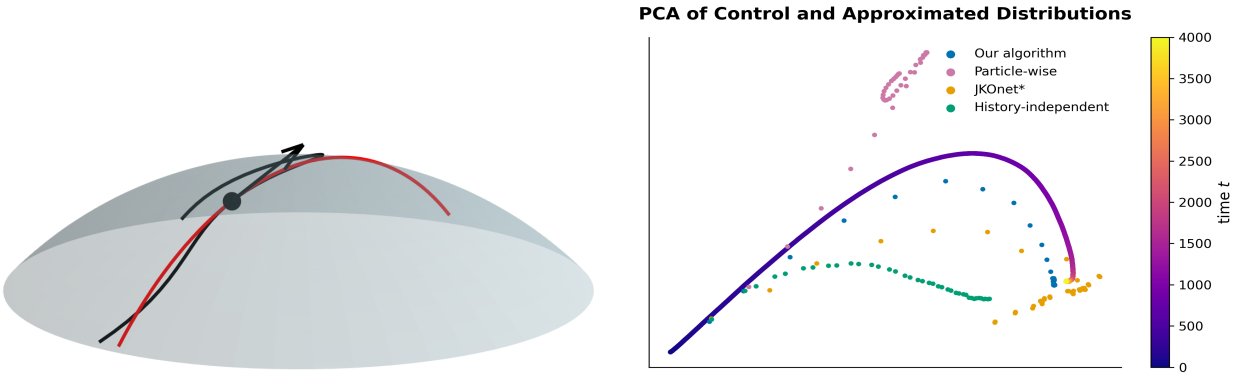

FIGURE 1. (Left) Schematic of one step of Euler's method for approximating a true curve (black) with a geodesic (red) in the direction of the tangent vector. (Right) Dimension reduction that shows our method (blue) more accurately following the trajectory of the true evolution of bacterial chemotaxis (rainbow curve) than other similar methods.

analyzing stability, and approximating steady states [45, 10, 14]. We detail an approach for efficiently approximating steady states in [48].

The fundamental observation that drives the work in [48] is that we can extend the analogy further than first-order methods. By taking more than one step with the micro-scale model, we can approximate the tangent vector at multiple initial measures, which then allows us to estimate a "second derivative." This opens up the possibility of developing analogous second-order methods for the equivalent of ODEs on the Wasserstein manifold. In [48], we present a Newton-type method that uses our optimal transport–based timestepper from [21] to approximate distributions which are relative steady states of the micro-scale model. However, since we only have access to a micro-scale timestepper, we are not able to approximate the Jacobian of the system itself, so we combine our method with a Krylov-subspace approach for solving a linear system [9, 22, 41]. With this approach, we only need to know the action of the Jacobian on particular perturbations, which we can approximate by perturbing the initial measure, applying the micro-scale timestepper, and computing a velocity field via optimal transport. In [48], we detail this approach, and we show how it can be used to construct new, smoother timesteppers using inverse CDFs, which we then use to quickly find steady-state distributions. We show that our method drastically reduces the number of micro-scale simulations steps required to find steady-state distributions of many example particle system models, such as bacterial chemotaxis, economic agents, and Langevin dynamics in both one and two dimensions. This shows that our timestepper is not just good for more efficiently and accurately simulating trajectories of curves of measures — it can be used to do many system-level tasks, and efficiently computing steady states is just one such application.

Machine Learning on Data Sets of Measures. We also use the formalism of the Wasserstein manifold to motivate a natural approach for "linearizing" the space of probability measures, which facilitates applying standard machine learning techniques on measure-valued data. For a fixed measure $\sigma \in \mathcal{P}_2(\mathbb{R}^d)$, the logarithm map into the tangent space

at σ is given by $\mu \mapsto T_\sigma^\mu - \mathrm{id}$, which means we can embed measures into a common linear space by computing optimal transport maps from a fixed reference [2, 43]. Intuitively, this embedding process "vectorizes" the measure-valued data in a canonical way. Using these embeddings, we can build linear classifiers, generate new data using weighted averages, and more. We also get a free approximation of the optimal transport distance between measures by computing the distance between their linear embeddings in the tangent space, and we call this approximation the "linearized optimal transport" (LOT) distance [36, 7]:

$$W_{2,\sigma}^{\mathrm{LOT}}(\mu,\nu) = \|T_\sigma^\mu - T_\sigma^\nu\|_{L^2(\sigma)} = \left(\int_{\mathbb{R}^d} \|T_\sigma^\mu(x) - T_\sigma^\nu(x)\|^2 \, d\sigma(x)\right)^{\frac{1}{2}}.$$

We can use this LOT "distance" to more efficiently run unsupervised clustering algorithms since it does not require computing an optimal transport map for each pair of points to obtain pairwise distances. We showcase the power of this approach in [29], in which we detail a Python library (pyLOT) we built to do many standard LOT methods and demonstrate its utility on an example dataset of 3D scans of primate teeth. In the paper, we see impressive scores for various machine learning tasks, like clustering teeth into primate species, classifying teeth by species, and generating new examples of teeth for each species.

While the formalism is less clear in the unbalanced case, we can also use the intuition developed for LOT to extend many of the methods to the case where the measures have different total mass. In such cases, there is a generalized notion of an optimal transport map which can be used in a similar way [6]. In particular, we can use unbalanced Wasserstein barycenters to build a dictionary for clustering, and we apply this approach to give unsupervised labeling of hyperspectral images [26]. We treat each hyperspectral image as a measure on the space of reflectances, and since these generically have different total mass, we either need to normalize each HSI spectrum or use unbalanced optimal transport. In [26], we demonstrate the practicality of the unbalanced approach and discuss the advantages and disadvantages to the normalized approach, which is theoretically more justified but loses important data about the total reflectivity [11, 12]. This loss of data suggests that generalizing of the geometry of the Wasserstein manifold to the unbalanced case could motivate better approaches, not only for hyperspectral images, but for many more unbalanced problems [6, 28].

General Machine Learning and AI. My recent bend towards optimal transport intersects nicely with my general interest in machine learning and artificial intelligence research in the form of diffusion models. One can understand the "forward process" in diffusion models as learning a nice way to push forward the unknown distribution to a known distribution (usually a Gaussian), and similarly the "reverse process" attempts to generate new samples from the known distribution and push them forward [18, 42]. These processes inherently involve distributions "evolving in time" (where time here is the process of (de-)noising), which means one can understand them in terms of the geometry of the Wasserstein manifold. This understanding also extends to the underlying embedding spaces, which can also be understood as spaces of distributions (i.e., point clouds).

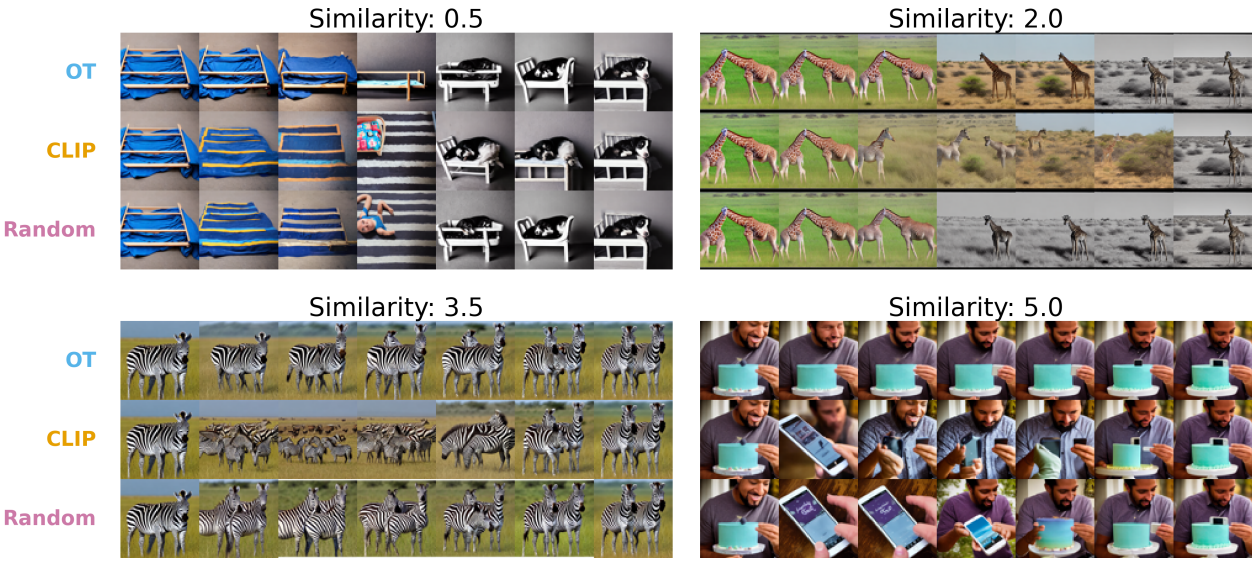

FIGURE 2. Images generated using equally spaced embeddings along different interpolating paths through embedding space.

In [20], we explore the geometry of embedding spaces by constructing different interpolating paths between embeddings and investigating how they affect the resulting image interpolations. As shown in Figure 2, we notice that the interpolations given by optimal transport give more "natural looking" image interpolations than other interpolation methods, which suggests that the correct way to understand the geometry of embedding space is as a space of measures. Our work hints at a deeper "distributional" perspective on diffusion models that may be helpful for understanding their behavior, training process, and vulnerabilities.

FUTURE DIRECTIONS

General Numerical Analysis on the Wasserstein Manifold. Our work in [21, 48], particularly the naturality of the analog with standard numerical methods in Euclidean space, suggests that many more numerical methods exist for analyzing curves of measures on the Wasserstein manifold. While Newton's method is one particular second-order method, there are other methods for, e.g., giving second-order approximations of curves in $\mathbb{R}^d$. The natural way to give a second-order approximation of a curve on $\mathcal{P}_2(\mathbb{R}^d)$ is to use the derivative of the tangent vector field $\mathbf{v}_t$, which is subtle to compute. Indeed, the obvious candidate of $\frac{d}{dt}v_t(x)$ is not the correct second derivative object to use in a second-order approximation. Instead, the relevant object is the second derivative of the flow curve $\gamma_x(t)$, which we use to prove our error bound of Euler's method in [21]. Better understanding its behavior would not only give a natural formulation for a second-order method, but it would also suggest ways to estimate an appropriate step size in Euler's method based only on local data, which would further improve our analysis.

There is also the question of how to appropriately balance the signal to noise ratio in the computation of the approximate vector field $\mathbf{v}_t^h$ used in our Euler-type method for particle systems. Experimentally, as evidenced by Figure 3, we observe that the accuracy of our

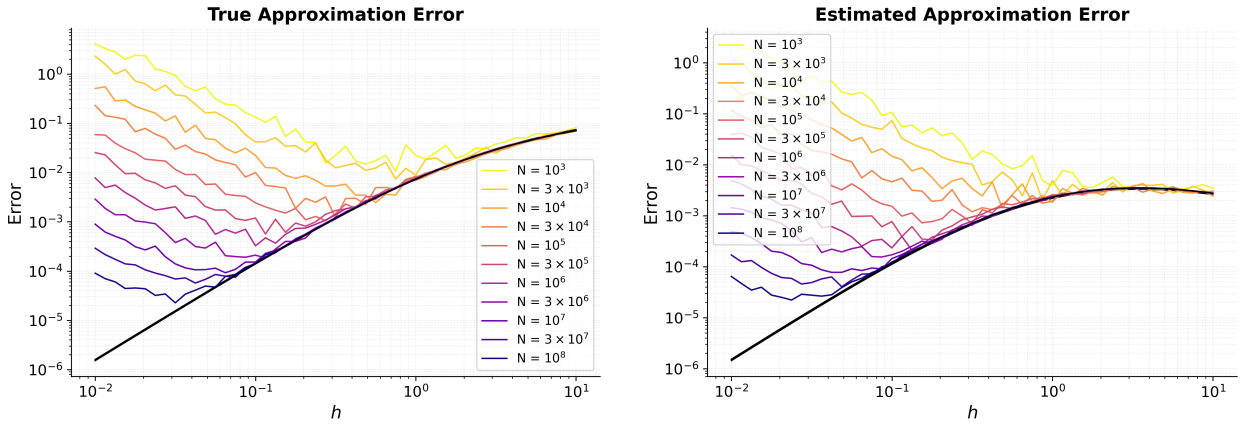

FIGURE 3. (Left) Approximation error of our vector field $\mathbf{v}_t^h$ versus the true underlying vector field $\mathbf{v}_t$ with different numbers of particles N. (Right) Estimation of the approximation error using the difference of consecutive approximations $\mathbf{v}_t^h$ and $\mathbf{v}_t^{2h}$.

approximation increases as the number of particles increases, but for a fixed number of particles, there is tension on the choice of h — for large h, there is too much error in the finite difference approximation of the derivative, but for small h, there is too much noise in the micro-scale timestepper. This is because the paths of the individual particles given by the micro-scale timestepper are not differentiable (and taking independent samples of the new distribution instead of pushing particles forward is even worse because then the particle-wise paths are not even continuous). Despite the apparent tension, we strongly believe that there is a balance between the number of particles and the time step h that makes our method successful. Moreover, we believe we can identify this balance without knowing the true vector field simply by computing how quickly the approximated vector field changes as we take additional steps. Developing the statistical rates needed to make these relationships precise is a subject of ongoing work.

Computational and Statistical Theory for Nearby OT Problems. The setting of particle timesteppers that we discuss in [21, 48] also motivates a natural question: Can one use knowledge of optimal transport maps of nearby distributions to speed up the computation of optimal transport maps for new distributions? In the computations in both papers, we are required to compute various optimal transport maps between slightly perturbed distributions. These computations are the most expensive computational bottleneck in our methods (especially in multiple dimensions), and the similarity of both the source and target distributions suggests that the optimal transport maps should also be similar [49]. Thus, one might hope that "warm starting" the computation with a previously-computed OT map could drastically reduce the number of iterations required to compute the new OT maps. If so, then we could not only improve our computational advantage over other methods, but we would also have a way to improve the efficiency of computing optimal transport maps in general, which has impacts far beyond our current work [8, 39].

Learning Diffusion Models as Evolving Measures. Our work in [21, 48] also intersects nicely with our work in [20] to raise questions about how optimal transport (and more

specifically particle timesteppers) can be used to understand how diffusion models are trained. There are various flow-based methods for diffusion models [46, 52, 30, 31, 1] which take a Lagrangian view on the evolution of the data in the forward and reverse processes. While optimal transport will likely never be a leading method by which diffusion models are trained, it has potential to elucidate understanding of the forward and reverse processes because of its ability to make sense of particle evolution, even when individual particle trajectories are not "smooth". In a similar way to our work in [20], a new distributional perspective could give insights into the training, the overall data landscape, and the specific behavior of the denoising processes that could hint at more useful theoretical guarantees for diffusion models.

References

[1] Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions, 2023.

[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics. ETH Zürich. Birkhäuser, Basel, 2nd edition, 2008.

[3] Hassan Arbabi and Ioannis G. Kevrekidis. Particles to partial differential equations parsimoniously. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 31(3):033137, 03 2021.

[4] Riccardo Bonalli and Alessandro Rudi. Non-parametric learning of stochastic differential equations with non-asymptotic fast rates of convergence. *Foundations of Computational Mathematics*, 25, 2025.

[5] R. Qi Charles, Hao Su, Mo Kaichun, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 77–85, 2017.

[6] Lénaïc Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Unbalanced optimal transport: Dynamic and kantorovich formulations. *Journal of Functional Analysis*, 274(11):3090–3123, 2018.

[7] Alexander Cloninger, Keaton Hamm, Varun Khurana, and Caroline Moosmüller. Linearized Wasserstein dimensionality reduction with approximation guarantees. *Applied and Computational Harmonic Analysis*, 74:101718, 2025.

[8] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013.

[9] Gianluca Fabiani, Nikolaos Evangelou, Tianqi Cui, Juan M Bello-Rivas, Cristina P Martin-Linares, Constantinos Siettos, and Ioannis G Kevrekidis. Task-oriented machine learning surrogates for tipping points of agent-based models. *Nature communications*, 15(1):4117, 2024.

[10] Gianluca Fabiani, Hannes Vandecasteele, Somdatta Goswami, Constantinos Siettos, and Ioannis G. Kevrekidis. Enabling local neural operators to perform equation-free system-level analysis, 2025.

[11] Scott Fullenbaum, Marshall Mueller, Abiy Tasissa, and James M. Murphy. Hyperspectral image clustering via learned representation in wasserstein space. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 8791–8796, 2024.

[12] Scott Fullenbaum, Marshall Mueller, Abiy Tasissa, and James M. Murphy. Nonlinear unmixing of hyperspectral images via regularized wasserstein dictionary learning. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 8289–8294, 2024.

[13] C. W. Gear, D. Givon, and I. G. Kevrekidis. Virtual slow manifolds: The fast stochastic case. *AIP Conference Proceedings*, 1168(1):17–20, 09 2009.

[14] C. W. Gear, Ioannis G. Kevrekidis, and Constantinos Theodoropoulos. 'Coarse' integration/bifurcation analysis via microscopic simulators: Micro-Galerkin methods. *Computers & Chemical Engineering*, 26(7–8):941–963, 2002.

[15] Nicola Gigli. On Hölder continuity-in-time of the optimal transport map towards measures along a curve. *Proceedings of the Edinburgh Mathematical Society*, 54(2):401–409, 2011.

[16] Somdatta Goswami, Aniruddha Bora, Yue Yu, and George Em Karniadakis. Physics-informed deep neural operator networks. In Timon Rabczuk and Klaus-Jürgen Bathe, editors, *Machine Learning in*

Modeling and Simulation: Methods and Applications, Computational Methods in Engineering & the Sciences, pages 219–254. Springer, Cham, 2023.

[17] Somdatta Goswami, Katiana Kontolati, Michael D. Shields, and George Em Karniadakis. Deep transfer operator learning for partial differential equations under conditional shift. *Nature Machine Intelligence*, 4(12):1155–1164, 2022.

[18] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020.

[19] Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the fokker–planck equation. *SIAM Journal on Mathematical Analysis*, 29(1):1–17, 1998.

[20] Nicholas Karris, Luke Durell, Javier Flores, and Tegan Emerson. Which way from b to a: The role of embedding geometry in image interpolation for stable diffusion. *TAG–Data Science Conference*, 2025.

[21] Nicholas Karris, Evangelos A. Nikitopoulos, Ioannis G. Kevrekidis, Seungjoon Lee, and Alexander Cloninger. Using linearized optimal transport to predict the evolution of stochastic particle systems. *Preprint, arXiv:2408.01857*, 2025.

[22] Carl T. Kelley, Ioannis G. Kevrekidis, and Liang Qiao. Newton-krylov solvers for time-steppers. *arXiv preprint math/0404374*, 2004.

[23] Ioannis G. Kevrekidis, C. William Gear, James M. Hyman, Panagiotis G. Kevrekidis, Olof Runborg, and Constantinos Theodoropoulos. Equation-free, coarse-grained multiscale computation: Enabling microscopic simulators to perform system-level analysis. *Communications in Mathematical Sciences*, 1(4):715–762, 2003.

[24] Nikola B. Kovachki, Samuel Lanthaler, and Hrushikesh Mhaskar. Data complexity estimates for operator learning. Preprint, arXiv:2405.15992 [cs.LG], 2024.

[25] Seungjoon Lee, Yorgos M. Psarellis, Constantinos I. Siettos, and Ioannis G. Kevrekidis. Learning black- and gray-box chemotactic PDEs/closures from agent based Monte Carlo simulation data. *Journal of Mathematical Biology*, 87(1):15, 2023.

[26] Joshua Lentz, Nicholas Karris, Alexander Cloninger, and James M.Murphy. Unbalanced optimal transport dictionary learning for unsupervised hyperspectral image clustering. In *2025 15th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2025.

[27] Ju Li, Panayotis G. Kevrekidis, C. William Gear, and Ioannis G. Kevrekidis. Deciding the nature of the coarse equation through microscopic simulations: The baby-bathwater scheme. *Multiscale Modeling & Simulation*, 1(3):391–407, 2003.

[28] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal transport in competition with reaction: The hellinger–kantorovich distance and geodesic curves. *SIAM Journal on Mathematical Analysis*, 48(4):2869–2911, 2016.

[29] Jun Linwu, Varun Khurana, Nicholas Karris, and Alexander Cloninger. Linearized optimal transport pyLOT library: A toolkit for machine learning on point clouds. *Preprint: arXiv:2502.03439*, 2025.

[30] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations*, 2023.

[31] Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023.

[32] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3:218–229, 2021.

[33] Quentin Mérigot, Alex Delalande, and Frédéric Chazal. Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 3186–3196. PMLR, 2020.

[34] H. N. Mhaskar. Local approximation of operators. *Applied and Computational Harmonic Analysis*, 64:194–228, 2023.

[35] Petr Mokrov, Alexander Korotin, Lingxiao Li, Aude Genevay, Justin M Solomon, and Evgeny Burnaev. Large-scale wasserstein gradient flows. *Advances in Neural Information Processing Systems*, 34, 2021.

[36] Caroline Moosmüller and Alexander Cloninger. Linear optimal transport embedding: Provable Wasserstein classification for certain rigid transformations and perturbations. *Information and Inference: A Journal of the IMA*, 12(1):363–389, 2023.

[37] Felix Otto. The geometry of dissipative evolution equations: The porus medium equation. *Communications in Partial Differential Equations*, 26(1–2):101–174, 2001.

[38] Se Rim Park, Soheil Kolouri, Shinjini Kundu, and Gustavo K. Rohde. The cumulative distribution transform and linear pattern classification. *Applied and Computational Harmonic Analysis*, 45(3):616 – 641, 2018.

[39] Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5–6):355–607, 2019.

[40] Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: deep hierarchical feature learning on point sets in a metric space. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, page 5105–5114, 2017.

[41] Liang Qiao, Radek Erban, Carl T. Kelley, and Ioannis G. Kevrekidis. Spatially distributed stochastic systems: Equation-free and equation-assisted preconditioned computations. *The Journal of chemical physics*, 125(20), 2006.

[42] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10674–10685, 2022.

[43] Filippo Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, volume 87 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhäuser, Cham, 2015.

[44] S. Setayeshgar, C. W. Gear, H. G. Othmer, and I. G. Kevrekidis. Application of coarse integration to bacterial chemotaxis. *Multiscale Modeling & Simulation*, 4(1):307–327, 2005.

[45] C. I. Siettos, C. C. Pantelides, and I. G. Kevrekidis. Enabling dynamic process simulators to perform alternative tasks: A time-stepper-based toolkit for computer-aided analysis. *Industrial & Engineering Chemistry Research*, 42:6795–6801, 2003.

[46] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations, 2021.

[47] Antonio Terpin, Nicolas Lanzetti, Martín Gadea, and Florian Dorfler. Learning diffusion at lightspeed. In *The Thirty-Eighth Annual Conference on Neural Information Processing Systems*, 2024.

[48] Hannes Vandecasteele, Nicholas Karris, Alexander Cloninger, and Ioannis G. Kevrekidis. Newton-Krylov methods for computing steady states of particle timesteppers via optimal transport. *In preparation.*, 2025.

[49] C. Villani. *Optimal Transport: Old and New*. Grundlehren der Mathematischen Wissenschaften. Springer Berlin Heidelberg, 2009.

[50] Wei Wang, Dejan Slepčev, Saurav Basu, John A. Ozolek, and Gustavo K. Rohde. A linear optimal transportation framework for quantifying and visualizing variations in sets of images. *International Journal of Computer Vision*, 101:254–269, 2013.

[51] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (TOG)*, 2019.

[52] Qinsheng Zhang and Yongxin Chen. Diffusion normalizing flow. In *Advances in Neural Information Processing Systems*, 2021.