Multimarginal Flow Matching with Adversarially Learnt Interpolants

Oskar Kviman[*1] **Kirill Tamogashev**[*2] **Nicola Branchini**[2]
Víctor Elvira[2] **Jens Lagergren**[1] **Nikolay Malkin**[2,3]
[*]Equal contribution [1]KTH, Stockholm [2]University of Edinburgh [3]CIFAR Fellow

Abstract

Learning the dynamics of a process given sampled observations at several time points is an important but difficult task in many scientific applications. When no ground-truth trajectories are available, but one has only snapshots of data taken at a few discrete time steps, the problem of modelling the dynamics – and thus inferring the underlying trajectories – can be solved by multimarginal generalisations of flow matching algorithms. This paper proposes a novel flow matching method that overcomes certain limitations of existing multimarginal trajectory inference algorithms. Our proposed method, ALI-CFM, uses a GAN-inspired adversarial loss to fit neurally parametrised interpolant curves between source and target points such that the marginal distributions at intermediate time points are close to the observed distributions. The resulting interpolants are smooth trajectories that, as we show, are unique under mild assumptions. These interpolants are subsequently marginalised by a flow matching algorithm, yielding a trained vector field for the underlying dynamics. We showcase the versatility and scalability of our method by achieving strong results on trajectory prediction in single-cell RNA sequencing data.

1 Introduction

Modelling the time-dependent dynamics of a system given experimental observations is a central task in many scientific problems in biology (see Schiebinger et al. [2019], Bunne et al. [2023]), medicine (see Oeppen and Vaupel [2002], Hay et al. [2021]), and other areas. The problem involves a collection of data snapshots taken at various time steps that together provide an empirical account of some process. Examples of such processes include recordings of health measurements, evolution of a disease [Waddington, 1942, Hay et al., 2021], and single-cell RNA sequencing (scRNA-seq; Macosko et al. [2015], Klein et al. [2015]).

Formally, a (deterministic) system in $\mathbb{R}^n$ can be described by an ordinary differential equation (ODE) $dx_t = v_t(x_t)\,dt$, where v_t is a time-dependent vector field. Given a pair of marginal distributions q_0, q_1, we aim to find v_t such that the ODE's integration map from $t = 0$ to $t = 1$ pushes q_0 to q_1. In the multimarginal case, the marginal distributions p_t induced by the dynamics with initial conditions $p_0 = q_0$ should also satisfy intermediate conditions, namely, $p_{t_i} = q_{t_i}$ for a set of times $0 = t_0 < t_1 < \cdots < t_K = 1$. The intermediate distributions q_{t_i} are provided by datasets of samples.

While this problem can be solved by applying flow matching [Lipman et al., 2023, Albergo et al., 2023, Liu et al., 2023] to learn flows for every pair of consecutive marginals $q_{t_i}, q_{t_{i+1}}$ [Tong et al., 2024], several methods specialised to multimarginal problems have also been proposed. One work [Rohbeck et al., 2025] proposes to use cubic splines to build trajectories that pass through samples from intermediate distributions. However, spline interpolation methods do not seem to scale well to high dimensions Lee et al. [2025]. Another line of work [Neklyudov et al., 2024, Kapuśniak et al., 2024] assumes that the samples from all of the target marginal distributions q_{t_i} follow a common

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Frontiers in Probabilistic Inference: Sampling Meets Learning.

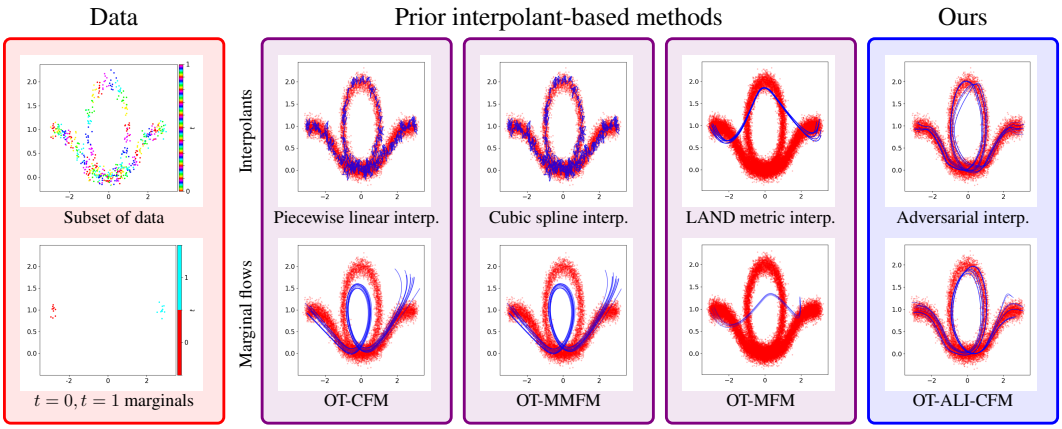

Figure 1: Comparison of CFM Tong et al. [2024], MFM Kapuśniak et al. [2024], MMFM Rohbeck et al. [2025] and our ALI-CFM method on a synthetic 2D 'knot' distribution. See §3 for details.

geometry. These methods fit an interpolant that can later be used to learn the underlying dynamics. However, they are not suitable when the data geometry varies with time (Fig. 1).

To overcome the limitations of previous methods, we propose to learn the interpolations between distributions using a GAN-like adversarial objective Goodfellow et al. [2014], Huang et al. [2024] called adversarially learnt interpolants (ALI). ALI directly matches the target intermediate marginals with those of the learnt interpolants. The interpolants can then be marginalised by a conditional flow matching (CFM) loss, allowing us to model complex time-dependent behaviour while using available intermediate-time information. The full algorithm, called ALI-CFM, approximates the data distribution at each time step, as opposed to prior methods that rely on the provided samples explicitly, which makes our method especially useful in cases where the provided dataset is noisy.

Our contributions in this paper are the following:

(1) We propose the ALI objective for learning of interpolants that applies to multimarginal problems with either a discrete set of time snapshots or those in continuous time.
(2) We show the versatility of ALI-CFM on synthetic examples that clearly depict its advantages.
(3) We demonstrate the scalability of ALI-CFM on a scRNA-seq trajectory inference problem.

2 Method

Background. We summarise CFM, roughly following the setting and exposition of Lipman et al. [2023], Tong et al. [2024], Pooladian et al. [2023]. All statements below hold under regularity conditions whose details are not important in this paper.

We assume a dynamical system in $\mathbb{R}^n$ given by an ordinary differential equation (ODE) $dx_t = v_t(x_t)\,dt$. Its integration map $\psi_t : \mathbb{R}^n \to \mathbb{R}^n$ from time 0 to time t satisfies

$$\frac{d}{dt}\psi_t(x_0) = v_t(\psi_t(x_0)), \quad \psi_0(x_0) = x_0. \tag{1}$$

The integration map, together with initial conditions $\psi_0(x_0) = x_0 \sim p_0$, defines a probability path $p_t = (\psi_t)_{\#}p_0$, where p_t is the marginal distribution of x_t.

In the bimarginal FM setting, one observes samples from two marginal distributions, $x_0 \sim q_0$ and $x_1 \sim q_1$, and fixes a vector field v_t such that its integration map ψ_t satisfies $(\psi_1)_{\#}q_0 = q_1$. (This vector field is not tractably computable, but is described by interpolants, as we discuss below.) Having fixed v_t as a target, the goal is to approximate it with a neural net u_t^θ with weights θ. This could be done via the FM objective

$$L_{\text{FM}} = \mathbb{E}_{t\sim U[0,1], x\sim p_t}\|u_t^\theta(x) - v_t(x)\|_2^2. \tag{2}$$

However, since v_t and p_t are intractable to compute and sample from, the loss (2) is intractable.

Instead, one assumes a family of *interpolant* curves, one for every pair x_0, x_1 in the support of $q_0 \otimes q_1$. These curves are denoted $G(x_0, x_1, t)$ and should satisfy $G(x_0, x_1, t) = x_t$ for $t = 0, 1$. For example, a linear interpolant is described by $G(x_0, x_1, t) = tx_1 + (1-t)x_0$. We define

$v_t(x_t \mid x_0, x_1) := \frac{d}{dt} G(x_0, x_1, t)$. For any joint distribution π over $\mathbb{R}^n \times \mathbb{R}^n$ whose marginals are q_0 and q_1, respectively, it can then be shown that the marginal vector field

$$v_t(x_t) := \mathbb{E}\left[v_t(x_t \mid x_0, x_1) \mid x_t = G(x_0, x_1, t)\right], \quad (x_0, x_1) \sim \pi, \tag{3}$$

pushes q_0 to q_1 and can thus be used as the learning target for u_t^θ. The marginals p_t of the resulting dynamics are tractably sampled by drawing $(x_0, x_1) \sim \pi$ and set $x_t = G(x_0, x_1, t)$. While v_t itself is still not tractable, one can replace (2) by the following *conditional flow matching* (CFM) objective:

$$L_{\text{CFM}} = \mathbb{E}_{t\sim U[0,1],(x_0,x_1)\sim\pi} \|u_t^\theta(x_t) - v_t(x_t|x_0, x_1)\|_2^2, \text{where } x_t = G(x_0, x_1, t). \tag{4}$$

It is easily shown that the gradients of (4) and (2) coincide, and (4) thus provides a tractable way to learn the target vector field Lipman et al. [2023].

The above setting leaves two choices open: the **coupling** π and the **interpolants** G. Past work has proposed an independent coupling $\pi = q_0 \otimes q_1$, giving objectives equivalent to those in Liu et al. [2023], Albergo et al. [2023], or couplings computed by (possibly minibatch or entropic) optimal transport Tong et al. [2024], Pooladian et al. [2023]. The latter has been shown to result in straighter integration curves and solve the dynamic optimal transport problem. For the interpolants, linear and trigonometric Albergo et al. [2023] curves have been proposed, as well as those trained to pass through areas of high data distribution density Kapuśniak et al. [2024]. In the multimarginal setting, piecewise linear Tong et al. [2024] and cubic spline Rohbeck et al. [2025] interpolants have been used. See Appendix A for a discussion.

Adversarial learning of interpolants. We move from the bimarginal to the multimarginal setting, where data has been collected from a sequence of K marginal distributions, $x_{t_i} \sim q_{t_i}$ with corresponding time stamps $0 = t_0 < t_1 < \cdots < t_K = 1$. Given a coupling π between q_0 and q_1 and a neural network f_ϕ, we model interpolants of the following form

$$G_\phi(x_0, x_1, t) = (1-t)x_0 + tx_1 + t(1-t)f_\phi(x_0, x_1, t). \tag{5}$$

Our aim is to match the intermediate distributions at t_i of the interpolants when $(x_0, x_1) \sim \pi$ to the given marginals q_{t_i}, that is, to enforce

$$(G_\phi(\cdot, \cdot, t_i))_\# \pi = q_{t_i}. \tag{6}$$

(The parametrisation (5) guarantees $G_\phi(x_0, x_1, t) = x_t$ for $t = 0, 1$, so (6) holds automatically for $i = 0, K$.) In order to approximately enforce (6), we use an adversarial learning scheme. Let $D_\gamma(x_t, t)$ be a second neural network, tasked to discriminate between marginal samples, $x_t \sim q_t$, and the learnable interpolants in (5). Optimising the min-max GAN objective [Goodfellow et al., 2014] for each t_i,

$$\min_{G_\phi} \max_{D_\gamma} \underbrace{\mathbb{E}_{(x_0,x_1)\sim\pi}\left[\log(1 - D_\gamma(G_\phi(x_0, x_1, t_i), t_i))\right] + \mathbb{E}_{q_{t_i}}\left[\log D_\gamma(x_{t_i}, t_i)\right]}_{L_{\text{GAN}}(G_\phi, D_\gamma; t_i)}, \tag{7}$$

is then equivalent to minimising the Jensen-Shannon divergence Goodfellow et al. [2014] between q_{t_i} and, in this case, $G_\phi(\cdot, \cdot, t_i)_\# \pi$. Notably, our 'generators' – the interpolants – are conditioned on scalar-valued time inputs, associated with the targeted q_{t_i}. While the noise in GANs typically comes from a fixed distribution, in (7) the pair $(x_0, x_1) \sim \pi$ plays the role of 'noise'.

The solutions to the min-max problem in (7) are not unique and can induce arbitrarily curved interpolants. To this end, we introduce a regularising term in the learning objective which penalises deviations from the linear interpolant between coupled samples from the end-marginals. Letting $\ell(x_0, x_1, t) = (1-t)x_0 + tx_1$, we define

$$L_{\text{reg}}(G_\phi; t_i) = \mathbb{E}_{(x_0,x_1)\sim\pi}\left[\|G_\phi(x_0, x_1, t_i) - \ell(x_0, x_1, t)\|^2\right] \tag{8}$$

The problem of matching marginals (6) while minimising the regulariser (8) enjoys unique solutions:

Theorem 2.1. *Fix $t \in (0, 1)$, q_t, and a coupling π between q_0 and q_1 such that the distribution $\ell(\cdot, \cdot, t)_\#\pi$ is absolutely continuous (a.c.) w.r.t. the Lebesgue measure. Then the interpolant $G(\cdot, \cdot, t) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ minimising*

$$\mathbb{E}_{(x_0,x_1)\sim\pi}\|G(x_0, x_1, t) - \ell(x_0, x_1, t)\|^2 \tag{9}$$

subject to $G(\cdot, \cdot, t)_\#\pi = q_t$, exists and is unique on the support of π up to almost-everywhere equality.

We provide the proof in Appendix B. Note that the assumption of an a.c. target interpolant is satisfied under a number of conditions, e.g., if π is the solution to entropic OT with squared-euclidean cost (or either of its limiting cases: (nonentropic) OT or the independent coupling) and either q_0 or q_1 is a.c.

Table 1: Trajectory inference on 5D PCA scRNA-seq data. Accuracy measured in EMD (smaller is better) w.r.t. the left-out marginal distributions, averaged over five independent runs.

Algorithm ↓ | Dataset →	Cite	EB	Multi
I-CFM	1.236±0.050	1.156±0.42	1.150±0.091
OT-CFM	1.142±0.085	0.809±0.16	0.975±0.045
OT-MFM	0.793±0.019	0.711±0.050	0.890±0.123
I-MMFM (Cubic splines)	2.068±0.390	4.740±0.650	1.528±0.040
OT-MMFM (Cubic splines)	1.099±0.043	3.530±0.194	1.807±0.085
OT-ALI-CFM (ours)	0.910±0.024	0.742±0.022	0.925±0.018

Table 2: Trajectory inference on 50D and 100D PCA scRNA-seq data. The accuracies here are measured in the same way as in Table 1.

Dim. →	50		100	
Algorithm ↓ | Dataset →	Cite	Multi	Cite	Multi
I-CFM	42.478±0.930	51.098±0.340	49.929±0.391	57.801±0.365
OT-CFM	38.367±0.295	47.205±0.184	45.148±0.207	54.630±0.456
I-MFM	41.172±0.269	48.415±0.793	46.339±0.618	53.667±0.768
OT-MFM	36.471±0.480	45.879±0.438	42.232±0.249	51.169±0.523
OT-ALI-CFM (ours)	42.977±0.739	51.55±0.511	49.363±0.771	58.347±0.449

Complete method: ALI-CFM. The loss (7) and the above result motivate the full *adversarially learnt interpolants (ALI) objective*:

$$L_{\text{ALI}}(G_\phi, D_\gamma) = \mathbb{E}_{i\sim\text{Unif}(\{1,\dots,K-1\})}\left[L_{\text{GAN}}(G_\phi, D_\gamma; t_i) + \lambda L_{\text{ref}}(G_\phi; t_i)\right], \tag{10}$$

where $\lambda > 0$ is a regularisation weight.

Once the interpolants G_ϕ have been trained using ALI, they can be marginalised using the CFM objective (4), using the same coupling π.[1] At convergence, this yields a dynamical system, defined by a vector field u_t^θ, whose marginals p_t at each t match the interpolants' marginals $(G_\phi(\cdot,\cdot,t))_\#\pi$. If learnt perfectly, the resulting flow solves the multimarginal transport problem: $p_{t_i} = q_{t_i}$ for all i.

We refer to the complete method as ALI-CFM: it consists of (i) learning interpolants G_ϕ using ALI, then (ii) marginalising them to yield the time-dependent vector field u_t^θ. The prefixes I- and OT- in the algorithm names specify if the coupling π uses an independent or a (minibatch) optimal transport plan.

3 Experiments

Synthetic data. First, we showcase the flexibility of our ALI-CFM method on synthetic data: a sequence of 1200 marginal distributions centred along a knot (Fig. 1). The experiment shows that ALI-CFM is the only method capable of accurately capturing the time-dependent geometry of the 'knot' distribution. See Appendices C and D for details of experiment setup.

Single-cell trajectory inference. As a real-world example, we experiment with time-series scRNA-seq data, predicting the gene expression levels of cell populations that evolve over time. The samples from the K consecutive populations are naturally unpaired, and the dynamics of the gene expressions are unknown. We follow preceding works Tong et al. [2024] on predicting the marginal distributions of held-out intermediate-time data, measuring accuracy via the earth-mover distance (EMD) between the CFM trajectories and the held-out data at the corresponding time stamp. That is, using the remaining data, we train ALIs by optimising the objective in (10) and then learn ALI-CFMs by minimising (4). In Tables 1 and 2 we compare the performance of OT-ALI-CFM to other CFM methods on Embryoid body (EB; Moon et al. [2019]) data, and on Cite-seq (Cite) and Multiome (Multi) data from Lance et al. [2021]. More details are given in Appendix D.2.

4 Discussion

We have proposed ALI-CFM, a novel approach to learning interpolants in multimarginal flow matching. We have demonstrated the capability of our ALIs to fit complex dynamics where existing interpolants fail (Fig. 1). In the 5D scRNA-seq trajectory inference tasks, OT-ALI-CFM mostly performs within the margin of error of OT-MFM (Table 1). However, we believe that regularising our interpolants to stay close to $\ell(x_0, x_1, t)$ is overly restrictive, causing our interpolants to miss the data manifold – especially in higher dimensions, which explains the results in Table 2. In our future work, we plan to explore alternative regularisation objectives and apply OT-ALI-CFM to real datasets akin to the data in Fig. 1, e.g., using cell-tracking or LiDAR data.

[1]The objective requires differentiating the learnt interpolants with respect to t, which is done with little overhead using autograd. Note that the interpolants' parameters ϕ are fixed at this stage.

References

Michael S Albergo, Nicholas M Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint arXiv:2303.08797*, 2023.

Charlotte Bunne, Stefan G Stark, Gabriele Gut, Jacobo Sarabia Del Castillo, Mitch Levesque, Kjong-Van Lehmann, Lucas Pelkmans, Andreas Krause, and Gunnar Rätsch. Learning single-cell perturbation responses using neural optimal transport. *Nature methods*, 20(11):1759–1768, 2023.

Michael AA Cox and Trevor F Cox. Multidimensional scaling. In *Handbook of data visualization*, pages 315–347. Springer, 2008.

Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

James A Hay, Lee Kennedy-Shaffer, Sanjat Kanjilal, Niall J Lennon, Stacey B Gabriel, Marc Lipsitch, and Michael J Mina. Estimating epidemiologic dynamics from cross-sectional viral load distributions. *Science*, 373(6552):eabh0635, 2021.

Nick Huang, Aaron Gokaslan, Volodymyr Kuleshov, and James Tompkin. The GAN is dead; long live the GAN! a modern GAN baseline. *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Kacper Kapuśniak, Peter Potaptchik, Teodora Reu, Leo Zhang, Alexander Tong, Michael Bronstein, Avishek Joey Bose, and Francesco Di Giovanni. Metric flow matching for smooth interpolations on the data manifold. *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Allon M Klein, Linas Mazutis, Ilke Akartuna, Naren Tallapragada, Adrian Veres, Victor Li, Leonid Peshkin, David A Weitz, and Marc W Kirschner. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, 161(5):1187–1201, 2015.

Christopher Lance, Malte D Luecken, Daniel B Burkhardt, Robrecht Cannoodt, Pia Rautenstrauch, Anna Laddach, Aidyn Ubingazhibov, Zhi-Jie Cao, Kaiwen Deng, Sumeer Khan, et al. Multimodal single cell data integration challenge: results and lessons learned. *Advances in Neural Information Processing Systems (NeurIPS) Competitions and Demonstrations*, 2021.

Justin Lee, Behnaz Moradijamei, and Heman Shakeri. Multi-marginal stochastic flow matching for high-dimensional snapshot data at irregular time points. *International Conference on Machine Learning (ICML)*, 2025.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *International Conference on Learning Representations (ICLR)*, 2023.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *International Conference on Learning Representations (ICLR)*, 2023.

Evan Z Macosko, Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, Allison R Bialas, Nolan Kamitaki, Emily M Martersteck, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214, 2015.

Kevin R Moon, David Van Dijk, Zheng Wang, Scott Gigante, Daniel B Burkhardt, William S Chen, Kristina Yim, Antonia van den Elzen, Matthew J Hirn, Ronald R Coifman, et al. Visualizing structure and transitions in high-dimensional biological data. *Nature biotechnology*, 37(12): 1482–1492, 2019.

Kirill Neklyudov, Rob Brekelmans, Alexander Tong, Lazar Atanackovic, Qiang Liu, and Alireza Makhzani. A computational framework for solving Wasserstein lagrangian flows. *International Conference on Machine Learning (ICML)*, 2024.

Jim Oeppen and James W Vaupel. Broken limits to life expectancy, 2002.

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky TQ Chen. Multisample flow matching: Straightening flows with minibatch couplings. *International Conference on Machine Learning (ICML)*, 2023.

Martin Rohbeck, Edward De Brouwer, Charlotte Bunne, Jan-Christian Huetter, Anne Biton, Kelvin Y Chen, Aviv Regev, and Romain Lopez. Modeling complex system dynamics with flow matching across time and conditions. *International Conference on Learning Representations (ICLR)*, 2025.

Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.

Alexander Tong, Jessie Huang, Guy Wolf, David Van Dijk, and Smita Krishnaswamy. TrajectoryNet: A dynamic optimal transport network for modeling cellular dynamics. *International Conference on Machine Learning (ICML)*, 2020.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024.

Conrad H Waddington. The epigenotype. *Endeavour*, 1:18–20, 1942.

Eric Xing, Michael Jordan, Stuart J Russell, and Andrew Ng. Distance metric learning with application to clustering with side-information. *Advances in Neural Information Processing Systems (NeurIPS)*, 2002.

A Related works

Closest to our approach are [Kapuśniak et al., 2024, Neklyudov et al., 2024], who also learn nonlinear interpolants parameterised by a time-dependent neural network. However, there are differences in both the approach and the underlying motivations. Firstly, none of these approaches explicitly approximate a divergence between the intermediate time marginals and the interpolants, and they do not use adversarial training. Kapuśniak et al. [2024] motivates their approach by addressing distributions that are supported on a manifold with a metric learning approach Cox and Cox [2008], Xing et al. [2002]. However, their metric is time-independent, which contrasts with our generator and discriminator networks that are both time-dependent, and we do not specify an underlying metric.

Moreover, Rohbeck et al. [2025] explicitly addresses the multi-marginal problem; however, they, along with the other piecewise interpolants Tong et al. [2024], Neklyudov et al. [2024], Kapuśniak et al. [2024], impose constraints on the interpolants to pass through the marginal samples point-wise, diverging from our approach. Furthermore, the smooth spline-based interpolations used in Rohbeck et al. [2025], Lee et al. [2025] may not scale well in high dimensions Lee et al. [2025].

B Proof of uniqueness theorem

For completeness, we restate the theorem here.

Theorem 2.1. *Fix $t \in (0,1)$, q_t, and a coupling π between q_0 and q_1 such that the distribution $\ell(\cdot,\cdot,t)_{\#}\pi$ is absolutely continuous (a.c.) w.r.t. the Lebesgue measure. Then the interpolant $G(\cdot,\cdot,t) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ minimising*

$$\mathbb{E}_{(x_0,x_1)\sim\pi}\|G(x_0,x_1,t) - \ell(x_0,x_1,t)\|^2 \tag{9}$$

subject to $G(\cdot,\cdot,t)_{\#}\pi = q_t$, exists and is unique on the support of π up to almost-everywhere equality.

Proof of Theorem 2.1. Let $G(\cdot,\cdot,t)$ be any function satisfying the constraint. Consider the following joint distribution on $\mathcal{X} \times \mathcal{X}$,

$$\Pi(dx_t, dx_t') = (\ell(\cdot,\cdot,t), G(\cdot,\cdot,t))_{\#}\pi.$$

The marginals of Π over the first and second components are the a.c. distributions $\ell(\cdot,\cdot,t)_{\#}\pi$ and q_t, respectively. The distribution Π is a Kantorovich plan between these two marginals whose cost is given by (9). This cost has a unique minimiser over all transport maps Π (up to a.e. equality on the

supports) because the first marginal is a.c., and the minimiser has Π deterministic over the second component given the first, i.e., Π is given by $(\mathrm{Id}, T)_{\#}\ell(\cdot, \cdot, t)$ for some function $T : \mathcal{X} \to \mathcal{X}$.

This minimum is indeed uniquely achieved by $G = T \circ \ell - \ell$. Conversely, the optimal transport plan from $\ell(\cdot, \cdot, t)_{\#}\pi$ to q_t yields an interpolant G in the obvious manner, showing existence. □

C The knot distribution

Before normalising t to be of unit length, we let $t \in [0, 3]$, assert that the number of marginals, K, is a multiple of three, and set

$$\tilde{t} = 3\frac{t}{t_{\text{max}}} - 1.5. \tag{11}$$

We then partition $\tilde{t}$ into three equally sized segments $\tilde{t}_1 = \tilde{t}_{1:K/3}$, $\tilde{t}_2 = \tilde{t}_{K/3+1:2K/3}$ and $\tilde{t}_3 = \tilde{t}_{2K/3+1:K}$. For each time stamp, there are two random variables

$$X(t) \sim q_t(\mu_X(t), \sigma^2) \tag{12}$$

and

$$Y(t) \sim q_t(\mu_Y(t), \sigma^2) \tag{13}$$

where $\sigma = 0.1$ and

$$\mu_X(t) = \begin{cases} 3(\tilde{t} + 0.5), & \forall \tilde{t}(t) \in \tilde{t}_1 \\ \cos(2\pi(\tilde{t} - 0.75) & \forall \tilde{t}(t) \in \tilde{t}_2 \\ 3(\tilde{t} - 0.5), & \forall \tilde{t}(t) \in \tilde{t}_3 \end{cases} \tag{14}$$

and

$$\mu_Y(t) = \begin{cases} -0.5\tanh(5(\tilde{t} + 1)) + 0.5, & \forall \tilde{t}(t) \in \tilde{t}_1 \\ \sin(2\pi(\tilde{t} - 0.75)), & \forall \tilde{t}(t) \in \tilde{t}_2 \\ 0.5\tanh(5(\tilde{t} - 1)) + 0.5, & \forall \tilde{t}(t) \in \tilde{t}_3 \end{cases}. \tag{15}$$

Finally, we divide t by three and collect ten samples from the bivariate distribution of $(X(t), Y(t))$ at each of the K time stamps.

D Additional experiment details

D.1 Knot experiment

In this section, we detail the networks' hyperparameters for the 'knot' experiment presented in Fig. 1.

The trainable vector field is parameterised with a two-layer MLP with 32 hidden units in every experiment, and the Adam optimiser is used for training. All methods are trained for 30k epochs with a learning rate of 10^{-4}.

ALI-CFM interpolant is trained using two-layer MLPs with 128 hidden units and ELU activation function for both f_ϕ and D_γ. We use $\lambda = 1$ in (10), and train the interpolant for 50k epochs with a batch size of 128. The nets f_ϕ and D_γ were optimised using separate Adam optimisers, with a learning rate of 0.001.

For OT-MFM, we follow the setup used in their LiDAR experiment Kapuśniak et al. [2024], where the metric is inferred from all the data, while the interpolants are learnt using only samples from q_0 and q_1.

D.2 scRNA-seq experiments

The datasets had already mostly been preprocessed, and we downloaded the Cite and Multi data at `https://data.mendeley.com/datasets/hhny5ff7yj/1`, and the EB data in the repository associated with Tong et al. [2020], specifically here: `https://github.com/KrishnaswamyLab/TrajectoryNet/blob/master/data/eb_velocity_v5.npz`. Following Tong et al. [2024], we additionally whitened the data in the 5D experiments.

Hyperparameters for ALI-CFM Regarding neural network parameterisations, we let f_ϕ and D_γ be two-layered MLPs with 64 (for the 5D experiments) or 1024 (for the 50 or 100D experiments) hidden units in each layer with ELU activations. In accordance with Kapuśniak et al. [2024], we used

a three-layered MLP with SELU activations to model u_t^θ. The number of hidden units in the layers of u_t^θ was 64 or 1024, depending on the data dimensions, as for f_ϕ and D_γ above.

Beyond hyperparameters, we found that normalising the data and adding small noise ($0.01 \cdot \epsilon$, $\epsilon \sim \mathcal{N}(0, 1)$) on the time-input to G_ϕ both helped when training the GAN. Importantly, these techniques were only applied when training the ALIs, not ALI-CFM. Instead, when learning the vector field using the ALIs, we denormalised the interpolants and the target vector field.

At inference time, we pushed all the samples from marginal $i - 1$ to the time associated with the held-out marginal t_i, as in Tong et al. [2024].

Reproduced results The reported scores I/OT-MFM and I/OT-CFM were obtained by running the code provided by Kapuśniak et al. [2024] without any changes made to the code.

The CFMs trained with the cubic splines interpolants, i.e. I/OT-MMFM, had not been previously applied to these datasets. We implemented the cubic splines interpolants without the Gaussian probability path to align with the other considered methods. We only ran I/OT-MMFM on the 5D experiments.