FINSLER FLOW MATCHING: DYNAMICS-AWARE GEODESIC INTERPOLATION FOR SINGLE-SNAPSHOT TRAJECTORY INFERENCE

Anonymous authors
Paper under double-blind review

ABSTRACT

Single-cell snapshot data can resolve a continuum of cellular states but do not uniquely determine the dynamics governing transitions between them. However, additional dynamical information can often be encoded in a cell-cell Markov transition kernel. Existing generative approaches for single cell trajectory inference either infer transport only from population marginals, impose a symmetric geometry on the state space, or incorporate directionality through a single velocity vector at each observed state. We introduce Finsler Flow Matching (FFM), a framework for learning continuous stochastic dynamics from discrete Markov transition graphs. We use the first and second local moments to construct a Finsler structure motivated by the Freidlin–Wentzell action, where the second moment determines anisotropic accessibility and the first moment introduces a preferred direction of motion. We learn neural approximations of the resulting directed geodesics, use their Finsler cost to construct source-target couplings, and define geometry-aware stochastic conditional paths that can be distilled into a continuous generative process through simulation-free score and flow matching. Across synthetic and single-cell trajectory inference benchmarks, FFM improves recovery of withheld intermediate populations, particularly when the transition dynamics are strongly directional or anisotropic. Our results provide a principled route from discrete transition probabilities to continuous generative dynamics while retaining both directional and diffusive structure.

1 INTRODUCTION

Understanding how cells change their state in response to intrinsic programs and extrinsic perturbations is a central challenge in biology. Such transitions underlie development and differentiation, but also cellular reprogramming, disease progression, and responses to therapeutic intervention. Single-cell genomics methods have made it possible to probe these processes at high resolution by resolving a continuum of cellular states in a single cross-sectional experiment (Trapnell et al., 2014). However, because single-cell sequencing methods are destructive, individual cells cannot be followed through time and transitions between observed states are not directly measured. As a result, the continuous dynamics underlying cell state changes can generally not be determined from these measurements alone (Weinreb et al., 2018).

Several experimental and computational strategies can provide additional information about developmental directionality and dynamics. RNA velocity infers the direction of local transcriptional change (La Manno et al., 2018; Bergen et al., 2020), lineage tracing tracks ancestry (Weinreb et al., 2020; Wagner & Klein, 2020; McKenna et al., 2016) and metabolic or fluorescent labelling can introduce an absolute timescale (Battich et al., 2020; Cao et al., 2020; Gehart et al., 2019). Measurements collected across multiple experimental time points also carry temporal information, although the correspondence between cells at successive time points is lost upon sampling and must itself be inferred (Schiebinger et al., 2019). Despite their different origins, these sources of dynamical information can often be translated into a common representation, a cell-cell Markov transition kernel, whose entries quantify the likelihood of transitions between observed cellular states (Lange et al., 2022; Weiler et al., 2024). Such kernels have proven highly effective for inferring terminal states, lineage relationships, and fate probabilities (Fleck et al., 2023; Lange et al., 2024). However, the

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

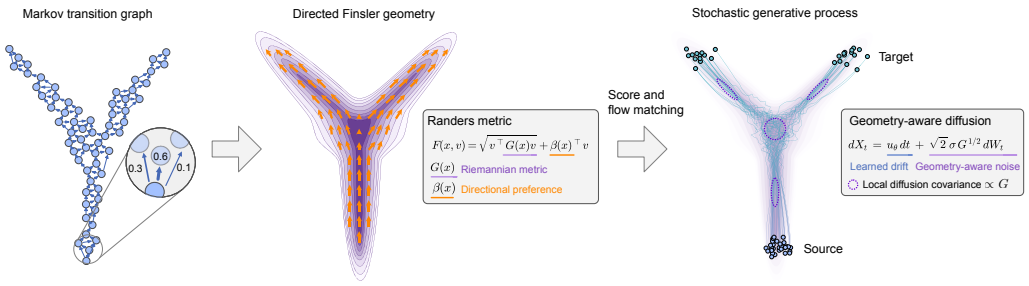


Figure 1: **Overview of Finsler Flow Matching (FFM).** A discrete Markov transition graph is converted into a continuous directed Finsler geometry, which guides score and flow matching to define a stochastic generative model with geometry-aware diffusion.

resulting dynamics remain defined on the finite graph of observed cells. A Markov kernel encodes transition preferences between observed states, but this does not specify how the system evolves continuously through state space, particularly between or beyond the observed cells.

Generative modeling of population dynamics provides a complementary approach (Neklyudov et al., 2023). Given samples at multiple time points, optimal transport formulations recover a coupling between the observed marginals (Schiebinger et al., 2019; Cuturi, 2013), while flow and Schrödinger Bridge matching provide simulation-free frameworks for learning continuous dynamics (Tong et al., 2024b; Lipman et al., 2023; Rohbeck et al., 2025; Tong et al., 2024a). Some variants additionally incorporate the geometry of the state or distribution space (Chen & Lipman, 2023; Haviv et al., 2025). Metric Flow Matching replaces Euclidean interpolation with geodesics under a learned Riemannian metric (Kapusniak et al., 2024), capturing the anisotropic geometry of the data manifold but not a preferred direction of motion. Curly Flow Matching incorporates RNA velocity through a nonzero-drift reference process and can therefore capture directed and non-gradient dynamics (Petrović et al., 2025), but takes as input a single drift vector at each state, and therefore cannot exploit the second-order structure of a local transition distribution.

In this paper, we use a discrete Markov transition graph to define a continuous dynamical prior for endpoint-constrained generative transport. We introduce Finsler Flow Matching (FFM), which converts the directional and diffusive structure of a Markov transition graph into a smooth Finsler geometry and uses this geometry to learn continuous stochastic transport through score and flow matching. Under the additional assumption that the Markov chain discretizes an Itô diffusion, the graph-derived local moments recover its drift and diffusion. Freidlin–Wentzell (Freidlin & Wentzell, 2012; Heymann & Vanden-Eijnden, 2008) theory then motivates a Randers metric (Bao et al., 2000), in which local diffusion determines anisotropic accessibility and drift introduces directional preference. We then learn a neural path interpolant to approximate directed Randers geodesics and use the resulting path cost to construct source–target couplings consistent with the transition dynamics. Finally, we use these couplings and geodesic interpolants to construct tractable conditional probability paths and learn an endpoint-constrained stochastic generative process through simulation-free score and flow matching (Tong et al., 2024b), whose local noise geometry is shaped by the graph-derived second moment.

To summarize, our main contributions are:

- **A principled construction of Finsler geometry from Markov graphs.** Under a short-time Itô diffusion model, we show that the first and second local moments of a cell-cell Markov kernel recover the drift and diffusion of an underlying stochastic process. We use these estimates to derive a Randers geometry from the Freidlin-Wentzell minimum-action principle (Freidlin & Wentzell, 2012), jointly encoding anisotropic accessibility and directional dynamics.
- **A framework for learning continuous stochastic dynamics from discrete Markov graphs.** We introduce Finsler Flow Matching, which learns a neural path interpolant under the induced Finsler geometry, uses directed geodesic costs to construct source-target couplings, and defines stochastic conditional paths whose local noise structure is shaped by the

graph-derived diffusion. These paths are then distilled into a continuous generative process through simulation-free score and flow matching.

- **Flexible inference of single-cell trajectory dynamics.** By using cell-cell Markov graphs as a common representation of transition dynamics, FFM can incorporate information derived from various experimental and computational sources, including RNA velocity, lineage tracing, labeling experiments, or temporally sampled populations within a unified generative framework. We show that our model outperforms previous methods in intermediate state prediction on various benchmarks.

2 PRELIMINARIES

In this section, we establish the background on Finsler geometry, optimal transport, and continuous generative modeling via flow and score matching. More information can be found in A

Finsler Geometry: Riemannian geometry measures local distance via a symmetric metric tensor $G(x)$. Finsler geometry generalizes this to a direction-dependent, asymmetric cost. We introduce a special class of Finsler structures called a *Randers metric*:

$$F(x, v) = \sqrt{v^\top G(x) v} + \beta(x)^\top v,$$

where $G(x) \in \mathrm{SPD}(d)$ is a Riemannian metric tensor and $\beta(x)$ is a 1-form giving directionality. For an absolutely continuous path $\gamma : [0, 1] \to \mathbb{R}^d$, the Finsler length $L_F(\gamma)$ and energy $E_F(\gamma)$ functionals are defined as:

$$L_F(\gamma) \coloneqq \int_0^1 F(\gamma(t), \dot\gamma(t))\mathrm{d}t, \qquad E_F(\gamma) \coloneqq \int_0^1 F(\gamma(t), \dot\gamma(t))^2\mathrm{d}t.$$

Curves minimizing L_F are called *(Finsler) geodesics* and are locally shortest paths connecting two points. The *Finsler geodesic distance* $c_F(x_0, x_1)$ is defined as:

$$c_F(x_0, x_1) \coloneqq \inf_{\gamma(0)=x_0\ \gamma(1)=x_1} \int_0^1 F(\gamma(t), \dot\gamma(t))\mathrm{d}t. \tag{1}$$

Optimal Transport and Schrödinger Bridge Problem: Optimal Transport (OT) aims to transport a probability distribution into another one in the most cost efficient way (Villani, 2009). Given source and target densities $p_0, p_1 \in \mathcal{P}(\mathbb{R}^d)$ and a cost function $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, the *Kantorovich problem* seeks an optimal transport plan $\pi \in \Pi(p_0, p_1)$ minimizing total cost:

$$\pi^\star = \arg\inf_{\pi\in\Pi(p_0,p_1)} \int_{\mathbb{R}^d\times\mathbb{R}^d} c(x_0, x_1)\mathrm{d}\pi(x_0, x_1),$$

where $\Pi(p_0, p_1)$ is the set of joint distributions with marginals p_0 and p_1. Using the squared Finsler geodesic distance as cost function and the elementary fact $c_F(x_0, x_1)^2 = \inf_{\gamma:\gamma(0)=x_0,\gamma(1)=x1} E_F(\gamma)$ [1] implies that

$$W^2_{2,F}(p_0, p_1) = \inf_{\pi\in\Pi(p_0,p_1)} \inf_{\gamma:\gamma(0)=x_0,\gamma(1)=x_1} \int_{\mathbb{R}^d\times\mathbb{R}^d} \int_0^1 F^2(\gamma(t), \dot\gamma(t))\ \mathrm{d}t\ \mathrm{d}\pi(x_0, x_1). \tag{2}$$

A stochastic counterpart to the OT problem is given by the Schrödinger Bridge Problem (SBP) (Léonard, 2013). Given a reference measure on $\Omega = C([0, 1], \mathbb{R}^d)$ of a stochastic process R, the SBP searches for the closest path measure to R that has endpoint marginals given by p_0 and p_1. More precisely

$$Q^* = \operatorname*{arg\,min}_{Q:Q_0=p_0,;Q_1=p_1} \mathrm{KL}(Q\|R). \tag{3}$$

A classical result due to Föllmer (1988) tells us that the solution can be decomposed as

$$Q^* = \int R^{x_0,x_1}\mathrm{d}\pi^*(x_0, x_1), \text{ where } \pi^* = \operatorname*{arg\,min}_{\pi\in\Pi(p_0,p_1)} \mathrm{KL}\left(\pi\|R_{01}\right), \tag{4}$$

[1] It follows from the Cauchy-Schwarz inequality and the fact that minimizers of the Energy are constant speed geodesics.

where $R^{x_0,x_1} = R(\cdot \mid Y_0 = x_0, Y_1 = x_1)$ denotes the reference process conditioned on its endpoints and R_{01} its endpoint law. For a Brownian reference process $dY_t = \sigma_B dW_t$, one can see that $R_t^{x_0,x_1} = \mathcal{N}\left((1-t)x_0 + tx_1, \sigma^2 t(1-t)I\right)$. For non-Brownian reference processes, such as SDEs with anisotropic diffusion and non-trivial drift

$$dY_t = b_0(Y_t), dt + \Sigma(Y_t) dW_t,$$

the conditional bridges are in general neither Gaussian nor available in closed form. In the small noise regime, their most-probable transition paths are instead characterized by minimizers of the Freidlin-Wentzell action, a connection that we use in B.

Score and Flow Matching: Flow Matching (FM) is a simulation-free method for training continuous-time generative models. Given a target probability path $p_t(x)$ pushed forward by an unknown vector field $u_t(x)$, FM fits a neural vector field $v_\theta(t, x)$ via a regression objective:

$$\mathcal{L}_{\mathrm{FM}}(\theta) = \int_0^1 \mathbb{E}_{x \sim p_t(x)} \left[|v_\theta(t,x) - u_t(x)|^2 \right] \mathrm{d}t. \tag{5}$$

Conditional Flow Matching (CFM) circumvents the intractability of the marginal vector field $u_t(x)$ by conditioning on endpoint pairs $(x_0, x_1) \sim q(x_0, x_1)$ (Lipman et al., 2023; Albergo & Vanden-Eijnden, 2023). The network is trained on tractable conditional vector fields $v_t(x \mid x_0, x_1)$ pushing probability paths along conditional trajectories $\mu_t(x_0, x_1)$. FM is often combined with minibatch optimal transport to pair source and target samples, enabling it to approximate dynamic optimal transport (Tong et al., 2024a). Following the approach of Tong et al. (2024b) one may jointly training the flow and score through the combined objective

$$\mathcal{L}_{[\mathrm{SF}]^2\mathrm{M}}(\theta) = \mathcal{L}_{\mathrm{CFM}}(\theta) + \lambda \mathcal{L}_{\mathrm{CSM}}(\theta).$$

The learned flow v_θ and score s_ψ can be used with an arbitrary matrix-valued diffusion schedule $\mathbf{g} : [0,1] \to \mathbb{R}^{d \times d}$ to define the stochastic generative process

$$dX_t = \left[v_\theta(t, X_t) + \frac{1}{2} \mathbf{g}(t) \mathbf{g}(t)^\top s_\psi(t, X_t) \right] dt + \mathbf{g}(t)\, dW_t, \tag{6}$$

which satisfies $X_t \sim p_t$ for all $t \in [0,1]$.

3 Related Work

Single-cell trajectory inference. OT methods infer developmental trajectories from populations observed at multiple time points by coupling successive snapshot distributions (Schiebinger et al., 2019; Tong et al., 2020; Bunne et al., 2022; Zhang et al., 2025). Relatedly, Huguet et al. (2022) combine manifold-aware OT with neural differential equations to learn continuous cellular dynamics from snapshot data. Complementary approaches such as CellRank construct cell-cell Markov transition kernels from RNA velocity, lineage information, and other dynamical measurements to infer fate probabilities and developmental relationships (Lange et al., 2022; Weiler et al., 2024). FFM builds on these representations by using local transition statistics to construct a smooth directed geometry and continuous generative dynamics beyond the observed cells.

Stochastic generative transport. Simulation-free Schrödinger bridge methods combine score and flow matching to learn stochastic dynamics between endpoint distributions (Tong et al., 2024b). Related work studies closed-form Gaussian Schrödinger bridges and smooth Schrödinger bridges for trajectory inference (Bunne et al., 2023; Hong et al., 2025). FFM builds on this framework by using the transition-derived geometry to determine both the interpolating paths and the local anisotropic noise structure.

Geometric and directional flow matching. Kapusniak et al. (2024) replace Euclidean conditional paths with neural approximations of geodesics under a data-dependent Riemannian metric. FFM extends this idea to a directed Randers geometry derived from local transition statistics. Relatedly, Bamberger et al. (2026) uses local Markov kernels to construct anisotropic covariance for geometry-aware flow matching, whereas FFM additionally incorporates first-moment directionality through a Randers geometry and geodesic coupling.Petrović et al. (2025) uses a nonzero-drift reference process to model directed and non-gradient dynamics from local velocity fields. FFM instead derives directionality from the first moment of a transition kernel and uses its second moment to capture anisotropic transition structure.

Lineage-informed Finsler geometry. Zweig et al. (2026) recently use Finsler geometry to incorporate directed lineage-tree priors into trajectory inference, penalizing motion inconsistent with a prescribed adjacency between cell types. FFM instead derives its Finsler geometry directly from cell-cell transition probabilities, using the probabilistic transition graph to encode both directional preference and local transition accessibility rather than relying on a categorical lineage prior.

4 Finsler Flow Matching

Problem formulation. Let $D = \{x_i\}_{i=1}^N \subseteq \mathbb{R}^d$ be a collection of observations sampled from an unknown distribution p. Let $D_0 = \{x_i^0\}_{i=1}^{N_0}$ and $D_1 = \{x_i^1\}_{i=1}^{N_1}$ denote two subpopulations from source and target distributions p_0 and p_1. In addition, we are given a row-stochastic transition matrix $P \in \mathbb{R}^{N\times N}$, whose entries represent transition likelihoods between observed states.

Our goal is to infer a transport mechanism from p_0 to p_1 and a family of intermediate distributions $(p_t)_{t\in[0,1]}$ that are consistent with the dynamics encoded by P. More formally we can define this as an OT problem of the form 2

$$\begin{aligned}(\pi^*, \gamma^*) &= \arg\min_{(\pi,\gamma)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \int_0^1 F_P^2(\gamma(t), \dot\gamma(t))\, \mathrm{d}t\, \mathrm{d}\pi(x_0, x_1) \\ &\text{where } \pi \in \Pi(p_0, p_1) \text{ and } \gamma(0) = x_0, \gamma(1) = x_1\end{aligned} \tag{7}$$

and $F_P(x, v) = \sqrt{v^\top G_P(x) v} + \beta_P(x)^\top v$ is a P-derived Randers metric (Definition 7). The corresponding intermediate distributions are induced by transporting the optimal coupling along the optimal paths.

From now on we consider the special case where the P-derived Randers metric is constructed from spatial moments of P. More precisely we define at each observed state x_i moments as:

$$m(x_i) \coloneqq \sum_{j=1}^N P_{ij}(x_j - x_i), \text{ and } C(x_i) \coloneqq \sum_{j=1}^N P_{ij}(x_j - x_i)(x_j - x_i)^\top. \tag{8}$$

[2]$m(x_i)$ captures the net preferred displacement and therefore provides directionality whereas $C(x_i)$ captures the direction on the data manifold with high transition probabilities. To evaluate the metric along continuous paths, we construct smooth extensions of m and C, for example using kernel smoothing, neural networks, or support vector regression (Smola & Schölkopf, 2004), while preserving positive semidefiniteness of C. We use kernel smoothing for simplicity details are provided in C. To obtain a non-degenerate Finsler structure, we define the regularized Freidlin–Wentzell (FW) action (Freidlin & Wentzell, 2012). Firstly, set $C_\rho(x) \coloneqq C(x) + \rho I$ with $\rho > 0$ and define

$$F_P(x, v) \coloneqq \underbrace{\sqrt{v^\top C_\rho(x)^{-1} v}}_{\text{Anisotropic accessibility}} \underbrace{\sqrt{m(x)^\top C_\rho(x)^{-1} m(x) + \lambda^2}}_{\text{first-moment calibration}} - \underbrace{v^\top C_\rho(x)^{-1} m(x)}_{\text{directional alignment}}, \tag{9}$$

where $\lambda > 0$ ensures Randers admissibility 5. The anisotropic accessibility term encourages geodesics to remain close to regions and directions that are well supported by the transition probabilities, directional alignment ensures that geodesics follow the preferred directionality and the first-moment calibration ensures that both penalty terms are on the same scale. The corresponding action is

$$S(\gamma) = \int_0^1 F_P(\gamma(t), \dot\gamma(t))\, \mathrm{d}t. \tag{10}$$

The following proposition formalizes how geodesics under F_P align with the transition structure encoded by P:

[2]At finite graph scale $C = \mathrm{Cov}_P(\Delta x) + mm^\top$, we interpret C as an anisotropic transition-accessibility rather then diffusion covariance.

Algorithm 1 Training the Finsler Geodesic Model

Require: Source and target distributions p_0, p_1, Finsler metric F_P, interpolant network φ_η, LR γ.
1: **while** not converged **do**
2: Sample $x_0 \sim p_0$, $x_1 \sim p_1$, $t \sim \mathcal{U}[0,1]$ and compute $x_{t,\eta}(x_0, x_1)$ and $\dot{x}_{t,\eta}(x_0, x_1)$
3: $\mathcal{L}_{\text{geo}} \leftarrow F_P(x_{t,\eta}, \dot{x}_{t,\eta})^2$
4: $\eta \leftarrow \eta - \gamma \nabla_\eta \mathcal{L}_{\text{geo}}$
5: **end while**

Proposition 1 (Informal 8)**.** *Suppose the endpoints can be connected along a transition-supported corridor at a Finsler cost lower than that required to leave and return to a prescribed surrounding neighborhood. Then every minimizing F_P-geodesic remains within that neighborhood. Moreover, locally the metric favors alignment with the first transition moment.*

In the special case where P is the discrete observation of an SDE 14, we can conclude even more about geodesics. By Freidlin & Wentzell (2012) and Heymann & Vanden-Eijnden (2008) we can show that geodesics under the resulting metric F_P follow the maximum likelihood paths of the stochastic process.

Theorem 2 (Informal 18)**.** *If P is a discretization of a SDE, then minimizers of the empirical FW Finsler action converge to the asymptotically most probable transition path in the small-noise limit.*

FFM is defined for arbitrary row-stochastic transition matrices, whereas its diffusion interpretation requires Assumption 11. Connections to the limits of local graph kernels, and the scope of these results for velocity-based constructions, are discussed in B.

Learning Geodesics: For endpoints $(x_0, x_1) \sim p_0 \otimes p_1$ and $t \in [0,1]$, parameterize a candidate curve by

$$x_{t,\eta}(x_0, x_1) \;=\; (1-t)\, x_0 + t\, x_1 + t(1-t)\, \varphi_\eta(t, x_0, x_1), \tag{11}$$

where $\varphi_\eta : [0,1] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is some neural network with parameters η. The boundary correction $t(1-t)$ enforces $x_{0,\eta} = x_0$ and $x_{1,\eta} = x_1$. The time derivative is given by,

$$\dot{x}_{t,\eta} = x_1 - x_0 + (1-2t)\varphi_\eta(t, x_0, x_1) + t(1-t)\partial_t \varphi_\eta(t, x_0, x_1).$$

This interpolant was previously proposed in Kapusniak et al. (2024), and similarly to their approach, we minimize the Finsler Energy functional,

$$\mathcal{L}_{\text{geo}}(\eta) = \mathbb{E}_{\substack{(x_0,x_1)\sim p_0\otimes p_1 \\ t\sim\mathcal{U}[0,1]}} \left[F_P(x_{t,\eta}(x_0,x_1), \dot{x}_{t,\eta}(x_0,x_1))^2 \right]. \tag{12}$$

Since energy minimizers are constant-speed reparameterizations of geodesics, minimizing (12) over η yields an approximation of the pairwise P-derived Finsler geodesics. We denote the resulting parameters by η^*.

Finsler Optimal Transport: Having learned an approximation of the geodesic connecting arbitrary endpoint pairs, we can approximate the inner minimization in the transport problem 7. For each (x_0, x_1), define the learned Finsler transport cost

$$c_{\eta^*}(x_0, x_1) := \int_0^1 F_P(x_{t,\eta^*}(x_0,x_1), \dot{x}_{t,\eta^*}(x_0,x_1))^2 \, \mathrm{d}t.$$

Thus,

$$c_{\eta^*}(x_0, x_1) \approx \inf_{\substack{\gamma(0)=x_0 \\ \gamma(1)=x_1}} \int_0^1 F_P(\gamma(t), \dot{\gamma}(t))^2 \, \mathrm{d}t = c_F^2(x_0, x_1).$$

We can solve the corresponding static optimal transport problem using costs. In particular, we consider its entropy-regularized form

$$\pi_\varepsilon^* = \arg\min_{\pi \in \Pi(\hat{p}_0, \hat{p}_1)} \left\{ \sum_{i=1}^m \sum_{j=1}^m \pi_{ij} c_{ij} + \varepsilon \, \mathrm{KL}(\pi \,\|\, \hat{p}_0 \otimes \hat{p}_1) \right\}, \tag{13}$$

where $\varepsilon > 0$ controls the strength of the entropic regularization. This problem can be solved efficiently using Sinkhorn iterations (Cuturi, 2013). The interpolant x_{t,η^*} and the coupling π_ε^* provide an approximation of the original Finsler OT problem 7 up to the approximation errors.

Flow and Stochastic Model: Assume we have computed $(\pi_\varepsilon^*, x_{\cdot,\eta^*})$. Drawing $(x_0, x_1) \sim \pi_\varepsilon^*$ and transporting each pair along its learned geodesic gives the probability path $p_t(x) \coloneqq (x_{t,\eta^*})_{\#}\, \pi_\varepsilon^*(x)$. By construction, p_0 and p_1 are the endpoint marginals. The corresponding conditional probability distribution path is $p_t(x|x_0, x_1) = \delta_{x_{t,\eta(x_0,x_1)}}$. Furthermore, define the conditional path $\mu_t(x_0, x_1) := x_{t,\eta^*}(x_0, x_1)$ We then train a model v_θ using the conditional flow matching objective

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{\substack{(x_0,x_1)\sim\pi_\varepsilon^\star \\ t\sim\mathcal{U}[0,1]}} \left[\|v_\theta(t, \mu_t(x_0, x_1)) - \dot{\mu}_t(x_0, x_1)\|_2^2 \right].$$

Motivated by Föllmer (1988) we approximate each conditional bridge with an anisotropic Gaussian path. For $(x_0, x_1) \sim \pi_\varepsilon^\star$, define $C_{\rho,t} := C_\rho(\mu_t)$ and $p_t(x \mid x_0, x_1) = \mathcal{N}\big(x;\, \mu_t,\, \sigma^2 t(1-t) C_{\rho,t}\big)$. We evaluate C at the deterministic interpolant μ_t rather than at the state x, so that the conditional path can be sampled in closed form. The interpretation of this path is that noise is injected along accessible directions according to P. The conditional probability-flow vector field is

$$u_t(x \mid x_0, x_1) = \dot{\mu}_t + \frac{1-2t}{2t(1-t)}(x - \mu_t) + \frac{1}{2}\dot{C}_{\rho,t} C_{\rho,t}^{-1}(x - \mu_t),$$

See 21 for a derivation. We distill the conditional bridges into a marginal vector field v_θ and score s_ψ using simulation-free conditional flow and score matching (Tong et al., 2024b),

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{\substack{(x_0,x_1)\sim\pi_\varepsilon^\star,\ t\sim\mathcal{U}(0,1) \\ x\sim p_t(\cdot|x_0,x_1)}} \left[\|\tilde{v}_\theta(t, x) - u_t(x \mid x_0, x_1)\|_2^2 \right],$$

$$\mathcal{L}_{\mathrm{CSM}}(\psi) = \mathbb{E}_{\substack{(x_0,x_1)\sim\pi_\varepsilon^\star,\ t\sim\mathcal{U}(0,1) \\ x\sim p_t(\cdot|x_0,x_1)}} \left[\lambda(t)^2 \, \|s_\psi(t, x) - s_t(x \mid x_0, x_1)\|_2^2 \right]$$

where the final loss is given by $\mathcal{L}_{[\mathrm{SF}]^2\mathrm{M}} = \mathcal{L}_{\mathrm{CFM}} + \mathcal{L}_{\mathrm{CSM}}$ and the conditional score is

$$s_t(x \mid x_0, x_1) = \nabla_x \log p_t(x \mid x_0, x_1) = -\frac{1}{\sigma^2 t(1-t)} C_{\rho,t}^{-1}(x - \mu_t).$$

Inference: We sample $x_0 \sim p_0$ and integrate the learned dynamics from $t = 0$ to $t = 1$. In the deterministic setting, we integrate the ODE

$$dX_t = v_\theta(t, X_t)\, dt.$$

In the stochastic setting, we use the score-and-flow SDE introduced in 6 and make a slight generalization by setting $\mathbf{g}(t, x) = g(t)\, C_\rho^{1/2}(x)$, for some scalar-valued noise function $g(t)$. Substituting this choice into the stochastic dynamics gives

$$dX_t = \left[\tilde{v}_\theta(t, X_t) + \frac{1}{2} g(t)^2\, C_\rho(x) s_\psi(t, X_t) + \frac{1}{2} g(t)^2 \,\mathrm{div}(C_\rho(x)) \right] dt + g(t)\, C_\rho(x)^{1/2}\, dW_t,$$

where $\mathrm{div}(C_\rho(x))_i = \sum_j^d \partial_j C_{ij}(x)$ and which we integrate from $X_0 \sim p_0$ to $t = 1$. The scalar schedule $g(t)$ controls the overall noise strength, whereas $C_\rho(x)$ determines its local anisotropic geometry.

5 Experiments

Experimental Setup: The transition graph and derived geometric quantities are recomputed after removing the evaluation cells. Our primary metric is Euclidean W_2 at $t = 1/2$, supplemented by endpoint W_2 and the geometry-dependent $W_{2,F}$. Because $t = 1/2$ is a benchmark convention rather than an established correspondence with biological midpoint time, on Pancreas we additionally report $\overline{W}_2 := \frac{1}{|\mathcal{T}|} \sum_{t\in\mathcal{T}} W_2(p_t, p_{\mathrm{withheld}})$. We compare against OT-CFM, MFM/LAND, Curly-FM under matched architectures and training budgets. More details C.

Method	mid $W_2 \downarrow$	end $W_2 \downarrow$	mid $W_{2,F} \downarrow$
OT-CFM	0.7112 ± 0.0233	0.1734 ± 0.0216	1.0305 ± 0.0263
MFM (LAND)	0.6843 ± 0.0074	0.2357 ± 0.0258	0.8645 ± 0.0961
Curly-FM	0.5221 ± 0.0016	$\mathbf{0.1405 \pm 0.0057}$	0.5236 ± 0.1657
FFM (ours)	0.1588 ± 0.0096	0.2343 ± 0.0292	0.2892 ± 0.0090
FFM (ours), with noise	$\mathbf{0.1449 \pm 0.0116}$	0.2636 ± 0.0682	$\mathbf{0.2742 \pm 0.0314}$
sampling floor	*0.0646*	*0.1896*	*0.2812*

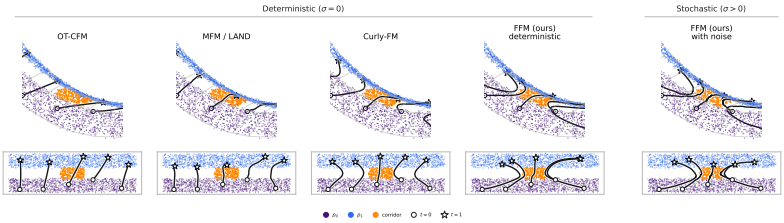


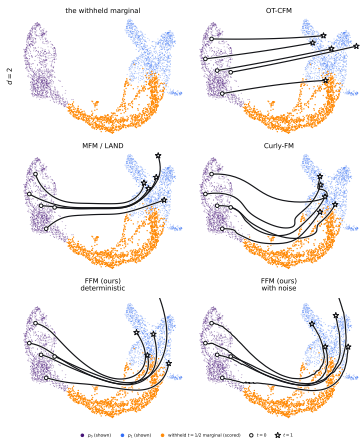
Figure 2: $W_{2,F}$ denotes the Wasserstein distance induced by the Finsler cost c_F. Sampling floor is computed between disjoint halves of the corresponding reference cloud.

FFM recovers directed and anisotropic transition structure: To test whether FFM can recover intermediate dynamics when both directionality matters and when the data satisfy the manifold hypothesis, we construct the synthetic Sheet benchmark, a saddle-shaped manifold. We score against a narrow corridor, which represents a subset of the manifold corresponding to high probability transition regions. We define a directed transition kernel that drives motion towards that middle corridor before crossing and away from it afterward. Curly-FM receives the first transition moment as its local velocity field, whereas FFM additionally uses the second moment to capture anisotropic accessibility. FFM substantially improves midpoint recovery, achieving $W_2 = 0.1588$ and 0.1449 for deterministic and stochastic FFM, respectively, compared with 0.5221 for Curly-FM 2. We further performed and ablation of the Finsler metric, which demonstrates that both the directional and anisotropic components contribute to this improvement 4b. On a complementary synthetic bifurcation benchmark, FFM also preserves cell-level fate assignments across both branches 8.

FFM improves recovery of pancreatic differentiation in high-dimensional representations: To test whether FFM's graph-derived geometry improves trajectory reconstruction on real RNA-velocity data, we evaluate FFM on 3,696 pancreatic cells (Bastidas-Ponce et al., 2019) partitioned into latent-time tertiles and withhold the intermediate population, which is enriched for Ngn3-high pre-endocrine cells. FFM receives an RNA-velocity-derived transition kernel computed using Lange et al. (2022), whereas Curly-FM receives per-cell velocities. We found that relative performance depended strongly on representation dimensionality. In two dimensions, Curly-FM achieved lower midpoint W_2 than FFM, whereas FFM achieved substantially lower midpoint W_2 in the 20 and 50dimensional PCA 3. endpoint rankings vary across methods 3

FFM integrates diverse sources of dynamical information through a common transition representation: To demonstrate that FFM can integrate heterogeneous dynamical information through a common transition representation, we evaluate 2,901 lineage-recorded cells from the iTracer hindbrain organoid dataset (He et al., 2022), withholding the middle diffusion-pseudotime tertile. Lineage information is encoded by barcode families and finer-grained scar families. We construct transition kernels from expression alone and augment them with RNA velocity, lineage similarity derived from barcode and scar families, or both. Each FFM variant achieves lower mean midpoint W_2 than OT-CFM and MFM across the evaluated representations 1. We further found that the optimal kernel depends on the representation dimensionality. Velocity performs best in two dimensions, lineage plus velocity in 20 dimensions, and expression alone in 50 dimensions. These results show that FFM can incorporate heterogeneous sources of transition information, while the utility of each modality depends on how well it informs the transition kernel.

Additional analyses: Appendix C reports runtime and parameter-sensitivity analyses, additional results on erythroid differentiation (Pijuan-Sala et al., 2019) as used in used in Petrović et al. (2025), and further details of the benchmarks above.



Method	$W_2(\frac{1}{2}) \downarrow$	$\overline{W}_2^{[1/3,2/3]} \downarrow$	$W_2(1) \downarrow$	$W_{2,F}(\frac{1}{2}) \downarrow$
$d = 2$				
OT-CFM	1.8062 ± 0.0197	1.7869 ± 0.0230	**0.1740 ± 0.0338**	0.6420 ± 0.0059
MFM / LAND	1.4682 ± 0.0128	1.4642 ± 0.0131	0.3769 ± 0.0346	**0.5399 ± 0.0024**
Curly-FM	**0.9773 ± 0.1311**	**0.9686 ± 0.1725**	0.2581 ± 0.0371	0.7493 ± 0.0258
FFM (ours), deterministic	1.6692 ± 0.0154	1.6175 ± 0.0348	0.3283 ± 0.1293	0.5681 ± 0.0100
FFM (ours), with noise	1.6957 ± 0.0305	1.6434 ± 0.0555	0.2272 ± 0.0239	0.5730 ± 0.0082
sampling floor	*0.1113*	*0.1113*	*0.1022*	*0.4162*
$d = 20$				
OT-CFM	12.3484 ± 0.0485	12.4957 ± 0.0681	5.2262 ± 0.1583	10.9277 ± 0.0412
MFM / LAND	12.0705 ± 0.0095	12.1656 ± 0.0522	7.8205 ± 0.7863	10.4768 ± 0.0637
Curly-FM	12.2962 ± 0.0414	12.3920 ± 0.0687	**4.8562 ± 0.0576**	11.0235 ± 0.0520
FFM (ours), deterministic	10.4319 ± 0.0896	10.7413 ± 0.1927	5.3793 ± 0.1848	**8.9033 ± 0.0400**
FFM (ours), with noise	**10.4036 ± 0.0279**	**10.7355 ± 0.0560**	5.0605 ± 0.1654	8.9213 ± 0.0453
sampling floor	*4.1781*	*4.1781*	*4.4483*	*4.3608*
$d = 50$				
OT-CFM	13.5145 ± 0.0466	13.6811 ± 0.1314	8.1252 ± 0.1103	15.5900 ± 0.2287
MFM / LAND	12.9526 ± 0.0605	13.0996 ± 0.1029	8.6687 ± 0.3880	14.6185 ± 0.1085
Curly-FM	13.3381 ± 0.1482	13.4440 ± 0.2069	**7.2931 ± 0.2558**	15.0427 ± 0.1562
FFM (ours), deterministic	12.5784 ± 0.0235	**12.7715 ± 0.0719**	7.7975 ± 0.1005	13.9368 ± 0.1217
FFM (ours), with noise	**12.5732 ± 0.0573**	12.7723 ± 0.0952	7.7411 ± 0.1297	**13.7800 ± 0.0861**
sampling floor	*6.7529*	*6.7529*	*7.0275*	*7.2123*

Method	Shared coupling	Method-specific coupling
Graph-drift FM (control)	11.9409 ± 0.1110	11.9223 ± 0.0909
Anisotropic fixed-time FM 25	11.5712 ± 0.0880	11.4344 ± 0.1067
Second-moment FFM ($m = 0$)	11.5424 ± 0.1226	11.4967 ± 0.0878
FFM, one-form deleted (existing ablation)	11.1740 ± 0.0759	11.1183 ± 0.0774
FFM (ours)	**10.3967 ± 0.1371**	**10.3960 ± 0.1396**

Figure 3: Pancreas trajectory reconstruction. Left: Two-dimensional representation. Top right: W_2 at $t = 1/2$, the average $\overline{W}_2$ over 11 fixed evaluation times in $[1/3, 2/3]$, endpoint W_2, and $W_{2,F}$ Bottom right: same-information objective ablation at $d = 20$. "Shared" fixes the endpoint coupling "method-specific" constructs the coupling separately for each method. Ablation values are $W_2(t = 1/2)$, mean ± population standard deviation over five seeds.

Method	$d = 2$ mid W_2	$d = 20$ mid W_2	$d = 50$ mid W_2
OT-CFM	1.3631 ± 0.0368	15.9054 ± 0.0804	18.8807 ± 0.1172
MFM / LAND	1.3563 ± 0.0537	15.0519 ± 0.2250	18.2036 ± 0.2663
FFM, expression only	1.7162 ± 0.9919	12.8597 ± 0.1007	**16.6979 ± 0.3020**
FFM, + velocity	**1.0204 ± 0.1040**	12.7288 ± 0.1253	16.7371 ± 0.2566
FFM, + lineage	1.3074 ± 0.1083	12.7368 ± 0.1455	16.7415 ± 0.2426
FFM, + lineage + velocity	1.1773 ± 0.1404	**12.7121 ± 0.1314**	16.9196 ± 0.2700

Table 1: Intermediate-distribution recovery on iTracer R2 for different transition-kernel constructions. Lower is better.

6 Discussion

We present Finsler Flow Matching (FFM), a framework for transforming discrete Markov transition graphs into continuous stochastic dynamics through a transition-derived Finsler geometry. Its joint first- and second-moment construction captures directional preference and anisotropic accessibility, which guide source–target coupling, geodesic interpolation, and stochastic conditional paths.

Our experiments suggest that this is particularly useful for strongly directional or anisotropic dynamics. On the synthetic Sheet benchmark, ablating either component degrades intermediate-state recovery, while on Pancreas the advantage of FFM becomes more pronounced in higher-dimensional representations. The iTracer experiment further demonstrates the flexibility of the Markov transition kernel as an interface between heterogeneous measurements and generative dynamics. Expression, RNA velocity, lineage information, and their combinations can all be incorporated through the same representation. Nevertheless, additional modalities do not uniformly improve recovery. FFM therefore inherits both the information and the limitations of the supplied transition kernel, highlighting the importance of how experimental measurements are translated into transition probabilities.

These findings should be interpreted within the scope of the supplied dynamical prior. FFM does not resolve the non-identifiability of cellular dynamics from snapshots alone, and interpreting its geodesics as most-probable transition curves requires the stated diffusion-consistency and small-noise assumptions. Moreover, the Gaussian conditional paths used for stochastic training are tractable approximations rather than exact bridges of the graph-implied reference diffusion. Future work could explicitly model uncertainty in the transition graph and develop richer conditional path families. Another promising direction is conditional FFM, in which experimental covariates or perturbations parameterize the induced geometry, allowing the model to capture condition-specific changes in trajectory dynamics.

AI USE STATEMENT

We used generative AI tools to assist with code generation, in particular code for generating synthetic datasets, as well as sentence reformulation and language polishing throughout the manuscript. Furthermore, we used deep research tools to discover papers. In particular, it found the connections to FW Theory, which we subsequently investigated, verified, and developed in the context of our method. AI tools were further used to assist with completing and checking mathematical proofs. The theorem statements and underlying proof strategies were developed by the authors, and all AI-assisted derivations were independently checked for correctness. We reviewed all the AI generated output and take full responsibility for all of the content in this paper.

REPRODUCIBILITY STATEMENT

We code artifact containing the implementation of Finsler Flow Matching, preprocessing and transition-kernel construction, synthetic-data generation, training and evaluation scripts, benchmark-specific configurations, hyperparameter settings, and the records used to generate the reported tables and figures. Software dependencies are pinned in requirements-lock.txt to facilitate reconstruction of the experimental environment. Experimental architecture, optimization settings, training budgets, validation procedures, and numbers of random seeds are reported in Cc. Dataset sources and preprocessing procedures are documented for each benchmark, and validation and test populations are kept disjoint during model and hyperparameter selection. We additionally provide the configurations used for all reported methods and hyperparameter searches, including the data-dependent scaling of the Finsler regularization parameters. Evaluation settings, including any subsampling used for optimal-transport calculations, are recorded alongside the experiments. Because some GPU operations are nondeterministic, repeated training runs are not guaranteed to be bitwise identical, we therefore report results across multiple random seeds. The complete repository will be made publicly available upon publication.

REFERENCES

Michael S. Albergo and Eric Vanden-Eijnden. Building Normalizing Flows with Stochastic Interpolants. In *International Conference on Learning Representations*, 2023. doi: 10.48550/arXiv.2209.15571. URL https://arxiv.org/abs/2209.15571. arXiv:2209.15571 [cs.LG].

Georgios Arvanitidis, Lars Kai Hansen, and Søren Hauberg. A Locally Adaptive Normal Distribution. In *Conference on Neural Information Processing Systems*, June 2016. URL https://arxiv.org/abs/1606.02518v3.

Jacob Bamberger, Iolo Jones, Dennis Duncan, Michael Bronstein, Pierre Vandergheynst, and Adam Gosztolai. Carré du champ flow matching: better quality-generalisation tradeoff in generative models. *International Conference on Learning Representations*, 2026:135754–135779, April 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/dbc9f9d486bd1714896111b0fcd78cd1-Abstract-Conference.html?utm_source=chatgpt.com.

D. Bao, S.-S. Chern, and Z. Shen. *An Introduction to Riemann-Finsler Geometry*, volume 200 of *Graduate Texts in Mathematics*. Springer, New York, NY, 2000. ISBN 978-1-4612-7070-6 978-1-4612-1268-3. doi: 10.1007/978-1-4612-1268-3. URL http://link.springer.com/10.1007/978-1-4612-1268-3.

Aimée Bastidas-Ponce, Sophie Tritschler, Leander Dony, Katharina Scheibner, Marta Tarquis-Medina, Ciro Salinno, Silvia Schirge, Ingo Burtscher, Anika Böttcher, Fabian J. Theis, Heiko Lickert, and Mostafa Bakhti. Comprehensive single cell mRNA profiling reveals a detailed roadmap for pancreatic endocrinogenesis. *Development*, 146(12):dev173849, June 2019. ISSN 1477-9129. doi: 10.1242/dev.173849.

Nico Battich, Joep Beumer, Buys de Barbanson, Lenno Krenning, Chloé S. Baron, Marvin E. Tanenbaum, Hans Clevers, and Alexander van Oudenaarden. Sequencing metabolically labeled transcripts in single cells reveals mRNA turnover strategies. *Science*, 367(6482):1151–1156, March 2020. doi: 10.1126/science.aax3072. URL https://www.science.org/doi/10.1126/science.aax3072.

Volker Bergen, Marius Lange, Stefan Peidli, F. Alexander Wolf, and Fabian J. Theis. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology*, 38 (12):1408–1414, December 2020. ISSN 1546-1696. doi: 10.1038/s41587-020-0591-3. URL https://www.nature.com/articles/s41587-020-0591-3.

Tyrus Berry and Timothy Sauer. Local Kernels and the Geometric Structure of Data, January 2015. URL http://arxiv.org/abs/1407.1426. arXiv:1407.1426 [math.CA].

Charlotte Bunne, Laetitia Papaxanthos, Andreas Krause, and Marco Cuturi. Proximal Optimal Transport Modeling of Population Dynamics. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, pp. 6511–6528. PMLR, May 2022. URL https://proceedings.mlr.press/v151/bunne22a.html.

Charlotte Bunne, Ya-Ping Hsieh, Marco Cuturi, and Andreas Krause. The Schrödinger Bridge between Gaussian Measures has a Closed Form. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, pp. 5802–5833. PMLR, April 2023. URL https://proceedings.mlr.press/v206/bunne23a.html.

Junyue Cao, Wei Zhou, Frank Steemers, Cole Trapnell, and Jay Shendure. Sci-fate characterizes the dynamics of gene expression in single cells. *Nature Biotechnology*, 38(8):980–988, August 2020. ISSN 1546-1696. doi: 10.1038/s41587-020-0480-9.

Ricky T. Q. Chen and Yaron Lipman. Flow Matching on General Geometries. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=g7ohDlTITL.

Marco Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/hash/af21d0c97db2e27e13572cbf59eb343d-Abstract.html.

Jonas Simon Fleck, Sophie Martina Johanna Jansen, Damian Wollny, Fides Zenk, Makiko Seimiya, Akanksha Jain, Ryoko Okamoto, Malgorzata Santel, Zhisong He, J. Gray Camp, and Barbara Treutlein. Inferring and perturbing cell fate regulomes in human brain organoids. *Nature*, 621 (7978):365–372, September 2023. ISSN 1476-4687. doi: 10.1038/s41586-022-05279-8. URL https://www.nature.com/articles/s41586-022-05279-8.

Mark I. Freidlin and Alexander D. Wentzell. *Random Perturbations of Dynamical Systems*, volume 260 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 2012. ISBN 978-3-642-25846-6 978-3-642-25847-3. doi: 10.1007/978-3-642-25847-3. URL https://link.springer.com/10.1007/978-3-642-25847-3.

Hans Föllmer. Random fields and diffusion processes. In Persi Diaconis, David Elworthy, Hans Föllmer, Edward Nelson, George Papanicolaou, Srinivasa R. S. Varadhan, and Paul-Louis Hennequin (eds.), *École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, pp. 101–203, Berlin, Heidelberg, 1988. Springer. ISBN 978-3-540-46042-8. doi: 10.1007/BFb0086180.

Helmuth Gehart, Johan H. van Es, Karien Hamer, Joep Beumer, Kai Kretzschmar, Johanna F. Dekkers, Anne Rios, and Hans Clevers. Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping. *Cell*, 176(5):1158–1173.e16, February 2019. ISSN 1097-4172. doi: 10.1016/j.cell.2018.12.029.

Doron Haviv, Aram-Alexandre Pooladian, Dana Pe'er, and Brandon Amos. Wasserstein Flow Matching: Generative modeling over families of distributions, May 2025. URL http://arxiv.org/abs/2411.00698. arXiv:2411.00698 [cs.LG].

Zhisong He, Ashley Maynard, Akanksha Jain, Tobias Gerber, Rebecca Petri, Hsiu-Chuan Lin, Malgorzata Santel, Kevin Ly, Jean-Samuel Dupré, Leila Sidow, Fatima Sanchis Calleja, Sophie M. J. Jansen, Stephan Riesenberg, J. Gray Camp, and Barbara Treutlein. Lineage recording in human cerebral organoids. *Nature Methods*, 19(1):90–99, January 2022. ISSN 1548-7105. doi: 10.1038/s41592-021-01344-8. URL https://www.nature.com/articles/s41592-021-01344-8.

Matthias Heymann and Eric Vanden-Eijnden. The geometric minimum action method: A least action principle on the space of curves. *Communications on Pure and Applied Mathematics*, 61(8):1052–1117, August 2008. ISSN 0010-3640, 1097-0312. doi: 10.1002/cpa.20238. URL https://onlinelibrary.wiley.com/doi/10.1002/cpa.20238.

Peter Holderrieth and Ezra Erives. An Introduction to Flow Matching and Diffusion Models, 2026. URL https://diffusion.csail.mit.edu/.

Wanli Hong, Yuliang Shi, and Jonathan Niles-Weed. Trajectory Inference with Smooth Schrödinger Bridges. In *Proceedings of the 42nd International Conference on Machine Learning*, pp. 23722–23750. PMLR, October 2025. URL https://proceedings.mlr.press/v267/hong25f.html.

Guillaume Huguet, D S Magruder, Alexander Tong, Oluwadamilola Fasina, Manik Kuchroo, Guy Wolf, and Smita Krishnaswamy. Manifold Interpolating Optimal-Transport Flows for Trajectory Inference. In *Neural Information Processing Systems*, 2022.

Kacper Kapusniak, 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. In *Conference on Neural Information Processing Systems*, 2024.

Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, Peter Lönnerberg, Alessandro Furlan, Jean Fan, Lars E. Borm, Zehua Liu, David van Bruggen, Jimin Guo, Xiaoling He, Roger Barker, Erik Sundström, Gonçalo Castelo-Branco, Patrick Cramer, Igor Adameyko, Sten Linnarsson, and Peter V. Kharchenko. RNA velocity of single cells. *Nature*, 560(7719):494–498, August 2018. ISSN 1476-4687. doi: 10.1038/s41586-018-0414-6. URL https://www.nature.com/articles/s41586-018-0414-6.

Marius Lange, Volker Bergen, Michal Klein, Manu Setty, Bernhard Reuter, Mostafa Bakhti, Heiko Lickert, Meshal Ansari, Janine Schniering, Herbert B. Schiller, Dana Pe'er, and Fabian J. Theis. CellRank for directed single-cell fate mapping. *Nature Methods*, 19(2):159–170, February 2022. ISSN 1548-7105. doi: 10.1038/s41592-021-01346-6. URL https://www.nature.com/articles/s41592-021-01346-6.

Marius Lange, Zoe Piran, Michal Klein, Bastiaan Spanjaard, Dominik Klein, Jan Philipp Junker, Fabian J. Theis, and Mor Nitzan. Mapping lineage-traced cells across time points with moslin. *Genome Biology*, 25(1):277, October 2024. ISSN 1474-760X. doi: 10.1186/s13059-024-03422-4. URL https://doi.org/10.1186/s13059-024-03422-4.

Tiejun Li, Jifan Shi, Yichong Wu, and Peijie Zhou. On the Mathematics of RNA Velocity I: Theoretical Analysis. *CSIAM Transactions on Applied Mathematics*, 2(1):1–55, February 2021. ISSN 2708-0560, 2708-0579. doi: 10.4208/csiam-am.SO-2020-0001. URL https://www.global-sci.com/csiam-am/article/view/7770.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. Flow Matching for Generative Modeling. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=PqvMRDCJT9t.

Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport, August 2013. URL http://arxiv.org/abs/1308.0215. arXiv:1308.0215 [math.PR].

Aaron McKenna, Gregory M. Findlay, James A. Gagnon, Marshall S. Horwitz, Alexander F. Schier, and Jay Shendure. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science*, 353(6298):aaf7907, July 2016. doi: 10.1126/science.aaf7907. URL https://www.science.org/doi/10.1126/science.aaf7907.

Kirill Neklyudov, Rob Brekelmans, Daniel Severo, and Alireza Makhzani. Action Matching: Learning Stochastic Dynamics from Samples. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 25858–25889. PMLR, July 2023. URL https://proceedings.mlr.press/v202/neklyudov23a.html.

Shin-ichi Ohta. Finsler interpolation inequalities. *Calculus of Variations and Partial Differential Equations*, 36(2):211–249, October 2009. ISSN 0944-2669, 1432-0835. doi: 10.1007/s00526-009-0227-4. URL http://link.springer.com/10.1007/s00526-009-0227-4.

Shin-ichi Ohta and Karl-Theodor Sturm. Heat flow on Finsler manifolds. *Communications on Pure and Applied Mathematics*, 62(10):1386–1433, 2009. ISSN 1097-0312. doi: 10.1002/cpa.20273. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.20273.

Katarina Petrović, Lazar Atanackovic, Viggo Moro, Kacper Kapuśniak, İsmail İlkan Ceylan, Michael Bronstein, Avishek Joey Bose, and Alexander Tong. Curly Flow Matching for Learning Non-gradient Field Dynamics. In *Conference on Neural Information Processing Systems*, 2025. URL http://arxiv.org/abs/2510.26645. arXiv:2510.26645 [cs.LG].

Blanca Pijuan-Sala, Jonathan A. Griffiths, Carolina Guibentif, Tom W. Hiscock, Wajid Jawaid, Fernando J. Calero-Nieto, Carla Mulas, Ximena Ibarra-Soria, Richard C. V. Tyser, Debbie Lee Lian Ho, Wolf Reik, Shankar Srinivas, Benjamin D. Simons, Jennifer Nichols, John C. Marioni, and Berthold Göttgens. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature*, 566(7745):490–495, February 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-0933-9.

Martin Rohbeck, Edward De Brouwer, Charlotte Bunne, Anne Biton, Kelvin Y Chen, Aviv Regev, and Romain Lopez. Modeling Complex System Dynamics with Flow Matching Across Time and Conditions. In *International Conference on Learning Representations*, 2025.

Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, Lia Lee, Jenny Chen, Justin Brumbaugh, Philippe Rigollet, Konrad Hochedlinger, Rudolf Jaenisch, Aviv Regev, and Eric S. Lander. Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell*, 176(4):928–943.e22, February 2019. ISSN 1097-4172. doi: 10.1016/j.cell.2019.01.006.

Alex J. Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222, August 2004. ISSN 1573-1375. doi: 10.1023/B:STCO.0000035301.49549.88. URL https://doi.org/10.1023/B:STCO.0000035301.49549.88.

Daniel Ting, Ling Huang, and Michael Jordan. An Analysis of the Convergence of Graph Laplacians, January 2011. URL http://arxiv.org/abs/1101.5435. arXiv:1101.5435 [stat.ML].

Alexander Tong, Jessie Huang, Guy Wolf, David van Dijk, and Smita Krishnaswamy. TrajectoryNet: A Dynamic Optimal Transport Network for Modeling Cellular Dynamics. *Proceedings of machine learning research*, 119:9526–9536, July 2020. ISSN 2640-3498. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC8320749/.

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. 2024a. doi: 10.48550/arXiv.2302.00482. URL http://arxiv.org/abs/2302.00482. arXiv:2302.00482 [cs.LG].

Alexander Y. Tong, Nikolay Malkin, Kilian Fatras, Lazar Atanackovic, Yanlei Zhang, Guillaume Huguet, Guy Wolf, and Yoshua Bengio. Simulation-Free Schrödinger Bridges via Score and Flow Matching. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, pp. 1279–1287. PMLR, April 2024b. URL https://proceedings.mlr.press/v238/tong24a.html.

Cole Trapnell, Davide Cacchiarelli, Jonna Grimsby, Prapti Pokharel, Shuqiang Li, Michael Morse, Niall J. Lennon, Kenneth J. Livak, Tarjei S. Mikkelsen, and John L. Rinn. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, 32(4):381–386, April 2014. ISSN 1546-1696. doi: 10.1038/nbt.2859.

Cédric Villani. *Optimal Transport*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 2009. ISBN 978-3-540-71049-3 978-3-540-71050-9. doi: 10.1007/978-3-540-71050-9. URL http://link.springer.com/10.1007/978-3-540-71050-9.

Daniel E. Wagner and Allon M. Klein. Lineage tracing meets single-cell omics: opportunities and challenges. *Nature Reviews. Genetics*, 21(7):410–427, July 2020. ISSN 1471-0064. doi: 10.1038/s41576-020-0223-2.

Philipp Weiler, Marius Lange, Michal Klein, Dana Pe'er, and Fabian Theis. CellRank 2: unified fate mapping in multiview single-cell data. *Nature Methods*, 21(7):1196–1205, July 2024. ISSN 1548-7105. doi: 10.1038/s41592-024-02303-9.

Caleb Weinreb, Samuel Wolock, Betsabeh K. Tusi, Merav Socolovsky, and Allon M. Klein. Fundamental limits on dynamic inference from single-cell snapshots. *Proceedings of the National Academy of Sciences*, 115(10):E2467–E2476, March 2018. doi: 10.1073/pnas.1714723115. URL https://www.pnas.org/doi/10.1073/pnas.1714723115.

Caleb Weinreb, Alejo Rodriguez-Fraticelli, Fernando D. Camargo, and Allon M. Klein. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science*, 367(6479): eaaw3381, February 2020. doi: 10.1126/science.aaw3381. URL https://www.science.org/doi/10.1126/science.aaw3381.

Zhenyi Zhang, Tiejun Li, and Peijie Zhou. Learning stochastic dynamics from snapshots through regularized unbalanced optimal transport. In *International Conference on Learning Representations*, 2025. doi: 10.48550/arXiv.2410.00844. URL http://arxiv.org/abs/2410.00844.

Aaron Zweig, Mingxuan Zhang, David A. Knowles, and Elham Azizi. Learning Lineage-guided Geodesics with Finsler Geometry, March 2026. URL http://arxiv.org/abs/2603.16708. arXiv:2603.16708 [cs.LG].

A PRELIMIARIES

Finsler Geometry: For a complete introduction to Finsler geometry we refer the reader to Bao et al. (2000). Let M be an $m-$dimensional manifold embedded in $\mathbb{R}^d$. The *tangent bundle* $TM := \bigsqcup_{x\in M} T_xM$ is the $2m$-dimensional manifold composed of pairs (x, v) with $x \in M$ and $v \in T_xM$, where T_xM is the *tangent space* at x.

Definition 3 (Finsler structure)**.** A *Finsler structure* on M is a continuous function $F : TM \to \mathbb{R}_{\geq 0}$, which is smooth on $TM \setminus \{0\}$, such that for all $(x, v) \in TM$:

(F1) *Positive homogeneity:* $F(x, \lambda v) = \lambda F(x, v)$ for all $\lambda > 0$.

(F2) *Strong convexity:* the Hessian matrix $g_{ij}(x, v) = \frac{1}{2}\partial^2 F^2/\partial v^i \partial v^j$ is positive definite for $v \neq 0$.

(F3) *Positivity:* $F(x, v) > 0$ for $v \neq 0$.

In general, we have that $F(x, -v) \neq F(x, v)$ hence, the cost of moving in direction v is different from the cost of moving in direction $-v$. A Riemannian metric is the special case $F(x, v) = \sqrt{v^\top G(x)v}$, which additionally satisfies $F(x, -v) = F(x, v)$. Analogously to Riemannian Geometry, one can define a notion of length and geodesics.

Definition 4 (Length and geodesics)**.** For an absolutely continuous curve $\gamma : [0, 1] \to M$, define

$$L_F(\gamma) = \int_0^1 F(\gamma(t), \dot\gamma(t))\,\mathrm{d}t, \qquad E_F(\gamma) = \int_0^1 F(\gamma(t), \dot\gamma(t))^2\,\mathrm{d}t.$$

A *geodesic* between $x_0, x_1 \in \mathbb{R}^d$ is a curve minimizing L_F subject to $\gamma(0) = x_0$, $\gamma(1) = x_1$. Note that energy minimizers are constant-speed parametrized geodesics. Furthermore, we define the *length metric* on M as:

$$c_F(x_0, x_1) = \inf_\gamma \int_0^1 F(\gamma_t, \dot\gamma_t)\,dt,$$

where the infimum is taken over curves satisfying $\gamma_0 = x_0,\ \gamma_1 = x_1$.

We are particularly interested in the special case called a *Randers Metric*, where $G(x)$ is a smooth Riemannian metric and $\beta(x)$ a smooth 1-form and we define

$$F(x, v) = \sqrt{v^\top G(x)v}\ +\ \beta(x)^\top v. \tag{14}$$

Lemma 5. *[Bao et al. (2000)] The Randers structure* (14) *is a Finsler structure if and only if* $\|\beta(x)\|_{G^{-1}(x)} < 1$ *pointwise, where* $\|u\|^2_{G^{-1}} = u^\top G^{-1}u$.

Remark 6 (Approximation of the data manifold). Although the underlying dynamics may be concentrated on an unknown low-dimensional manifold $\mathcal{M} \subset \mathbb{R}^d$, we do not estimate $\mathcal{M}$. Instead, we construct the Finsler metric on $\mathbb{R}^d$ so that motion along directions supported by the local data and transition geometry has low cost, whereas motion away from the data manifold, in particular along weakly supported normal directions, has high cost. Thus, the Finsler structures in this work will be defined on $\mathbb{R}^d$.

Definition 7 (P-derived Randers metric)**.** A P-derived Randers metric is a Randers metric

$$F_P(x_i, v) = \sqrt{v^\top G_P(x_i)v} + \beta_P(x_i)^\top v$$

whose metric tensor $G_P(x_i)$ and one-form $\beta_P(x_i)$ are constructed from local transition statistics of P, with

$$G_P(x_i) \in \mathrm{SPD}(d), \qquad \|\beta_P(x_i)\|_{G_P(x_i)^{-1}} < 1.$$

For example, if $C_P(x_i)$ denotes a local transition second-moment tensor, one may consider the family

$$G_{P,q}(x_i) = \big(\tau I + C_P(x_i)\big)^{-q}, \qquad q > 0,$$

together with any admissible transition-derived one-form β_P. For $q = 1$, this is closely related to local-kernel geometries, where the second moment determines the induced Riemannian metric (Berry & Sauer, 2015).

Optimal Transport: Optimal transport (OT) was originally formulated by Monge as the problem of finding a deterministic transport map $T : \mathbb{R}^d \to \mathbb{R}^d$ pushing p_0 onto p_1 while minimizing the transport cost:

$$\inf_{T:\, T_{\#}p_0 = p_1} \int_{\mathbb{R}^d} c(x, T(x))\, \mathrm{d}p_0(x).$$

Since a transport map cannot split mass, such a minimizers does not exist in general. For example, if

$$p_0 = \delta_0, \qquad p_1 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1,$$

no map T can satisfy $T_{\#}p_0 = p_1$. The Kantorovich relaxation resolves this issue by optimizing instead over transport plans $\pi \in \Pi(p_0, p_1)$:

$$\inf_{\pi \in \Pi(p_0,p_1)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c(x_0, x_1)\, \mathrm{d}\pi(x_0, x_1).$$

In the example above, an admissible plan would be

$$\pi = \frac{1}{2}\delta_{(0,-1)} + \frac{1}{2}\delta_{(0,1)}.$$

Under the Euclidean squared distance $c(x_0, x_1) = \frac{1}{2}\|x_0 - x_1\|^2$ (absolute continuity of p_0) ***Brenier's theorem*** guarantees that the optimal plan $\pi^\star$ exists and is unique. Furthermore, there is a unique optimal transport map $T = \nabla\varphi$, for some convex function φ. The classical ***Benamou–Brenier theorem*** establishes that this static transport problem is equivalent to a dynamic formulation over density paths ρ_t and velocity fields v_t:

$$W_2^2(p_0, p_1) = \inf_{(\rho,v)} \int_0^1 \int_{\mathbb{R}^d} |v_t(x)|^2 \mathrm{d}\rho_t(x), \mathrm{d}t \quad \text{s.t.} \quad \partial_t\rho_t + \nabla\cdot(\rho_t v_t) = 0, \quad \rho_0 = p_0, \quad \rho_1 = p_1.$$

When substituting the Euclidean cost with the Finsler geodesic action cost $c_F(x_0, x_1)$ defined in (1), the static Kantorovich problem remains uniquely attained (Ohta, 2009, Theorem 4.10). Moreover, the static problem coincides with the dynamic Finsler Benamou-Brenier formulation:

$$W_{2,F}^2(p_0, p_1) \coloneqq \inf_{\pi \in \Pi(p_0,p_1)} \int_{\mathbb{R}^d \times \mathbb{R}^d} c_F(x_0, x_1)^2 \mathrm{d}\pi(x_0, x_1) = \inf_{(\rho,v)} \int_0^1 \int_{\mathbb{R}^d} F(x, v_t(x))^2 \mathrm{d}\rho_t(x)\mathrm{d}t$$

subject to the continuity equation $\partial_t\rho_t + \nabla\cdot(\rho_t v_t) = 0$ with endpoint conditions $\rho_0 = p_0$ and $\rho_1 = p_1$ (Ohta & Sturm, 2009, Theorem 7.3 and Corollary 7.5).

Flow Matching: We give a brief introduction to Flow Matching for a more in depth treatment consider Holderrieth & Erives (2026). The idea behind flow matching is to find a vector field that transports one distribution into another and can generate samples by integrating said vector field. More formally one can think of it as trying to find a vector field $u_t : [0, 1) \times \mathbb{R}^n \to \mathbb{R}^n$ such that

$$\begin{cases} \dfrac{d}{dt}x_t = u_t(x_t), \\ x_0 \sim p_0, x_1 \sim p_1. \end{cases}$$

Once such a vector field has been learned, samples from p_1 can be generated by sampling $x_0 \sim p_0$ and integrating the vector field from $t = 0$ to $t = 1$. In general, this vector field is not unique, which gives us some choice. For example, if $p_0 = p_1 = p$, then any vector field satisfying $\nabla\cdot(pu) = 0$ leaves the distribution p invariant. More generally, the endpoint distributions p_0 and p_1 do not uniquely determine the intermediate probability path. We exploit this freedom by choosing a probability density path $(p_t)_{t\in[0,1]}$ connecting p_0 and p_1, and then finding a vector field whose flow realizes this path. Even for a fixed path p_t, such a vector field need not be unique (one could make the choice optimal by choosing u_t as the minimizer of some action).

In general, we do not have p_0 and p_1 available in closed form, and hence the corresponding marginal path p_t is intractable. The idea is to condition the path on samples; then the marginal distributions become Dirac distributions. We therefore introduce a coupling $\pi \in \Pi(p_0, p_1)$, i.e., a

probability distribution on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals p_0 and p_1, and define a tractable conditional probability path $p_t(x \mid x_0, x_1)$ with marginals δ_{x_0} and δ_{x_1}. The corresponding marginal path is

$$p_t(x) = \int_{\mathbb{R}^n \times \mathbb{R}^n} p_t(x \mid x_0, x_1)\, d\pi(x_0, x_1).$$

For every endpoint pair (x_0, x_1), we define a conditional vector field $u_t(x \mid x_0, x_1)$ that solves

$$\begin{cases} \dfrac{d}{dt} x_t = u_t(x_t \mid x_0, x_1), \\ x_t \sim p_t(\cdot \mid x_0, x_1) \text{ for all } t \in [0, 1]. \end{cases}$$

Surprisingly, although the marginal vector field is generally intractable, the Conditional Flow Matching objective

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{\substack{(x_0,x_1)\sim\pi,\, t\sim\mathcal{U}[0,1], \\ x\sim p_t(\cdot|x_0,x_1)}} \left[\| v_\theta(t, x) - u_t(x \mid x_0, x_1) \|_2^2 \right]$$

has the same gradient with respect to θ as the corresponding unconditional Flow Matching objective. Hence, we can learn the marginal vector field using only samples from tractable conditional probability paths, resulting in a simulation-free training procedure.

Simulation free Score and Flow Matching: Here we introduce some of the ideas of Tong et al. (2024b) [3]. Similarly to the previous section, we want to find a vector field $\tilde{u}_t$ that satisfies.

$$\begin{cases} \mathrm{d}x_t = \tilde{u}_t(x_t)\mathrm{d}t + g_t dW_t, \\ x_0 \sim p_0, x_1 \sim p_1. \end{cases}$$

where g_t is some scalar valued function controlling the diffusion strength and W_t is a standard Brownian motion, i.e. a continuous stochastic process with independent Gaussian increments. Hence, instead of an ODE we treat a stochastic differential equation (SDE).

In a first step we go back again the setting of the previous section and choose a conditional probability path $p_t(x \mid x_0, x_1)$ and learn a conditional vector field such that the corresponding ODE

$$\frac{d}{dt} x_t = u_t(x_t)$$

satisfies $x_t \sim p_t$ for every t. We additionally introduce the conditional score of the probability path,

$$s_t(x \mid x_0, x_1) := \nabla \log p_t(x \mid x_0, x_1).$$

Note that similarly to flow matching, also the conditional score matching loss has the same gradient unconditional one. Using this score, the deterministic flow can be transformed into a stochastic process,

$$\mathrm{d}x_t = \left[v_t(x_t) + \frac{1}{2} g_t g_t^\top s_t(x_t) \right] \mathrm{d}t + g_t\, dW_t.$$

The Fokker-Planck equation of this SDE is given by

$$\begin{aligned} \partial_t p_t &= -\nabla \cdot \left(p_t \left[v_t + \frac{1}{2} g_t g_t^\top s_t \right] \right) + \frac{1}{2} \nabla \cdot \left(g_t g_t^\top \nabla p_t \right) \\ &= -\nabla \cdot (p_t v_t), \end{aligned}$$

where we used $p_t s_t = \nabla p_t$. Hence, the ODE and the corresponding SDE satisfy the same Fokker–Planck equation and therefore have the same marginals $X_t \sim p_t$. The score thus allows us to add stochasticity to the learned flow without changing the prescribed probability path.

[3] Which the authors highly suggest reading in full.

B THEORETICAL CONTRIBUTIONS

Transition-supported geodesics. For a threshold $\tau > 0$, define the high-probability transition edges

$$E_\tau := \{(i,j) : P_{ij} \geq \tau\}.$$

For endpoints $x_0, x_1 \in \mathcal{D}$, define the transition-supported subset

$$\mathcal{S}_\tau(x_0, x_1) := \{x_i \in \mathcal{D} : x_i \text{ lies on a directed path from } x_0 \text{ to } x_1 \text{ using only edges in } E_\tau\}.$$

We assume that $\mathcal{S}_\tau(x_0, x_1)$ is non-empty. For $r > 0$, define its r-neighborhood by

$$U_r(\mathcal{S}_\tau) := \left\{x \in \mathbb{R}^d : \operatorname{dist}(x, \mathcal{S}_\tau) < r\right\}, \qquad \operatorname{dist}(x, \mathcal{S}_\tau) := \min_{y \in \mathcal{S}_\tau} \|x - y\|_2.$$

Furthermore, define

$$c_r := \min_{\substack{r/2 \leq \operatorname{dist}(x, \mathcal{S}_\tau) \leq r \\ \|v\|_2 = 1}} F_P(x, v) > 0.$$

The positivity follows since F_P is continuous and strictly positive for $v \neq 0$, while the domain of minimization is compact. For our particular P-derived Randers metric (9), c_r can be related directly to the transition moments. By Cauchy–Schwarz,

$$F_P(x, v) \geq \frac{\sqrt{\|m(x)\|^2_{C_\rho(x)^{-1}} + \lambda^2} - \|m(x)\|_{C_\rho(x)^{-1}}}{\sqrt{\lambda_{\max}(C_\rho(x))}} \|v\|_2.$$

Hence, regions and directions with large P-weighted transition covariance are relatively cheap, whereas regions with little transition support, corresponding to small eigenvalues of C_ρ, incur a larger cost. Thus annuli farther from $\mathcal{S}_\tau$ have a larger escape cost c_r.

Proposition 8 (Confinement and directional preference of geodesics)**.** *Let F_P be the Randers metric induced by the first and second transition moments m and C in* (9)*. Suppose there exists a curve $\bar{\gamma}$ connecting x_0 and x_1 inside $U_{r/2}(\mathcal{S}_\tau)$. If*

$$L_{F_P}(\bar{\gamma}) < c_r r,$$

then every minimizing geodesic connecting x_0 and x_1 is contained in $U_r(\mathcal{S}_\tau)$.

Moreover, for fixed local speed $\|v\|_{C_\rho(x)^{-1}}$, the Finsler cost is minimized when v is positively aligned with the first transition moment $m(x)$.

Proof. Suppose a minimizing geodesic $\gamma^\star$ leaves $U_r(\mathcal{S}_\tau)$. Since $x_0, x_1 \in \mathcal{S}_\tau$, it must cross the annulus

$$r/2 \leq \operatorname{dist}(x, \mathcal{S}_\tau) \leq r$$

once when leaving and once when returning. Since $x \mapsto \operatorname{dist}(x, \mathcal{S}_\tau)$ is 1-Lipschitz, its total Euclidean length in this annulus is at least r. By positive homogeneity and the definition of c_r,

$$L_{F_P}(\gamma^\star) \geq c_r r > L_{F_P}(\bar{\gamma}),$$

contradicting minimality. Hence $\gamma^\star \subset U_r(\mathcal{S}_\tau)$.

For the directional claim, write $A(x) = C_\rho(x)^{-1}$. For fixed $\|v\|_{A(x)}$, the first term of

$$F_P(x, v) = \sqrt{\|m(x)\|^2_{A(x)} + \lambda^2}\, \|v\|_{A(x)} - \langle v, m(x)\rangle_{A(x)}$$

is constant. Thus minimizing F_P is equivalent to maximizing $\langle v, m(x)\rangle_{A(x)}$, which by Cauchy–Schwarz is attained when v is a positive multiple of $m(x)$. □

Uniform Convergence of Empirical Drift and Diffusion: Let Y_T be a stochastic process in $\mathbb{R}^d$ governed by the Stochastic Differential Equation (SDE)

$$dY_T = b_0(Y_T)\, dt + \Sigma_{\text{sde}}(Y_T)\, dW_t, \tag{15}$$

where $b_0 \in C^\infty(\mathbb{R}^d, \mathbb{R}^d)$ is the drift vector, W_t is a standard d-dimensional Brownian motion and $D(x) \coloneqq \Sigma_{\text{sde}}(x)\Sigma_{\text{sde}}(x)^\top$, is the smooth diffusion matrix.

Definition 9 (Transition Operator and Infinitesimal Generator)**.** The transition operator (semigroup) T_t acting on a test function $f \in C_c^\infty(\mathbb{R}^d)$ is defined by

$$T_t f(x) \coloneqq \mathbb{E}\left[f(Y_T) \mid Y_0 = x\right].$$

The infinitesimal generator L of the process Y_T is the operator

$$Lf(x) \coloneqq \lim_{t\to 0^+} \frac{T_t f(x) - f(x)}{t} = \nabla f(x)^\top b_0(x) + \frac{1}{2}\mathrm{Tr}\left(D(x)\nabla^2 f(x)\right).$$

Suppose we observe samples $\{x_i\}_{i=1}^N \subset \mathbb{R}^d$ drawn i.i.d. from a continuous density $p(x)$ alongside a row-stochastic transition matrix $P \in \mathbb{R}^{N\times N}$ constructed at spatial scale $\varepsilon_N > 0$.Thus the discrete transition operator is

$$(Pf)(x_i) = \sum_{j=1}^N P_{ij} f(x_j).$$

To link discrete dataset observations to the continuous diffusion process, we need the following assumptions.

Assumption 10. The discrete time step Δt_N scales quadratically with the bandwidth parameter ε_N

$$\Delta t_N = c\varepsilon_N^2 \quad \text{for a fixed constant } c > 0. \text{ (Time-Space Scaling)}$$

Informally, we say that a Markov chain with transition matrix P is a discrete observation of a continuous stochastic process Y_T if its discrete transition operator matches the continuous process generator L over an infinitesimal time step. Formally, we assume:

Assumption 11 (Uniform discrete-generator consistency)**.** Define the empirical generator

$$L_{N,\varepsilon_N} f(x_i) \coloneqq \frac{1}{\Delta t_N}\left(\sum_{j=1}^N P_{ij} f(x_j) - f(x_i)\right).$$

We assume that for every $f \in C_c^\infty(\mathbb{R}^d)$,

$$\sup_{x_i \in K} |L_{N,\varepsilon_N} f(x_i) - Lf(x_i)| \to 0$$

as $N \to \infty$ and $\varepsilon_N \to 0$.

Assumption 12. We assume local support of the transition kernel on the region of interest: for every compact $K \subset \mathbb{R}^d$, there exists a compact neighborhood $K^+ \supset K$ such that, for all sufficiently large N,

$$x_i \in K, \quad P_{ij} > 0 \qquad \Longrightarrow \qquad x_j \in K^+.$$

Proposition 13 (Recovery of drift and diffusion from P)**.** *Assume that $P = P_{N,\varepsilon_N}$ is row-stochastic and that Assumptions 10 11 and 12 hold. Define*

$$\hat{b}_{0,N}(x_i) = \frac{1}{\Delta t_N}\sum_{j=1}^N P_{ij}(x_j - x_i) \text{ and } \hat{D}_N(x_i) = \frac{1}{\Delta t_N}\sum_{j=1}^N P_{ij}(x_j - x_i)(x_j - x_i)^\top.$$

Then

$$\max_{x_i\in K} \|\hat{b}_{0,N}(x_i) - b_0(x_i)\| \to 0 \text{ and } \max_{x_i\in K} \|\hat{D}_N(x_i) - D(x_i)\| \to 0.$$

Proof. Fix a compact set K and let K^+ be the compact neighborhood from 12. Choose a cutoff $\chi \in C_c^\infty(\mathbb{R}^d)$ such that $\chi \equiv 1$ on a neighborhood of K^+. For $a, b \in \{1, \ldots, d\}$ define

$$\widetilde{q}_a(x) \coloneqq \chi(x) x_a, \qquad \widetilde{q}_{ab}(x) \coloneqq \chi(x) x_a x_b.$$

Then $\widetilde{q}_a, \widetilde{q}_{ab} \in C_c^\infty(\mathbb{R}^d)$, so Assumption 11 applies. For $x_i \in K$ 12 implies that every x_j with $P_{ij} > 0$ lies in K^+, where $\chi \equiv 1$. Hence

$$L_{N,\varepsilon_N}\widetilde{q}_a(x_i) = L_{N,\varepsilon_N} q_a(x_i), \qquad L_{N,\varepsilon_N}\widetilde{q}_{ab}(x_i) = L_{N,\varepsilon_N} q_{ab}(x_i).$$

Moreover, since $\chi \equiv 1$ on a neighborhood of K,

$$L\widetilde{q}_a(x) = Lq_a(x), \qquad L\widetilde{q}_{ab}(x) = Lq_{ab}(x), \qquad x \in K.$$

Therefore Assumption 11 may be applied to the cutoff observables while yielding the desired generator identities for q_a and q_{ab} on K. Since P is row-stochastic,

$$L_{N,\varepsilon_N} q_a(x_i) = \frac{1}{\Delta t_N} \sum_{j=1}^{N} P_{ij}(x_{j,a} - x_{i,a}) = \hat{b}_{0,N,a}(x_i).$$

Moreover,

$$Lq_a(x) = b_{0,a}(x).$$

The drift convergence therefore follows directly from Assumption 11. Expanding the increments for the second moment gives

$$L_{N,\varepsilon_N} q_{ab}(x_i) = x_{i,a}\hat{b}_{0,N,b}(x_i) + x_{i,b}\hat{b}_{0,N,a}(x_i) + \hat{D}_{N,ab}(x_i).$$

On the other hand, applying the diffusion generator yields

$$Lq_{ab}(x) = x_a b_{0,b}(x) + x_b b_{0,a}(x) + D_{ab}(x).$$

Consequently,

$$\begin{aligned}\hat{D}_{N,ab}(x_i) - D_{ab}(x_i) = &\big(L_{N,\varepsilon_N} q_{ab}(x_i) - Lq_{ab}(x_i)\big) \\ &- x_{i,a}\big(L_{N,\varepsilon_N} q_b(x_i) - Lq_b(x_i)\big) \\ &- x_{i,b}\big(L_{N,\varepsilon_N} q_a(x_i) - Lq_a(x_i)\big).\end{aligned}$$

Since K is compact, its coordinates are uniformly bounded. Uniform generator consistency therefore implies uniform convergence of every entry of $\hat{D}_N$ to the corresponding entry of D. Since the dimension is finite, this implies convergence in any matrix norm. □

Generator consistency for kernels: The generator-consistency assumption 11 is closely related to continuum limits of local graph random walks (Ting et al., 2011, Theorems 2–3), whose limiting drift and diffusion are determined by the first two local moments (Berry & Sauer, 2015, Lemma 3.9 and Theorem 3.12). For neighbourhood kernels, sufficient conditions involve shrinking bandwidths, sufficiently dense sampling from a smooth positive density, and controlled local bandwidth estimates. For the generator consistency assumption the row-normalized kernel, including any velocity and lineage kernel, must satisfy

$$m_h = ch^2 b_0 + o(h^2), \qquad C_h = ch^2 D + o(h^2),$$

uniformly with higher-order displacement contributions divided by h^2 going to zero as h goes to zero. For the cosine and correlation constructions studied by Li et al. (2021) gives a first-order limit on the time scale h, rather than a diffusion limit on the time scale h^2. Their inner-product construction instead admits a second-order limit. A possible diffusive construction is to replace a fixed cosine weight $\exp(\nu \cos\theta)$ by $\exp(\nu_h \cos\theta)$ with $\nu_h/h \to \nu_0$, which could be explored in future work. FFM itself uses the transition moments directly, their interpretation as drift and diffusion.

First-order velocity-kernel limits: Suppose that for some time scale $\tau_h \to 0$, the discrete generators $L_h = (P_h - I)/\tau_h$ converge to

$$Lf = b^\top \nabla f + \tfrac{1}{2}\operatorname{Tr}(A\nabla^2 f).$$

Applying this convergence to coordinate and quadratic coordinate functions (as in 13) yields

$$m_h/\tau_h \to b, \qquad C_h/\tau_h \to A.$$

In particular, Li et al. (2021) show that cosine RNA-velocity kernels possess a first-order limit with $h = \sqrt{\varepsilon}$

$$(P_h - I)/h \to a(x)\widehat{v}(x)^\top \nabla,$$

with $a(x) > 0$. Consequently $m_h/h \to a(x)\widehat{v}(x)$, while $C_h/h \to 0$. Thus the first transition moment asymptotically recovers the RNA-velocity direction. For such kernels we therefore interpret C_h as a finite-scale measure of anisotropic transition accessibility rather than as an estimator of diffusion.

Extension of empirical drift and diffusion. The estimators in Proposition 13 are defined only at the observed states. To evaluate the induced metric along continuous curves, we extend them to a compact region $K \subset \mathbb{R}^d$ containing all admissible paths. Let K^+ be a compact neighborhood of K such that

$$K \subset \operatorname{int}(K^+).$$

We assume that Assumption 11 holds on K^+, so that Proposition 13 applies on this neighborhood.

Definition 14 (Uniform nodewise consistency)**.** Let $A \subset \mathbb{R}^d$ be compact and let f be a continuous vector- or matrix-valued field on A. Suppose that $\widehat{f}_N(x_i)$ estimates $f(x_i)$ at each observed state $x_i \in A$. We say that $\widehat{f}_N$ is uniformly nodewise consistent for f on A if

$$\max_{i:x_i \in A} \left\| \widehat{f}_N(x_i) - f(x_i) \right\| \longrightarrow 0$$

along the asymptotic regime under consideration. For matrix-valued functions, any matrix norm may be used, since all matrix norms are equivalent in finite dimension.

In particular, Proposition 13 shows that $\hat{b}_{0,N}$ and $\hat{D}_N$ are uniformly nodewise consistent for b_0 and D, respectively, on K^+.

Definition 15 (Kernel extension)**.** Let $h \in C_c^\infty([0,\infty))$ be nonnegative, supported in $[0,1]$, and strictly positive on a neighborhood of zero. Let $\eta_N \to 0$ and define

$$k_{\eta_N}(x,y) = h\left(\frac{\|x-y\|^2}{\eta_N^2}\right).$$

Define the fill distance of the sample cloud relative to K by

$$r_N(K) \coloneqq \sup_{x \in K} \min_{i:x_i \in K^+} \|x - x_i\|.$$

We assume that the samples become dense relative to the kernel bandwidth:

$$\frac{r_N(K)}{\eta_N} \longrightarrow 0.$$

Since h is positive near zero, this condition ensures that, for all sufficiently large N, every $x \in K$ receives positive kernel weight from at least one observed state in K^+. For such N, define

$$w_{i,N}(x) = \frac{k_{\eta_N}(x,x_i)}{\displaystyle\sum_{j:x_j \in K^+} k_{\eta_N}(x,x_j)}, \qquad x \in K, \quad x_i \in K^+,$$

and extend the nodewise estimators by

$$\widetilde{b}_{0,N}(x) = \sum_{i:x_i \in K^+} w_{i,N}(x)\hat{b}_{0,N}(x_i)$$

and

$$\widetilde{D}_N(x) = \sum_{i:x_i \in K^+} w_{i,N}(x)\hat{D}_N(x_i).$$

Consequently, for sufficiently large N, the extended fields are smooth on a neighborhood of K. Moreover, because every $\hat{D}_N(x_i)$ is positive semidefinite, $\widetilde{D}_N(x)$ is also positive semidefinite.

Proposition 16 (Uniform convergence of the kernel-extended estimators)**.** *Assume that b_0 and D are continuous on K^+ and that $\hat{b}_{0,N}$ and $\hat{D}_N$ are uniformly nodewise consistent on K^+ in the sense of Definition 14. If $\eta_N \longrightarrow 0$, $\frac{r_N(K)}{\eta_N} \longrightarrow 0$, then*

$$\sup_{x \in K} \left\| \widetilde{b}_{0,N}(x) - b_0(x) \right\| \longrightarrow 0 \text{ and } \sup_{x \in K} \left\| \widetilde{D}_N(x) - D(x) \right\| \longrightarrow 0.$$

Proof. Modulus of continuity for a continuos function f on K^+ is defined as:

$$\omega_f(r) \coloneqq \sup_{\substack{x,y \in K^+ \\ \|x-y\| \le r}} \|f(x) - f(y)\|.$$

Since K^+ is compact, f is uniformly continuous and therefore $\omega_f(r) \longrightarrow 0$ as $r \longrightarrow 0$. Consider first the drift field. Since the kernel weights are nonnegative and sum to one,

$$\left\|\widetilde{b}_{0,N}(x) - b_0(x)\right\| \leq \sum_{i:x_i \in K^+} w_{i,N}(x) \left\|\hat{b}_{0,N}(x_i) - b_0(x_i)\right\| + \sum_{i:x_i \in K^+} w_{i,N}(x) \left\|b_0(x_i) - b_0(x)\right\|.$$

Whenever $w_{i,N}(x) > 0$, the support condition on h implies

$$\|x_i - x\| \leq \eta_N.$$

Hence

$$\left\|\widetilde{b}_{0,N}(x) - b_0(x)\right\| \leq \max_{i:x_i \in K^+} \left\|\hat{b}_{0,N}(x_i) - b_0(x_i)\right\| + \omega_{b_0}(\eta_N).$$

Taking the supremum over $x \in K$, the first term converges to zero by uniform nodewise consistency, while the second converges to zero because $\eta_N \to 0$.

Applying the same argument to the matrix-valued field D gives

$$\sup_{x \in K} \left\|\widetilde{D}_N(x) - D(x)\right\| \leq \max_{i:x_i \in K^+} \left\|\hat{D}_N(x_i) - D(x_i)\right\| + \omega_D(\eta_N) \longrightarrow 0.$$

□

Freidlin–Wentzell Theory. Consider the small-noise diffusion

$$\mathrm{d}Y_T^\kappa = b_0(Y_T^\kappa)\,\mathrm{d}t + \sqrt{2\kappa}\, D(Y_T^\kappa)^{1/2}\,\mathrm{d}W_t,$$

where b_0 is smooth. For an absolutely continuous path $\phi : [0, T] \to \mathbb{R}^d$, the Freidlin–Wentzell (FW) (Freidlin & Wentzell, 2012) action is defined as:

$$S_T(\phi) = \frac{1}{4} \int_0^T (\dot{\phi}_t - b_0(\phi_t))^\top D(\phi_t)^{-1} (\dot{\phi}_t - b_0(\phi_t))\,\mathrm{d}t. \tag{16}$$

FW theory gives a path-space large-deviation principle with rate functional S_T. Informally speaking, among paths satisfying the same endpoint constraints, minimizers of S_T are the asymptotically most probable transition paths (See Theorem 4.1 Freidlin & Wentzell (2012)).

The fixed-time action S_T depends on the prescribed transition duration T. Minimizing over T yields the quasipotential

$$V(x_0, x_1) \coloneqq \inf_{T>0} \inf_{\phi:\phi(0)=x_0,\ \phi(T)=x_1} S_T(\phi).$$

By Proposition 2.1 and Equation (2.14) of Heymann & Vanden-Eijnden (2008),

$$V(x_0, x_1) = \inf_{T>0} \inf_{\phi:\phi(0)=x_0,\ \phi(T)=x_1} S_T(\phi) = \frac{1}{2} \inf_{\gamma:\gamma(0)=x_0,\ \gamma(1)=x_1} S_{\mathrm{FW}}(\gamma). \tag{17}$$

where $S_{\mathrm{FW}}(\gamma) = \int_0^1 F^{\mathrm{FW}}(\gamma(t), \dot{\gamma}(t))\mathrm{d}t$ with

$$F^{\mathrm{FW}}(x, v) \coloneqq \Big(v^\top \underbrace{\|b_0(x)\|^2_{D(x)^{-1}} D(x)^{-1}}_{G_{\mathrm{FW}}(x)} v\Big)^{1/2} + v^\top \underbrace{\big(-D(x)^{-1} b_0(x)\big)}_{\beta_{\mathrm{FW}}(x)}.$$

This has almost the form of a Finsler length functional 4, however, for $b_0(x) \neq 0$, $\|\beta_{\mathrm{FW}}(x)\|_{G_{\mathrm{FW}}(x)^{-1}} = 1$. Thus F^{FW} is a degenerate Randers-type structure rather than Finsler metric (See 5).

S_{FW} defines the parameterization-invariant FW action whose minimizers are the minimum-action transition curves. Let γ^* be the unique minimizer (up reparametrization) among curves connecting x_0 to x_1. We define the optimal transition time scale

$$T^* = \int_0^1 \frac{\mathrm{d}t}{\Lambda(\gamma^*(t), \dot{\gamma}^*(t))}, \qquad \text{where } \Lambda(x, v) = \sqrt{\frac{b_0(x)^\top D(x)^{-1} b_0(x)}{v^\top D(x)^{-1} v}},$$

and denote by $\tau_\delta(X^\kappa) = \inf\{t > 0 \mid Y_T^\kappa \in \bar{B}_\delta(x_1)\}$ the first hitting time of a closed δ-ball centered at x_1. Under the hypotheses of Heymann & Vanden-Eijnden (2008, Proposition 2.3), including uniqueness of the minimizer and of the corresponding constrained fixed-time minimizers, together with the stated compactness and convergence conditions, the following conditional concentration statement holds, for any error tolerance $\eta > 0$,

$$\lim_{\delta \to 0^+ T \to T^*} \liminf_{\kappa \to 0} \mathbb{P}_{x_0}\Big(\rho_F\Big(X^\kappa\big|_{[0,\tau_\delta(X^\kappa)]},\, \gamma^*\Big) < \eta \;\Big|\; \tau_\delta(X^\kappa) \le T\Big) = 1,$$

where ρ_F is the Fréchet distance between curves. Informally speaking, conditioned on the system completing the transition from x_0 into a small neighborhood of x_1 within a time frame close to T^*, the realization Y_T^κ concentrates with probability 1 in an arbitrarily narrow geometric tube around the minimizer γ^* as $\kappa \to 0$.

Empirical Finsler minimizers converge to geometric FW transition curves: We next show that minimizers of the regularized empirical Finsler action converge to the population geometric FW minimizer, which has the small-noise transition interpretation established above.

Assumption 17 (Compact path class)**.** Let $\Gamma = \{\gamma_\theta : \theta \in \Theta, \|\dot\gamma(t)\| \le M)\}$ be a class of smooth curves with $\gamma_\theta(0) = x_0$, $\gamma_\theta(1) = x_1$, Θ compact, and $(t, \theta) \mapsto (\gamma_\theta(t), \dot\gamma_\theta(t))$ continuous. Set $K_\Gamma = \overline{\bigcup_{\theta\in\Theta} \gamma_\theta([0,1])}$. We assume that $K_\Gamma \subset K$, where K is the compact set on which the uniform convergence in Proposition 16 holds.

For the unregularized geometric Freidlin–Wentzell interpretation in this subsection, we additionally assume that D is uniformly positive definite on the compact region K^+ containing the admissible paths. For numerical stability and to ensure uniform positive definiteness at finite N, we use the ridge-regularized estimator

$$\tilde D_{\rho_N}(x) \coloneqq \tilde D_N(x) + \rho_N I,$$

where $\rho_N > 0$ and $\rho_N \to 0$ as $N \to \infty$. Furthermore, we define the regularized FW metric as:

$$\tilde F^{\mathrm{FW}}_{N,\lambda}(x, v) \coloneqq \sqrt{v^\top \tilde D_{\rho_N}(x)^{-1} v}\, \sqrt{\|\tilde b_{0,N}(x)\|^2_{\tilde D_{\rho_N}(x)^{-1}} + \lambda_N^2} - v^\top \tilde D_{\rho_N}(x)^{-1} \tilde b_{0,N}(x).$$

, note that this ensures that $\tilde F^{\mathrm{FW}}_{N,\lambda}$ is a P-derived Randers 7 metric thus a Finsler structure and hence a for some sequence $\lambda_N > 0$ which converges to 0 as $N \to \infty$.

Theorem 18 (Consistency of regularized empirical geometric FW minimizers)**.** *Assume Assumption 17 and the assumptions of Proposition 16. Let $\rho_N > 0$ and $\lambda_N > 0$ satisfy $\rho_N \to 0, \lambda_N \to 0$. Define*

$$\tilde S_N(\gamma) \coloneqq \int_0^1 \tilde F^{\mathrm{FW}}_{N,\lambda_N}\left(\gamma(t), \dot\gamma(t)\right)\, \mathrm{d}t,$$

and let

$$\tilde\gamma_N^* \in \arg\min_{\gamma\in\Gamma} \tilde S_N(\gamma).$$

Then:

1. $$\sup_{\gamma\in\Gamma} \Big|\tilde S_N(\gamma) - S_{\mathrm{FW}}(\gamma)\Big| \longrightarrow 0.$$

2. *If $\mathcal{S}_{\mathrm{FW}}$ has a unique minimizer γ^* over Γ, up to reparameterization, then*
$$\rho_F\left(\tilde\gamma_N^*, \gamma^*\right) \longrightarrow 0.$$

Consequently, the regularized empirical Finsler minimizer converges to the population FW minimizer within Γ. If γ^* is also the global geometric FW minimizer connecting x_0 to x_1, and the hypotheses of Heymann & Vanden-Eijnden (2008, Proposition 2.3) hold, then γ^* is the asymptotically most probable transition curve in the small-noise sense stated above, and $\tilde\gamma_N^*$ converges to this curve.

Proof. By Proposition 16, we have uniform convergence on K_Γ:

$$\sup_{x\in K_\Gamma} \|\tilde{b}_{0,N}(x) - b_0(x)\| \to 0, \qquad \sup_{x\in K_\Gamma} \|\tilde{D}_N - D(x)\| \to 0.$$

Since $\rho_N \to 0$, we also have

$$\sup_{x\in K_\Gamma} \|\tilde{D}_{\rho_N}(x) - D(x)\| \le \sup_{x\in K_\Gamma} \|\tilde{D}_N(x) - D(x)\| + \rho_N \to 0.$$

Since D is uniformly positive definite on K_Γ, and since matrix inversion is continuous on the set of uniformly positive definite matrices, it follows that

$$\sup_{x\in K_\Gamma} \|\tilde{D}_{\rho_N}(x)^{-1} - D(x)^{-1}\| \to 0.$$

By Assumption 17, all curves in Γ have uniformly bounded velocity:

$$\|\dot\gamma(t)\| \le M \qquad \text{for all } \gamma \in \Gamma,\ t \in [0,1].$$

Therefore all pairs $(\gamma(t), \dot\gamma(t))$ lie in the compact set $K_\Gamma \times \overline{B_M(0)}$.

On this compact set, the map

$$(b, D, v, \lambda_N) \mapsto \sqrt{v^\top D^{-1} v}\, \sqrt{\|b\|^2_{D^{-1}} + \lambda_N^2} - v^\top D^{-1} b$$

is continuous. The uniform convergence of $\tilde{b}_{0,N}$, $\tilde{D}_{\rho_N}$ and $\lambda_N \to 0$ implies

$$\sup_{x\in K_\Gamma,\ \|v\|\le M} \left| \tilde{F}^{\mathrm{FW}}_{N,\lambda_N}(x,v) - F^{\mathrm{FW}}(x,v) \right| \to 0.$$

Consequently, for every $\gamma \in \Gamma$,

$$|\tilde{S}_N(\gamma) - S_{\mathrm{FW}}(\gamma)| \le \int_0^1 \left| \tilde{F}^{\mathrm{FW}}_{N,\lambda_N}(\gamma(t), \dot\gamma(t)) - F^{\mathrm{FW}}(\gamma(t), \dot\gamma(t)) \right| dt.$$

Taking the supremum over $\gamma \in \Gamma$, we obtain

$$\sup_{\gamma\in\Gamma} |\tilde{S}_N(\gamma) - S_{\mathrm{FW}}(\gamma)| \le \sup_{x\in K_\Gamma,\ \|v\|\le M} \left| \tilde{F}^{\mathrm{FW}}_{N,\lambda_N}(x,v) - F^{\mathrm{FW}}(x,v) \right| \to 0.$$

This proves the first claim. Let γ^* be the unique minimizer of S_{FW} over Γ, up to reparameterization. Fix $r > 0$, and define

$$\Gamma_r = \{\gamma \in \Gamma : \rho_F(\gamma, \gamma^*) \ge r\}.$$

Because Γ is compact and S is continuous on Γ, the minimum of S over Γ_r is attained. Since γ^* is the unique minimizer, there is a positive gap

$$\Delta_r \coloneqq \inf_{\gamma\in\Gamma_r} S_{\mathrm{FW}}(\gamma) - S_{\mathrm{FW}}(\gamma^*) > 0.$$

By the uniform convergence just proved, for sufficiently large N,

$$\sup_{\gamma\in\Gamma} |\tilde{S}_N(\gamma) - S_{\mathrm{FW}}(\gamma)| < \frac{\Delta_r}{3}.$$

For any $\gamma \in \Gamma_r$, this gives

$$\tilde{S}_N(\gamma) \ge S_{\mathrm{FW}}(\gamma) - \frac{\Delta_r}{3} \ge S_{\mathrm{FW}}(\gamma^*) + \Delta_r - \frac{\Delta_r}{3} = S_{\mathrm{FW}}(\gamma^*) + \frac{2\Delta_r}{3}.$$

On the other hand,

$$\tilde{S}_N(\gamma^*) \le S_{\mathrm{FW}}(\gamma^*) + \frac{\Delta_r}{3}.$$

Therefore

$$\tilde{S}_N(\gamma) > \tilde{S}_N(\gamma^*) \qquad \text{for all } \gamma \in \Gamma_r.$$

Hence no minimizer $\tilde\gamma^* \in \arg\min_{\gamma\in\Gamma} \tilde{S}_N(\gamma)$ can lie in Γ_r. Thus,

$$\rho_F(\tilde\gamma^*, \gamma^*) < r$$

for all sufficiently large N. Since $r > 0$ was arbitrary,

$$\rho_F(\tilde{\gamma}^*, \gamma^*) \to 0.$$

Finally, suppose that γ^* is also the global geometric FW minimizer connecting x_0 to x_1, and that the hypotheses of Heymann & Vanden-Eijnden (2008, Proposition 2.3) hold. The conditional concentration result above then identifies γ^* as the asymptotically most probable transition curve in the small-noise limit. Since

$$\rho_F\left(\tilde{\gamma}_N^*, \gamma^*\right) \longrightarrow 0,$$

the regularized empirical Finsler minimizer converges to this transition curve. □

Remark 19 (Fixed-ridge geometry). The positive-definiteness assumption above is needed only for passing to the unregularized limit $\rho_N \to 0$ and recovering the geometric FW action of the original diffusion. For any fixed $\bar{\rho} > 0$, the regularized population tensor $D_{\bar{\rho}}(x) := D(x) + \bar{\rho} I$ satisfies is unifromly positive definite. Hence, Proposition 16 implies

$$\sup_{x \in K} \left\| (\widetilde{D}_N(x) + \bar{\rho} I)^{-1} - (D(x) + \bar{\rho} I)^{-1} \right\| \longrightarrow 0.$$

Consequently, for fixed $\bar{\rho} > 0$ and $\bar{\lambda} > 0$, the same argument as in Theorem 18 gives uniform convergence of the empirical regularized action to the corresponding population regularized action, without requiring D itself to be uniformly positive definite.

Remark 20. For a common graph time step $\Delta t > 0$, the diffusion-normalized moments are $\widehat{b} = m/\Delta t$ and $\widehat{D} = C/\Delta t$. Writing $F_{\rho,\lambda}[m, C]$ for the regularized metric constructed from the moments m, C, direct substitution gives

$$F_{\rho,\lambda}[m, C](x, v) = F_{\rho/\Delta t,\, \lambda/\sqrt{\Delta t}}\left[\frac{m}{\Delta t}, \frac{C}{\Delta t}\right](x, v).$$

Thus, raw and diffusion-normalized moments yield exactly the same metric after rescaling the regularization parameters, and we absorb the time-step normalization into these hyperparameters in the implementation. The vanishing-regularization assumptions in the consistency theorem apply to the normalized parameters; in raw-moment notation, they require $\rho_N/\Delta t_N \to 0$ and $\lambda_N/\sqrt{\Delta t_N} \to 0$.

Additional Calculation:

Proposition 21 (Gaussian probability path)**.** *Let $\mu_t \in \mathbb{R}^d$ be continuously differentiable and let $G_t \in \mathrm{SPD}(d)$ be a continuously differentiable. Define, for $t \in (0, 1)$, $p_t(x \mid x_0, x_1) = \mathcal{N}(x; \mu_t,\, t(1-t)G_t)$. Then the vector field*

$$\tilde{u}_t(x \mid x_0, x_1) = \dot{\mu}_t + \frac{1 - 2t}{2t(1-t)}(x - \mu_t) + \frac{1}{2}\dot{G}_t G_t^{-1}(x - \mu_t)$$

satisfies $\frac{dx}{dt} = \tilde{u}_t(x_t \mid x_0, x_1)$ and $x_t \sim p_t(\cdot \mid x_0, x_1)$ for all $t \in (0, 1)$.

Proof. Let B_t be a differentiable matrix factor satisfying

$$B_t B_t^\top = G_t, \qquad \dot{B}_t B_t^{-1} = \frac{1}{2}\dot{G}_t G_t^{-1},$$

and let $\varepsilon \sim \mathcal{N}(0, I)$. Such a metric exists since G_t is symmetric positive definite (SPD) and by using existence and uniqueness of ODEs. A sample from p_t can then be written as

$$x_t = \mu_t + \sqrt{t(1-t)}\, B_t \varepsilon.$$

Solving for ε gives

$$\varepsilon = \frac{1}{\sqrt{t(1-t)}} B_t^{-1}(x_t - \mu_t).$$

Differentiating x_t with respect to t yields

$$\begin{aligned}\dot{x}_t &= \dot{\mu}_t + \frac{1-2t}{2\sqrt{t(1-t)}} B_t\varepsilon + \sqrt{t(1-t)}\,\dot{B}_t\varepsilon \\ &= \dot{\mu}_t + \frac{1-2t}{2t(1-t)}(x_t-\mu_t) + \dot{B}_t B_t^{-1}(x_t-\mu_t) \\ &= \dot{\mu}_t + \frac{1-2t}{2t(1-t)}(x_t-\mu_t) + \frac{1}{2}\dot{G}_t G_t^{-1}(x_t-\mu_t).\end{aligned}$$

Hence $\dot{x}_t = u_t(x_t)$. Since $x_t = \mu_t + \sqrt{t(1-t)}B_t\varepsilon$ has distribution $\mathcal{N}(\mu_t, t(1-t)G_t)$ by construction, the flow generated by u_t has marginals p_t for all $t \in (0,1)$. □

Remark 22. Note that in the case where $G_t = \mathrm{Id}$ and $\mu_t = tx_1 + (1-t)x_0$we recover the formula from Tong et al. (2024b).

C EXPERIMENTS

Experimental protocol: All models get the same networks that uses four hidden layers of width 64, SiLU activations, and a 32-dimensional sinusoidal time embedding. We use Adam with learning rate 2×10^{-3}, batch size 256, and gradient clipping at 5. Where applicable, geodesic training uses 2,500 iterations, followed by 1,500 flow-training iterations, stochastic training uses an additional 1,500 iterations. Default coupling settings use eight quadrature samples and entropic regularization equal to 0.1 times the median transport cost, with benchmark-specific overrides stated separately. Results report mean ± population standard deviation across three seeds, except Erythroid, which uses five. Hyperparameters were selected using a validation split of 15 percent and using the mean of mid and end W_2.

Kernel extension of transition moments: The first and second transition moments $m(x_i)$ and $C(x_i)$ are defined only at the observed states $\{x_i\}_{i=1}^N$. To evaluate the Finsler metric along continuous paths, we extend both moments using normalized non-negative kernel smoothing. Let $k_h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ be a smooth kernel with bandwidth $h > 0$. Assuming that $\sum_{\ell=1}^N k_h(x, x_\ell) > 0$ throughout the region of interest, define

$$w_i(x) \coloneqq \frac{k_h(x, x_i)}{\sum_{\ell=1}^N k_h(x, x_\ell)}.$$

The extended moment fields are

$$m(x) \coloneqq \sum_{i=1}^N w_i(x) m(x_i), \qquad C(x) \coloneqq \sum_{i=1}^N w_i(x) C(x_i). \tag{18}$$

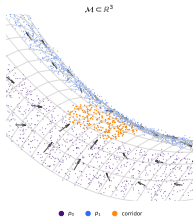
We choose kernel smoothing for its simplicity: it requires no additional model training and preserves the positive semidefiniteness of the second-moment field. Indeed, the weights are non-negative and sum to one, so $C(x)$ is a convex combination of positive semidefinite matrices. Both extended fields are smooth wherever the denominator is positive. Adding the ridge term $C_\rho(x) = C(x) + \rho I$, with $\rho > 0$, then ensures positive definiteness for the subsequent metric construction.

Sheet Dataset: We sample $N = 4000$ points from coordinates $u \sim \mathcal{U}(-1.5, 1.5)$ and $w \sim \mathcal{U}(-0.45, 0.45)$ and embed them on the saddle

$$z(u,w) = \frac{1}{2}\left[\left(\frac{3u}{2}\right)^2 - \frac{1}{2}\left(\frac{100w}{45}\right)^2\right],$$

with Gaussian noise of standard deviation 0.012. The prescribed transition probability matrix P forms an hourglass: trajectories are driven towards $u \approx 0$ and go outward after crossing. Source and target states are defined by $w \leq -0.13$ and $w \geq 0.13$, respectively.

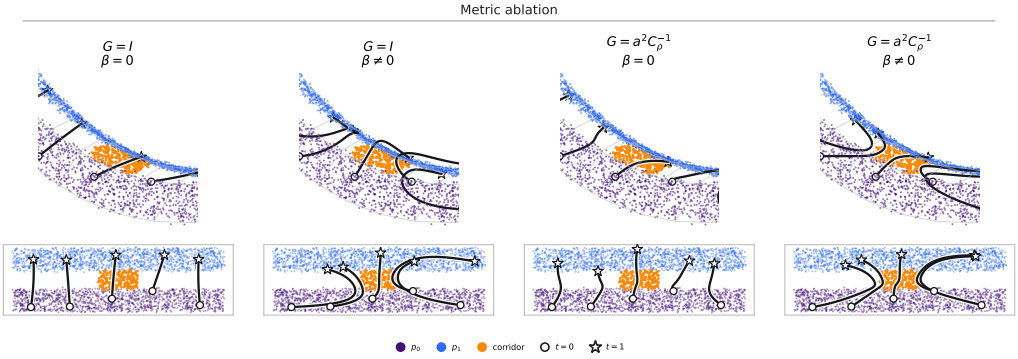
P is constructed bye directed 15-NN kernel combining spatial proximity with alignment to a



(a) Sheet benchmark.

G	Drift	c_{eff}	mid $W_2 \downarrow$	end $W_2 \downarrow$	$W_{2,F} \downarrow$
I	no	0.001	0.7165 ± 0.0136	$\mathbf{0.1751 \pm 0.0271}$	1.0840 ± 0.0371
I	yes	0.958	0.2541 ± 0.0375	0.2558 ± 0.0509	0.3676 ± 0.0773
C_ρ^{-1}	no	0.001	0.6672 ± 0.0160	0.2025 ± 0.0079	0.9437 ± 0.1068
C_ρ^{-1}	yes	0.958	$\mathbf{0.1588 \pm 0.0096}$	0.2343 ± 0.0292	$\mathbf{0.2892 \pm 0.0090}$
sampling floor			*0.0646*	*0.1896*	*0.2812*

(b) Metric ablation.



(c) Trajectories generated by the different methods.

Figure 4: **Sheet benchmark and additional results. (a)** Overview of the synthetic Sheet benchmark and its directed transition structure. **(b)** Ablation of the directional and anisotropic components of the Finsler metric. Directionality substantially improves intermediate recovery, with the full metric obtaining the lowest midpoint errors. **(c)** Qualitative comparison of generated trajectories.

Table 2: Wall-clock time in seconds per seed on a Tesla V100-SXM2-32GB on the Pancreas dataset. The stochastic FFM arm reuses the deterministic interpolant and coupling.

Method	$d = 2$	$d = 10$	$d = 20$	$d = 50$
OT-CFM	3	3	3	3
MFM / LAND	22	61	109	254
Curly-FM	41	106	190	445
FFM (ours), deterministic	26	63	109	256
FFM (ours), with noise	15	29	48	108

prescribed vector field ($\lambda = 6$), followed by row normalization. Source and target points are split $70/15/15$ into train/validation/test sets, whereas the withheld region is divided equally between validation and test. Hyperparameters are selected using the average of endpoint W_2 and intermediate-region W_2 on the validation set. We also perform a metric ablation where we compare only keeping the first and second moments 4b.

Pancreas Dataset: Hyperparameters are selected independently for each representation using a validation subset disjoint from the reported withheld cells. Runtime and sensitivity to ρ and λ are reported in **??** and 3. We additionally evaluate sensitivity to temporal alignment, the amount of intermediate-state support used to construct the geometry, and choices in the stochastic construction. To reduce dependence on the benchmark convention $t = 1/2$, we evaluate W_2 over a fixed 11-point grid spanning $t \in [1/3, 2/3]$ and report $\overline{W}_2^{[1/3,2/3]}$, the mean over this grid. We further report $W_{2,F}$ across representations, vary the fraction of intermediate-band cells available for metric construction, and ablate the conditional-path covariance and inference noise used by stochastic FFM. Unless otherwise stated, all Pancreas results in 3 use no intermediate-band cells for metric construction. 5 is an additional sensitivity analysis that deliberately varies this fraction.

(a) Sensitivity to model time

Method	$W_2(1/3)$	$W_2(0.433)$	$W_2(0.567)$	$W_2(2/3)$	$\overline{W}_2^{[1/3,2/3]}$
$d = 2$					
FFM, deterministic	1.9878 ± 0.0631	1.8184 ± 0.0326	1.4429 ± 0.0195	1.1770 ± 0.0919	1.6175 ± 0.0348
FFM, with noise	2.0052 ± 0.0705	1.8399 ± 0.0401	1.4723 ± 0.0601	1.2148 ± 0.0851	1.6434 ± 0.0555
OT-CFM	1.8508 ± 0.0172	1.7952 ± 0.0182	1.7607 ± 0.0195	1.7682 ± 0.0201	1.7869 ± 0.0230
MFM / LAND	1.5201 ± 0.0119	1.4878 ± 0.0117	1.4438 ± 0.0108	1.4002 ± 0.0098	1.4642 ± 0.0131
Curly-FM	1.0208 ± 0.1277	0.9952 ± 0.1378	0.9480 ± 0.1457	0.9009 ± 0.1505	0.9686 ± 0.1725
$d = 20$					
FFM, deterministic	11.5969 ± 0.1999	10.7814 ± 0.1811	10.3610 ± 0.1440	10.8331 ± 0.1471	10.7413 ± 0.1927
FFM, with noise	11.5392 ± 0.1640	10.7188 ± 0.1270	10.3671 ± 0.0477	10.9998 ± 0.2163	10.7355 ± 0.0560
OT-CFM	13.0727 ± 0.1325	12.5476 ± 0.0959	12.2577 ± 0.0231	12.3903 ± 0.0406	12.4957 ± 0.0681
MFM / LAND	12.8496 ± 0.1074	12.3470 ± 0.0706	11.8851 ± 0.0180	11.7331 ± 0.0277	12.1656 ± 0.0522
Curly-FM	12.8636 ± 0.0503	12.3801 ± 0.0478	12.2108 ± 0.0760	12.4096 ± 0.1064	12.3920 ± 0.0687
$d = 50$					
FFM, deterministic	13.2623 ± 0.1442	12.7423 ± 0.0966	12.5789 ± 0.0240	12.9036 ± 0.0495	12.7715 ± 0.0719
FFM, with noise	13.2718 ± 0.1524	12.7538 ± 0.1104	12.5768 ± 0.0639	12.8742 ± 0.0975	12.7723 ± 0.0952
OT-CFM	14.1359 ± 0.1441	13.6933 ± 0.1242	13.4941 ± 0.0923	13.6688 ± 0.0816	13.6811 ± 0.1314
MFM / LAND	13.8332 ± 0.1526	13.2557 ± 0.1216	12.7979 ± 0.0593	12.7353 ± 0.0364	13.0996 ± 0.1029
Curly-FM	13.8395 ± 0.1113	13.4287 ± 0.1521	13.2923 ± 0.2020	13.4719 ± 0.2307	13.4440 ± 0.2069

(b) Finsler-Wasserstein recovery

Method	$d = 2$	$d = 10$	$d = 20$	$d = 50$
OT-CFM	0.6420 ± 0.0059	10.9311 ± 0.0770	10.9277 ± 0.0412	15.5900 ± 0.2287
MFM / LAND	$\mathbf{0.5399 \pm 0.0024}$	$\mathbf{9.4377 \pm 0.0824}$	10.4768 ± 0.0637	14.6185 ± 0.1085
Curly-FM	0.7493 ± 0.0258	10.7482 ± 0.1305	11.0235 ± 0.0520	15.0427 ± 0.1562
FFM, deterministic	0.5681 ± 0.0100	9.5125 ± 0.5398	$\mathbf{8.9033 \pm 0.0400}$	13.9368 ± 0.1217
FFM, with noise	0.5730 ± 0.0082	9.6113 ± 0.5379	8.9213 ± 0.0453	$\mathbf{13.7800 \pm 0.0861}$
floor	*0.4162*	*5.3461*	*4.3608*	*7.2123*

$W_{2,F}(t = 1/2) \downarrow$.

(c) Intermediate support used for metric construction

Method	0%	5%	35%
$d = 2$			
OT-CFM	1.7607 ± 0.0262	1.7622 ± 0.0392	1.7666 ± 0.0297
FFM, deterministic	1.3833 ± 0.1410	1.3313 ± 0.0627	1.0927 ± 0.2477
FFM, with noise	1.3723 ± 0.1505	1.3525 ± 0.0420	1.1013 ± 0.2611
MFM / LAND	1.4506 ± 0.0025	1.2412 ± 0.5283	0.6878 ± 0.0378
Curly-FM	0.9751 ± 0.1704	0.9360 ± 0.1045	1.5865 ± 0.1369
$d = 20$			
OT-CFM	12.3291 ± 0.0685	12.3264 ± 0.0445	12.3491 ± 0.0850
FFM, deterministic	10.4208 ± 0.1028	8.2742 ± 0.3540	8.1989 ± 0.0481
FFM, with noise	10.3679 ± 0.0641	8.4860 ± 0.3098	8.2513 ± 0.2160
MFM / LAND	12.0368 ± 0.0696	11.8784 ± 0.0338	11.7247 ± 0.1739
Curly-FM	12.2067 ± 0.0844	12.2163 ± 0.1098	12.2431 ± 0.0665
$d = 50$			
OT-CFM	13.5319 ± 0.0480	13.5269 ± 0.0642	13.5577 ± 0.1075
FFM, deterministic	$\mathbf{12.5043 \pm 0.0544}$	$\mathbf{12.1225 \pm 0.1252}$	$\mathbf{11.3633 \pm 0.1356}$
MFM / LAND	13.5584 ± 0.0396	13.5208 ± 0.0275	13.5689 ± 0.1239
Curly-FM	13.3220 ± 0.1927	13.3291 ± 0.2501	13.2897 ± 0.1116

$W_2(t = 1/2) \downarrow$.

(d) Conditional-path noise width and covariance

Width mult.	$M_t = C_\rho(\mu_t)\, W_2(1/2)$	$M_t = C_\rho(\mu_t)\, W_2(1)$	$M_t = I\, W_2(1/2)$	$M_t = I\, W_2(1)$
0.003	10.4874 ± 0.0419	5.2457 ± 0.2301	10.4414 ± 0.1439	5.2980 ± 0.1595
0.010	10.4940 ± 0.1123	5.5062 ± 0.4344	10.5187 ± 0.1591	5.2420 ± 0.2684
0.030	10.4873 ± 0.0919	5.3305 ± 0.1619	10.4245 ± 0.1359	5.3180 ± 0.1351
0.100	10.3694 ± 0.0987	5.0662 ± 0.3449	10.3446 ± 0.1299	5.1521 ± 0.1702
0.300	10.4630 ± 0.2142	5.3692 ± 0.1590	10.4149 ± 0.1922	5.2433 ± 0.2451
1.000	14.0089 ± 0.0636	7.6047 ± 0.3015	13.8565 ± 0.1498	7.3363 ± 0.1651
3.000	36.3654 ± 0.5652	24.9668 ± 0.7730	33.9906 ± 1.1181	26.6449 ± 1.1385
10.000	118.0193 ± 4.7889	85.1409 ± 11.5354	143.4577 ± 2.9533	130.1162 ± 4.8789

(e) Path and inference covariance

Path	Inference	$W_2(1/2)$	$W_2(1)$
$C_\rho(\mu_t)$	$C_\rho(X_t)$	10.3694 ± 0.0987	5.0662 ± 0.3449
$C_\rho(\mu_t)$	I	10.3617 ± 0.1003	$\mathbf{4.9788 \pm 0.2102}$
I	$C_\rho(X_t)$	10.3557 ± 0.1112	5.1712 ± 0.2807
I	I	$\mathbf{10.3446 \pm 0.1299}$	5.1521 ± 0.1702

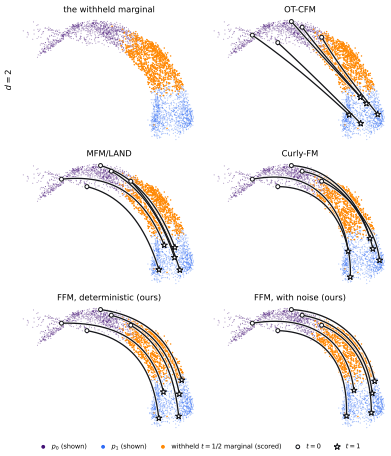
(f) Inference-noise schedule

$g(t)$	$M(X_t) = C_\rho(X_t)\, W_2(1/2)$	$M(X_t) = C_\rho(X_t)\, W_2(1)$	$M(X_t) = I\, W_2(1/2)$	$M(X_t) = I\, W_2(1)$
σ	10.3694 ± 0.0987	5.0662 ± 0.3449	10.3617 ± 0.1003	4.9788 ± 0.2102
$\propto (1 - t)$	10.3580	5.0072	10.3496	5.0492
$g \sim t$	10.3922	5.1977	10.3868	5.1169
$g \sim \sqrt{t(1-t)}$	10.3629	5.0814	10.3515	5.0035
0 after $t = 0.8$	10.3626	5.0440	10.3543	4.9913
0	10.4173	5.1955	10.4173	5.1955

Figure 5: **Pancreas robustness analyses.** (a) Intermediate-population recovery over model time. We show representative points from the fixed 11-point grid over $[1/3, 2/3]$ together with $\overline{W}_2^{[1/3,2/3]}$, the mean over all 11 points. (b) Geometry-dependent Wasserstein distance across representations. (c) Sensitivity to the fraction of intermediate-band cells made available for metric construction. Because the scored population changes with the support fraction, comparisons across columns should be interpreted descriptively. Methods within each column use the same scored cells. (d–f) Sensitivity of stochastic FFM at $d = 20$ to conditional-path noise width, conditional-path and inference covariance, and the scalar inference-noise schedule. Values are mean $\pm$ population standard deviation over three seeds unless stated otherwise. Lower is better.

Table 3: One-at-a-time sensitivity of FFM on Pancreas. The reported value is the validation selection objective; lower is better. The selected value is shown in bold.

d	Varied parameter	Fixed parameter	Values	Validation objective
2	ρ	$\lambda = 0.1$	0.01, 0.03, **0.1**, 0.3, 1.0	0.828, 0.749, **0.535**, 0.870, 1.131
2	λ	$\rho = 0.1$	0.01, 0.03, **0.1**, 0.3, 1.0	1.226, 0.951, **0.535**, 0.826, 0.873
20	ρ	$\lambda = 1.0$	0.003, 0.01, **0.03**, 0.1, 0.3	8.180, 8.210, **8.135**, 8.369, 8.288
20	λ	$\rho = 0.03$	0.03, 0.1, 0.3, **1.0**, 3.0	8.990, 8.535, 8.370, **8.135**, 8.667
50	ρ	$\lambda = 0.03$	3×10^{-4}, 10^{-3}, $\mathbf{3\times10^{-3}}$, 10^{-2}, 3×10^{-2}	10.455, 10.421, **10.314**, 10.390, 10.589
50	λ	$\rho = 0.003$	0.003, 0.01, **0.03**, 0.1, 0.3	10.379, 10.314, **10.314**, 10.554, 10.537



Method	Source	Cos. Dist. ↓	L_2 ↓	W_1 ↓	W_2 ↓
$d = 2$					
OT-CFM	Published	0.146 ± 0.001	2.704 ± 0.019	0.646 ± 0.006	–
OT-CFM	Reproduced	0.141 ± 0.002	2.447 ± 0.031	0.645 ± 0.009	0.679 ± 0.009
MFM (LAND)	Published	0.014 ± 0.001	1.999 ± 0.014	0.269 ± 0.004	–
MFM (LAND)	Reproduced	$\mathbf{0.010 \pm 0.001}$	1.858 ± 0.025	0.271 ± 0.007	0.286 ± 0.007
Curly-FM	Published	0.009 ± 0.000	1.663 ± 0.293	0.369 ± 0.090	–
Curly-FM	Reproduced	0.011 ± 0.001	2.151 ± 0.078	$\mathbf{0.122 \pm 0.002}$	$\mathbf{0.158 \pm 0.003}$
FFM (ours)	Ours	0.021 ± 0.004	$\mathbf{1.642 \pm 0.071}$	0.248 ± 0.034	0.291 ± 0.023
FFM (ours), with noise	Ours	0.019 ± 0.003	1.710 ± 0.081	0.226 ± 0.027	0.269 ± 0.019
$d = 20$					
OT-CFM	Published	0.489 ± 0.001	1.885 ± 0.020	6.103 ± 0.074	–
OT-CFM	Reproduced	$\mathbf{0.490 \pm 0.000}$	1.731 ± 0.014	6.106 ± 0.063	6.260 ± 0.066
MFM (LAND)	Published	0.495 ± 0.001	1.627 ± 0.040	4.855 ± 0.052	–
MFM (LAND)	Reproduced	0.494 ± 0.000	1.700 ± 0.021	4.759 ± 0.056	4.855 ± 0.054
Curly-FM	Published	0.488 ± 0.001	1.721 ± 0.035	6.124 ± 0.027	–
Curly-FM	Reproduced	$\mathbf{0.490 \pm 0.000}$	1.691 ± 0.011	6.112 ± 0.053	6.271 ± 0.058
FFM (ours)	Ours	0.491 ± 0.000	1.514 ± 0.016	$\mathbf{4.685 \pm 0.012}$	$\mathbf{4.782 \pm 0.014}$
FFM (ours), with noise	Ours	0.496 ± 0.000	$\mathbf{0.757 \pm 0.006}$	4.733 ± 0.035	4.823 ± 0.038
$d = 50$					
OT-CFM	Published	0.490 ± 0.000	2.215 ± 0.022	7.969 ± 0.029	–
OT-CFM	Reproduced	$\mathbf{0.491 \pm 0.000}$	2.053 ± 0.017	7.839 ± 0.016	7.974 ± 0.016
MFM (LAND)	Published	0.494 ± 0.000	1.971 ± 0.023	6.727 ± 0.022	–
MFM (LAND)	Reproduced	0.495 ± 0.000	2.004 ± 0.013	6.697 ± 0.052	6.819 ± 0.058
Curly-FM	Published	0.489 ± 0.000	2.045 ± 0.073	7.729 ± 0.046	–
Curly-FM	Reproduced	$\mathbf{0.491 \pm 0.000}$	1.990 ± 0.021	7.736 ± 0.025	7.864 ± 0.025
FFM (ours)	Ours	$\mathbf{0.491 \pm 0.000}$	1.772 ± 0.029	6.810 ± 0.022	6.907 ± 0.025
FFM (ours), with noise	Ours	0.499 ± 0.000	$\mathbf{0.915 \pm 0.006}$	$\mathbf{6.595 \pm 0.016}$	$\mathbf{6.663 \pm 0.015}$

Figure 6: Visualization of the learned erythroid trajectories in 2D and quantitative comparison across dimensions $d = 2, 20, 50$. "Published" denotes the values reported in Table 4 of Petrović et al. (2025), while "Reproduced" denotes our reproduction of the corresponding baselines using a common training harness at the published training budget. Due to a bug in the original Curly-FM evaluation, the Wasserstein quantity reported as W_2 in Table 4 of CurlyFM corresponds in fact to W_1. We therefore report these published values under W_1 and additionally compute the true W_2 distance for all runs evaluated with our harness. For $d = 20$ and $d = 50$, L_2 is reported in units of 10^3. Bold values indicate the best performance among methods evaluated with our common harness; published values are shown for reference and are not included when determining the best result.

Erythroid: We further evaluate FFM on the mouse erythroid differentiation dataset used by Petrović et al. (2025). The dataset consists of 9,815 cells along the erythroid developmental trajectory, partitioned into three temporal snapshots with the intermediate population withheld during training. We evaluate in the two-dimensional embedding and in $d \in \{20, 50\}$ dimensional representations. In addition to our methods, we report both the values originally published by Petrović et al. (2025) and our reproduction of the baselines using a common training harness at their published training budget.

iTracer: Let $\mathcal{N}_{15}(i)$ denote the 15 nearest neighbours of cell i, excluding itself, and let $j_7(i)$ denote its seventh nearest neighbour. All four kernel variants share the same directed, unsymmetrized support and self-tuning bandwidths. The expression factor is

$$K_{ij}^{\text{expr}} = \mathbf{1}_{\{j \in \mathcal{N}_{15}(i)\}} \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma_i \sigma_j}\right), \qquad \sigma_i = \|x_i - x_{j_7(i)}\|. \tag{19}$$

For barcode-family labels bf_i and scar-family labels sf_i, define

$$R_{ij} = 1 - \frac{1}{2}\left(\mathbf{1}_{\{\mathrm{bf}_i \neq \mathrm{bf}_j\}} + \mathbf{1}_{\{\mathrm{sf}_i \neq \mathrm{sf}_j\}}\right), \qquad K_{ij}^{\text{lin}} = \exp(\beta_{\text{lin}} R_{ij}). \tag{20}$$

Pairs involving a cell without a barcode call receive $R_{ij} = 0$. Barcode and scar families containing fewer than 100 and 50 cells, respectively, are excluded. For the RNA-velocity vector v_i, the velocity

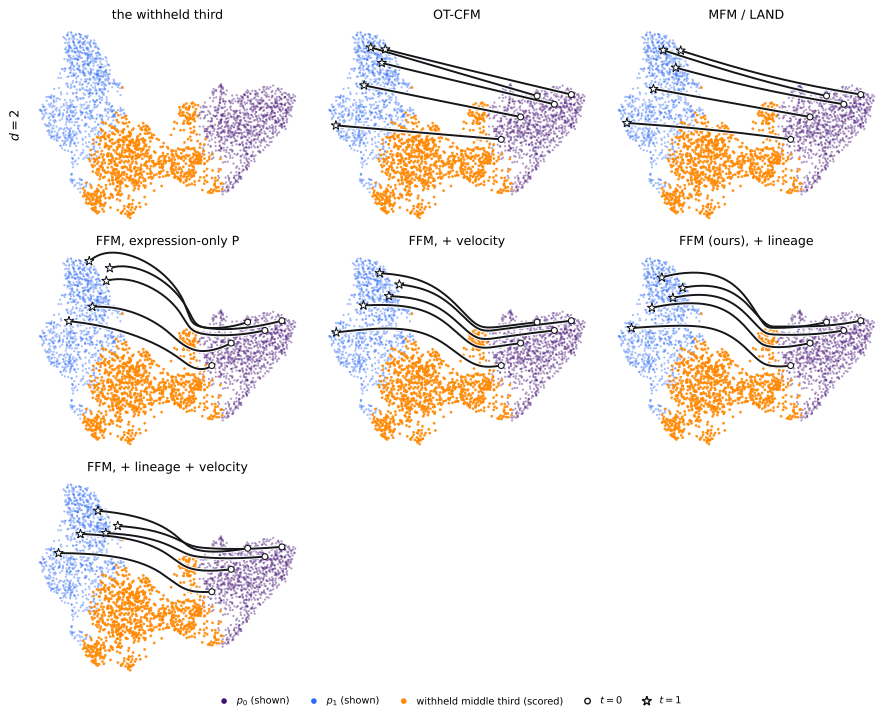


Figure 7: Learned trajectories on iTracer in the two-dimensional representation.

factor is

$$a_{ij} = \frac{\langle v_i, x_j - x_i \rangle}{\|v_i\| \, \|x_j - x_i\|}, \qquad K_{ij}^{\mathrm{vel}} = \exp(\nu a_{ij}). \tag{21}$$

The final transition matrix is

$$W_{ij} = K_{ij}^{\mathrm{expr}} K_{ij}^{\mathrm{lin}} K_{ij}^{\mathrm{vel}}, \qquad P_{ij} = \frac{W_{ij}}{\sum_{\ell} W_{i\ell}}, \qquad P_{ii} = 0. \tag{22}$$

The different choices

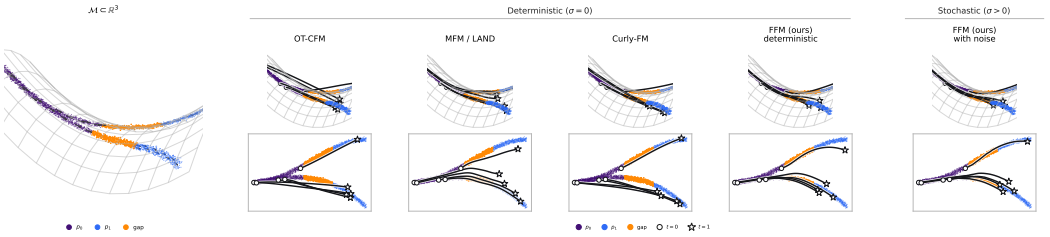
$$(\beta_{\mathrm{lin}}, \nu) \in \{(0,0), (0,3), (3,0), (3,3)\}$$

correspond to expression only, expression plus velocity, expression plus lineage, and all three factors, respectively.

Bifurcation: We evaluate a synthetic bifurcation comprising 3000 points in $\mathbb{R}^3$, where one source population splits into two target arms and the middle split is withheld 8. Hyperparameters are transferred from Sheet without retuning. To evaluate cell-level fate preservation, we integrate the 324 of 969 source cells with resolved initial arm labels and classify their endpoints as lower arm, upper arm, or outside. The target regions are constructed from balls of radius $1/2$ of the distance between arm centres Table 4. FFM maintains high recall on both arms, whereas OT-CFM and Curly-FM misassign lower-arm cells, and MFM loses upper-arm recall primarily through outside outcomes. A separate allocation sweep 5 shows that performance under imbalance remains configuration-dependent.

D Related Work

Relation to Metric Flow Matching. Metric Flow Matching (MFM) (Kapusniak et al., 2024) learns endpoint-conditioned approximate geodesics of a data-dependent Riemannian metric $g(x) = G(x; \mathcal{D})$ (Definition 1) as interpolants. This construction is contained in FFM as the symmetric special case in which no directional information is used. More precisely $G(x; \mathcal{D})$ is a



(a) Dataset. (b) Learned trajectories.

Figure 8: ForkSheet benchmark.

Table 4: ForkSheet fate preservation. Values are percentages, reported as mean ± population standard deviation over three seeds.

Method	True arm	Recall	Wrong arm	Outside
OT-CFM	Lower	69.4 ± 10.7	30.6 ± 10.7	0.0 ± 0.0
	Upper	100.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
MFM / LAND	Lower	91.4 ± 5.6	4.4 ± 3.9	4.2 ± 3.4
	Upper	68.1 ± 37.3	12.3 ± 17.4	19.6 ± 20.1
Curly-FM	Lower	48.8 ± 16.7	47.4 ± 15.2	3.8 ± 1.5
	Upper	100.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
FFM (ours), deterministic	Lower	75.5 ± 19.7	21.2 ± 19.4	3.4 ± 0.3
	Upper	100.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
FFM (ours), with noise	Lower	68.3 ± 19.8	20.8 ± 22.7	10.9 ± 12.7
	Upper	98.8 ± 0.9	0.0 ± 0.0	1.2 ± 0.9

special case of a $P-$derived Randers with $F_P(x, v) = \sqrt{v^T G(x; \mathcal{D}) v}$.

There is also an explicit connection between the LAND metric (Arvanitidis et al., 2016) considered in MFM and our P-derived metric (9). Consider a soft k-NN kernel

$$w_{ij} = \mathbf{1}_{\{j \in \mathcal{N}_k(i)\}} \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \qquad P_{ij} = \frac{w_{ij}}{Z_i}, \qquad Z_i = \sum_j w_{ij}.$$

The corresponding second transition moment is

$$C_N(x_i) = \frac{1}{Z_i} \sum_j w_{ij}(x_j - x_i)(x_j - x_i)^\top.$$

On the other hand, defining $h_\alpha(x_i) = \sum_j w_{ij}(x_j^\alpha - x_i^\alpha)^2$, we obtain

$$\operatorname{diag} C_N(x_i) = \frac{1}{Z_i} \operatorname{diag}(h(x_i)). \tag{23}$$

Consequently,

$$(\operatorname{diag} C_N(x_i) + \rho I)^{-1} = Z_i \left(\operatorname{diag}(h(x_i)) + \rho Z_i I\right)^{-1}. \tag{24}$$

Thus, apart from the normalization Z_i induced by requiring P to be row-stochastic, the diagonal of our second-moment geometry has exactly the form of the LAND metric on datapoints. In particular, when Z_i is approximately constant over the region of interest, choosing $\varepsilon = \rho Z$ recovers the LAND metric up to the global multiplicative factor Z, which does not change its geodesic curves.

FFM can therefore be viewed as a generalization of MFM incoporating a sense of directionality when given an transition probability matrix P.

Relation to Curly FM: Curly Flow Matching (Curly-FM) (Petrović et al., 2025) assumes access to a reference vector field $f_t(x)$ and learns endpoint conditioned paths whose velocities remain close

Table 5: ForkSheet recall under different nominal lower/upper fate allocations. Values are percentages, reported as mean ± population standard deviation over three seeds. Within each cell, the first line reports lower-arm recall and the second line reports upper-arm recall. These runs are separate from the evaluation in Table 4.

Method	50/50	70/30	40/60	20/80	10/90
OT-CFM	69.4 ± 10.7	71.1 ± 2.5	57.6 ± 11.7	59.0 ± 17.4	36.1 ± 16.4
	100.0 ± 0.0	100.0 ± 0.0	99.5 ± 0.7	99.1 ± 0.6	98.4 ± 1.3
MFM / LAND	91.2 ± 9.5	99.9 ± 0.2	92.0 ± 8.3	73.2 ± 23.6	50.9 ± 15.1
	76.0 ± 16.6	84.8 ± 1.8	98.5 ± 2.2	74.0 ± 34.9	96.9 ± 4.4
Curly-FM	48.8 ± 16.7	68.0 ± 7.1	37.2 ± 3.9	33.9 ± 18.8	0.0 ± 0.0
	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
FFM (ours), deterministic	95.6 ± 3.4	92.2 ± 10.4	89.4 ± 5.7	98.9 ± 0.8	77.8 ± 10.4
	94.1 ± 5.2	89.1 ± 13.9	94.2 ± 8.2	97.7 ± 0.8	98.3 ± 0.3
FFM (ours), with noise	97.3 ± 3.4	94.0 ± 8.5	93.8 ± 4.2	99.5 ± 0.8	79.6 ± 6.5
	94.5 ± 7.7	81.2 ± 19.5	98.5 ± 1.8	98.7 ± 0.4	93.1 ± 4.8

to this field, using an objective of the form

$$\mathcal{L}_{\mathrm{Curly}}(\eta) = \mathbb{E}_{x_0,x_1,t}\left[\|\dot{x}_{t,\eta} - f_t(x_{t,\eta})\|_2^2\right].$$

When the dynamics are instead specified by a transition matrix P, the natural first-order vector-field summary is its first spatial moment $m(x_i) = \sum_{j=1}^{N} P_{ij}(x_j - x_i)$,. If we replace the Euclidean norm in the Curly objective by the second moment C_ρ^{-1}

$$\|\dot{x}_{t,\eta} - m(x)\|^2_{C(x)^{-1}} = (\dot{x}_{t,\eta} - m(x))^T C_\rho(x)^{-1}(\dot{x}_{t,\eta} - m(x)) \tag{25}$$

which gives us the integrant of the fixed time FW functional 16. For data that follows the manifold hypothesis we have in particular that $C \neq I$. This observation also suggests a direct second-moment generalization of the Curly-FM path objective. Specifically, we replace its Euclidean velocity residual by the graph-derived anisotropic residual

$$\frac{1}{a}\big(\dot{x} - a\, m(x)\big)^\top C_\rho(x)^{-1}\big(\dot{x} - a\, m(x)\big),$$

so that the control has access to the same first and second transition moments as FFM. We evaluate this matched control on Pancreas in Fig. 3. Full FFM obtains lower $W_2(1/2)$ under both a shared endpoint coupling and method-specific couplings, indicating that the observed improvement is not explained solely by supplying the path objective with second-moment information.

Furthermore, FFM can also be used in settings with access to a reference vector field. Following velocity-kernel constructions such as those used in CellRank (Lange et al., 2022), a vector field f can be converted into a local transition matrix P^f. Let $W_{ij}^{(h)} = \mathbf{1}\{\|x_j - x_i\| \le h_i\}$, where h_i denotes the distance to the k-th nearest neighbor, and define

$$P_{ij}^f = \frac{W_{ij}^{(h)} \exp\left(\kappa \frac{f(x_i)^\top (x_j - x_i)}{\|f(x_i)\|\,\|x_j - x_i\|}\right)}{\sum_\ell W_{i\ell}^{(h)} \exp\left(\kappa \frac{f(x_i)^\top (x_\ell - x_i)}{\|f(x_i)\|\,\|x_\ell - x_i\|}\right)}.$$

For dense and locally isotropic sampling, the k-NN neighborhood of x_i is approximately rotationally symmetric. The cosine kernel assigns a displacement direction u a weight proportional to $\exp(\kappa\, \hat{f}(x_i)^\top u)$, where $\hat{f} = f/\|f\|$. By local isotropy, contributions orthogonal to $\hat{f}$ appear in pairs in opposite directions with and hence cancel eachother. Consequently,

$$m_N^f(x_i) = \sum_j P_{ij}^f (x_j - x_i) = c_i\, \hat{f}(x_i) + o(1), \qquad c_i > 0,$$

as the neighborhood becomes dense. Thus, the first moment of the cosine velocity kernel recovers the orientation of the vector field. FFM can therefore be applied also when the dynamical information is given as a vector field: its first moment retains the directional information used by Curly-FM,

while its second moment additionally describes the local distribution of accessible transitions. This is particularly helpful when the vector field measurement is noisy, as is often the case for RNA-velocity estimates. Then the construction of the transition kernel can be viewed as a denoising of the vector field.