Lagrangian Flow Matching: A Least-Action Framework for Principled Path Design

Shukai Du[1]* **Junzhe Zhang**[2]* **Yiming Li**[1]
[1]Department of Mathematics, Syracuse University
[2]Department of EECS, Syracuse University
{sdu113,jzhan403,yli658}@syr.edu

Abstract

Flow matching trains a neural velocity field by regression against a target velocity associated with a prescribed probability path connecting a simple initial distribution to the data distribution. A central design choice is the path itself. Existing constructions, including rectified and optimal-transport-based paths, transport samples along straight lines between coupled endpoints and thus cover only a narrow class of dynamics. We observe that this corresponds to the simplest case of the *least-action principle* in classical mechanics, in which the kinetic Lagrangian $\mathcal{L} = \frac{1}{2}\|\dot{x}\|^2$ yields free-particle straight-line trajectories. Building on this observation, we propose *Lagrangian flow matching*, a physics-based framework in which the probability path and velocity field are determined by minimizing the action of a general Lagrangian subject to the continuity equation and the prescribed endpoints. We show that this dynamic problem admits an equivalent static optimal transport (OT) formulation, yielding a family of simulation-free training objectives that recover OT-based flow matching as the kinetic special case and the trigonometric variance-preserving diffusion path as the harmonic-oscillator case. More general Lagrangians give rise to new probability paths and velocity fields, and numerical experiments show that they induce meaningful changes in the learned dynamics while remaining competitive with existing conditional flow matching models.

1 Introduction

A natural way to approach generative modeling is from a continuous-time perspective: rather than modeling the data distribution directly, one constructs a time-dependent probability path $(p_t)_{t\in[0,1]}$ that continuously transports a simple initial distribution p_{init} into the target distribution p_{data}. Two important classes of generative models built on this perspective are diffusion models [26, 14, 28] and flow-based models [9, 13, 19, 20, 29], which realize the transport through stochastic differential equations (SDE) and ordinary differential equations (ODE), respectively.

We focus on flow-based models. An influential early class is continuous normalizing flows (CNFs) [9, 13], which are expressive enough to represent a broad class of continuous probability paths [27] but are computationally expensive to train: likelihood-based training requires repeated forward and backward ODE solves. *Flow matching* provides an efficient alternative by directly regressing a neural velocity field $v_\theta(x,t)$ against a target velocity field $v_t(x)$ associated with a *prescribed probability path* [19, 20, 29]. By avoiding ODE solves during training, this simulation-free objective has enabled flow-based models to scale to high-dimensional data and large-scale applications.

The performance of a flow matching model, however, is determined not only by the regression objective itself but also by the prescribed probability path (p_t) and its associated velocity field v_t.

*Equal contribution.

Preprint.

Existing methods choose these objects from a relatively narrow design space: conditional Gaussian paths [19, 29], affine interpolations between coupled endpoints [20], and optimal-transport-based straight-line interpolations [29]. This raises a natural question: can the probability paths and velocity fields used in flow matching be selected from a more general, principled design space?

We address this question through the lens of the least-action principle, a foundational variational principle in classical mechanics [4]. Given a Lagrangian $\mathcal{L}(x, \dot{x}, t)$, the principle selects, between two fixed endpoints, the trajectory that minimizes the corresponding action. The choice of Lagrangian fully determines the dynamics: the kinetic case yields straight-line free-particle trajectories (Fig. 1, top-left), while a general Lagrangian produces curved paths governed by the Euler–Lagrange equations (Fig. 1, top-right).

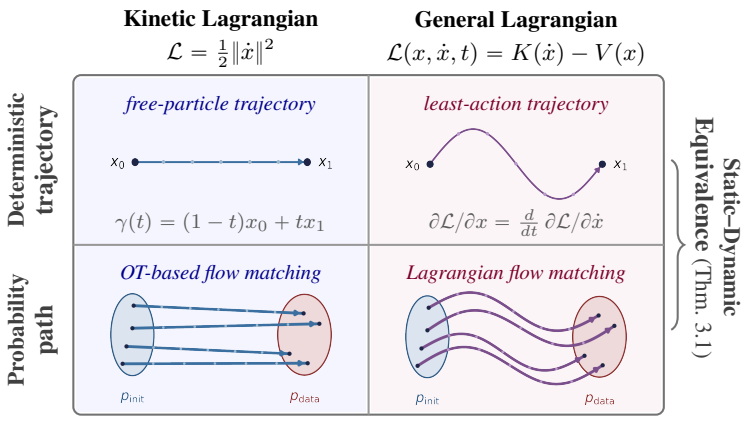


Figure 1: Lagrangian flow matching unifies trajectory selection (top row) and probability-path construction (bottom row) under the least-action principle, linked by the static–dynamic equivalence (Thm. 2.1). The kinetic Lagrangian (left column) yields straight-line trajectories and recovers OT-based flow matching; a general Lagrangian (right column) yields curved trajectories and induces new probability paths.

The same principle applies naturally to flow matching. Existing construction based on affine and OT-based paths transport samples along straight-line trajectories between coupled endpoints (Fig. 1, bottom-left), corresponding to the kinetic case. Replacing the kinetic Lagrangian by a more general one yields curved least-action trajectories between matched endpoints, which in turn induce new probability paths (Fig. 1, bottom-right). This transfer from trajectories to probability paths are made rigorous by the classical static–dynamic equivalence in optimal transport [6, 7, 30].

Several recent works have also explored variational or least-action formulations for learning transport dynamics [22, 23, 25, 15]. Comparatively little work, however, has aimed to cast least-action dynamics directly as a flow-matching problem. We aim to fill this gap by introducing a Lagrangian perspective to flow matching: the Lagrangian defines endpoint costs, selects couplings, and constructs explicit least-action paths. These paths are then used to define a standard flow-matching regression objective. In this sense, our approach aims to combine the flexibility of least-action-induced probability paths with the simplicity and efficiency of the simulation-free regression in flow matching.

Building on this perspective, we propose *Lagrangian flow matching*, a framework in which the probability path and velocity field are determined by minimizing a Lagrangian action subject to the continuity equation and the endpoint constraints $p_0 = p_{\text{init}}$ and $p_1 = p_{\text{data}}$. Our contributions are threefold. First, we propose a physics-based framework for selecting flow matching probability paths and velocity fields through a least-action principle. Second, we derive a family of simulation-free training objectives, and show that several existing constructions, including OT-based and conditional flow matching models, are recovered as special cases. Third, we derive new flow matching models from more general Lagrangians, in particular a *harmonic family* indexed by a frequency $\omega \in (0, \pi)$ that continuously deforms the least-action path, and recovers OT-based flow matching and the trigonometric variance-preserving schedule at its two endpoints. We validate the framework on synthetic 2D data, single-cell trajectory interpolation, and CIFAR-10 image generation, and show that different Lagrangians induce meaningful changes in the learned dynamics while remaining competitive with existing conditional flow matching models. All proofs of the results and details about the experimental setup are provided in Appendices D and H. Code is available at `https://github.com/junzhez/lagrangian-flow-matching`.

1.1 Preliminaries and Related Work

Throughout, p_{init} and p_{data} are probability densities on $\mathbb{R}^d$. The goal of generative modeling is to learn a transformation pushing p_{init} to p_{data}, given finitely many samples from p_{data}. This setting

includes both the standard case in which p_{init} is easily sampled (e.g., a standard Gaussian) and the more general case in which both p_{init} and p_{data} are accessed only through samples [2, 20, 29]. We focus on continuous-time approaches realizing this transformation through a probability path $(p_t)_{t\in[0,1]}$ with $p_0 = p_{\text{init}}$ and $p_1 = p_{\text{data}}$.

Flow matching trains a parametrized velocity field $v_\theta(x,t)$ by regressing against a target field $v_t(x)$:

$$\min_\theta \ \mathbb{E}_{t\sim\mathcal{U}[0,1],\, x\sim p_t} \big\| v_\theta(x,t) - v_t(x) \big\|^2, \tag{1}$$

where $(p_t)_{t\in[0,1]}$ satisfies the endpoint conditions $p_0 = p_{\text{init}}$, $p_1 = p_{\text{data}}$, and v_t generates this path through the continuity equation $\partial_t p_t + \nabla\cdot(p_t v_t) = 0$. Infinitely many pairs (p_t, v_t) satisfy these conditions, and the model's behavior depends on the choice.

Conditional flow matching addresses this intractability by expressing p_t as a mixture $p_t(x) = \mathbb{E}_{z\sim q}[p_t(x \mid z)]$ over a conditioning variable $z \sim q$ [19, 29]. If $v_t(x \mid z)$ generates $p_t(\cdot \mid z)$, then the marginal velocity is the posterior average $v_t(x) = \mathbb{E}[v_t(x \mid z) \mid x_t = x]$. The marginal objective (1) is equal, up to an additive constant, to

$$\min_\theta \ \mathbb{E}_{t\sim\mathcal{U}[0,1],\, z\sim q,\, x\sim p_t(\cdot|z)} \big\| v_\theta(x,t) - v_t(x \mid z) \big\|^2, \tag{2}$$

which is tractable when the conditional paths and velocities admit closed-form expressions. A representative example is the affine construction of [19]: with $z = y \sim p_{\text{data}}$ and $p_{\text{init}} = \mathcal{N}(0, I)$, the conditional trajectory is $\psi_t(x \mid y) = (1-(1-\varepsilon)t)x + ty$. We recover this construction as the $\omega \to 0$ limit of conditional Lagrangian flow matching (Example 2.4).

Dynamic optimal transport. The least-action perspective developed here builds on the dynamic formulation of optimal transport due to Benamou and Brenier [6]. The squared 2-Wasserstein distance between p_0 and p_1 admits the representation

$$W_2^2(p_0, p_1) = \inf_{(p_t, v_t)} \int_0^1 \int_{\mathbb{R}^d} \|v_t(x)\|^2 \, p_t(x)\, dx\, dt, \tag{3}$$

where the infimum is over pairs (p_t, v_t) satisfying the continuity equation and the endpoint constraints $p_{t=0} = p_0$, $p_{t=1} = p_1$. This is the kinetic special case of the Lagrangian framework developed below, and the basic viewpoint behind OT-based flow matching [29]. We defer readers to Appendix A for a more detailed discussion on related work.

2 Lagrangian Flow Matching

This section introduces our main framework. We first recall the elements of Lagrangian mechanics needed to state a least-action principle, then formulate the *Lagrangian optimal transport* problem and establish the static–dynamic equivalence that converts it into a tractable training objective (Sec. 2.1). The conditional formulation, which makes the framework practical at scale, is developed in Sec. 3.

Lagrangian mechanics provides a variational principle for selecting dynamics through an action functional, and has long served as a unifying language for describing physical systems [4]. A *Lagrangian* is a function $\mathcal{L}(x, v, t)$ of position $x \in \mathbb{R}^d$, velocity $v \in \mathbb{R}^d$, and time $t \in [0,1]$ encoding the dynamics of the system. Given a smooth curve $\gamma : [0,1] \to \mathbb{R}^d$, its *action* is

$$S[\gamma] \ = \ \int_0^1 \mathcal{L}\big(\gamma(t), \dot\gamma(t), t\big)\, dt.$$

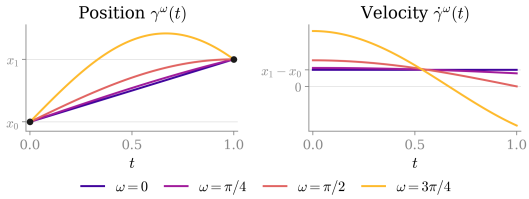


Figure 2: Position $\gamma^\omega(t)$ and velocity $\dot\gamma^\omega(t)$ of the harmonic least-action trajectory (4) for $\omega \in (0, \pi)$, with endpoints $x_0 = 0$ and $x_1 = 1$.

The *least-action principle* selects, between two fixed endpoints $\gamma(0) = x_0$ and $\gamma(1) = x_1$, the curve that minimizes $S[\gamma]$. A standard variational argument shows that any minimizer satisfies the *Euler–Lagrange equation*

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial v}\right) - \frac{\partial \mathcal{L}}{\partial x} \ = \ 0,$$

which, together with the boundary conditions, defines a boundary-value problem that admits a unique *closed-form* solution for some choices of Lagrangians, as illustrated by the following two examples.

Example 2.1 (Free particle)**.** The simplest Lagrangian is the kinetic energy of a free particle, $\mathcal{L}_{\text{free}}(x, v) = \frac{1}{2}\|v\|^2$. The Euler–Lagrange equation reduces to $\ddot{\gamma}(t) = 0$, so the least-action trajectory between x_0 and x_1 is the affine interpolation $\gamma_{x_0,x_1}(t) = (1-t)x_0 + tx_1$, with constant velocity $\dot{\gamma}_{x_0,x_1}(t) = x_1 - x_0$. (Fig. 2, $\omega = 0$ curve). The corresponding endpoint cost is the quadratic cost $c_{\mathcal{L}_{\text{free}}}(x_0, x_1) = \frac{1}{2}\|x_1 - x_0\|^2$, underlying classical Wasserstein optimal transport. ■

Example 2.2 (Harmonic oscillator)**.** A canonical non-kinetic example is $\mathcal{L}_\omega(x, v) = \frac{1}{2}\|v\|^2 - \frac{1}{2}\omega^2\|x\|^2$ with a frequency $0 < \omega < \pi$, adding a quadratic potential $V(x) = \frac{1}{2}\omega^2\|x\|^2$ to the kinetic term. The Euler–Lagrange equation $\ddot{\gamma} + \omega^2\gamma = 0$ has unique solution

$$\gamma^{\omega}_{x_0,x_1}(t) \;=\; \frac{\sin(\omega(1-t))}{\sin\omega}\, x_0 + \frac{\sin(\omega t)}{\sin\omega}\, x_1, \tag{4}$$

with time-varying velocity

$$\dot{\gamma}^{\omega}_{x_0,x_1}(t) = -\frac{\omega\cos(\omega(1-t))}{\sin\omega}\, x_0 + \frac{\omega\cos(\omega t)}{\sin\omega}\, x_1. \tag{5}$$

As $\omega \to 0$, $\sin(\omega s)/\sin\omega \to s$, so both expressions reduce to those of Example 2.1. The harmonic oscillator thus smoothly deforms the free-particle dynamics, with curvature controlled by the frequency ω (Fig. 2, $\omega = 0, \pi/4, \pi/2$ and $3\pi/4$ curve). ■

The least-action principle gives a recipe for selecting a deterministic trajectory between two fixed endpoints. To use this recipe in flow matching, where endpoints are samples from p_{init} and p_{data} rather than fixed points, we need a way to (i) couple endpoints across the two distributions, and (ii) assemble the resulting least-action trajectories into a probability path. This is the role of *Lagrangian optimal transport*, developed next, which generalizes the Benamou–Brenier dynamic formulation [6] from the kinetic Lagrangian to general Lagrangians [7, 30].

2.1 Lagrangian Optimal Transport

For $x_0, x_1 \in \mathbb{R}^d$, consider the least-action problem

$$\gamma_{x_0,x_1} \;:=\; \arg \min_{\substack{\gamma \in AC([0,1];\mathbb{R}^d) \\ \gamma(0)=x_0,\, \gamma(1)=x_1}} \int_0^1 \mathcal{L}(\gamma(t), \dot{\gamma}(t), t)\, dt, \tag{6}$$

where $AC([0,1];\mathbb{R}^d)$ denotes the space of absolutely continuous curves. We restrict throughout to Lagrangians of the classical separable form $\mathcal{L}(x, v, t) = K(v) - V(x)$, with K a quadratic form satisfying $K(v) \geqslant \alpha\|v\|^2$ for some $\alpha > 0$. The potential $V \geqslant 0$ is taken sufficiently regular and controlled by the kinetic term so that (6) admits a unique minimizer for every (x_0, x_1). The corresponding least-action cost is

$$c_{\mathcal{L}}(x_0, x_1) \;:=\; \int_0^1 \mathcal{L}\big(\gamma_{x_0,x_1}(t), \dot{\gamma}_{x_0,x_1}(t), t\big)\, dt. \tag{7}$$

Given probability densities p_{init} and p_{data} on $\mathbb{R}^d$, using $c_{\mathcal{L}}$ as the transport cost yields the *static* Lagrangian optimal transport problem

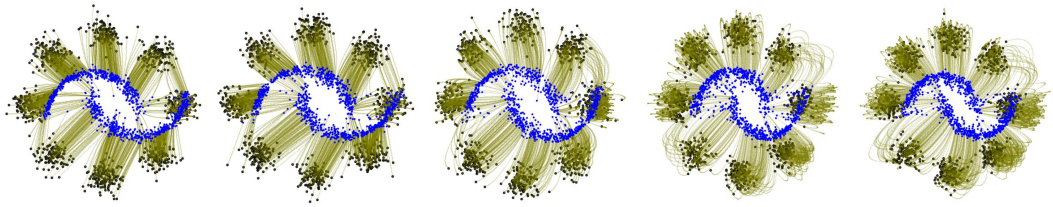
$$\mathcal{A}_{\text{stat}} \;:=\; \inf_{\pi \in \Pi(p_{\text{init}}, p_{\text{data}})} \int_{\mathbb{R}^d \times \mathbb{R}^d} c_{\mathcal{L}}(x_0, x_1)\, d\pi(x_0, x_1), \tag{8}$$

where $\Pi(p_{\text{init}}, p_{\text{data}})$ is the set of couplings between p_{init} and p_{data}. The corresponding *dynamic* formulation minimizes the action of the density flow:

$$\mathcal{A}_{\text{dyn}} \;:=\; \inf_{(p,v) \in \mathfrak{D}} \int_0^1 \!\!\int_{\mathbb{R}^d} \mathcal{L}(x, v(x,t), t)\, p(x,t)\, dx\, dt, \tag{9}$$

where $\mathfrak{D} := \big\{(p, v) :\, \partial_t p + \nabla \cdot (vp) = 0,\, p(\cdot, 0) = p_{\text{init}},\, p(\cdot, 1) = p_{\text{data}}\big\}$.

The two formulations are equivalent. Recall that for a measurable map $T : \mathbb{R}^m \to \mathbb{R}^n$ and a probability measure μ on $\mathbb{R}^m$, the pushforward $T_{\#}\mu$ is defined by $(T_{\#}\mu)(A) := \mu(T^{-1}(A))$, and that the evaluation map $e_t : AC([0,1];\mathbb{R}^d) \to \mathbb{R}^d$ is $e_t(\gamma) = \gamma(t)$.



(a) OT-CFM ($\omega \to 0$) (b) Harmonic $\omega = 0.001$ (c) Harmonic $\omega = 1$ (d) Harmonic $\omega = \pi/2$ (e) Stochastic interpolant

Figure 3: Mini-batch harmonic flow matching trajectories from a 8-Gaussian-mixture to a double-moon target, swept over $\omega \in (0, \pi/2]$ and compared against OT-CFM and the stochastic interpolant. Endpoints (black) are matched by quadratic-cost OT and connected along least-action trajectories (olive); the data manifold is shown in blue. See Examples 2.4 and 2.5 for the limiting cases.

Theorem 2.1 (Static–Dynamic Equivalence)**.** *The static and dynamic Lagrangian optimal transport problems have the same value:* $\mathcal{A}_{\mathrm{stat}} = \mathcal{A}_{\mathrm{dyn}}$*. Moreover, if* π^* *is an optimal plan for* (8)*, the probability measure* $\eta^* := (\gamma_{x_0,x_1})_{\#}\pi^*$ *on* $AC([0,1];\mathbb{R}^d)$ *induces an optimal pair for* (9) *via*

$$p^*(\cdot, t) := (e_t)_{\#}\eta^*, \qquad v^*(x,t) \;:=\; \mathbb{E}_{\gamma\sim\eta^*}\big[\dot\gamma(t) \,\big|\, \gamma(t) = x\big].$$

Thm. 2.1 shows that the least-action probability path is obtained by pushing an optimal static coupling forward along least-action trajectories. This converts training into a two-step procedure. First, solve the static problem (8) for an optimal coupling π^*. Then, for samples $(x_0, x_1) \sim \pi^*$, use the closed-form least-action trajectory γ_{x_0,x_1} as a velocity target:

$$\min_{\theta}\; \mathbb{E}_{t\sim\mathcal{U}[0,1]}\, \mathbb{E}_{(x_0,x_1)\sim\pi^*} \big\| v_\theta\big(\gamma_{x_0,x_1}(t), t\big) - \dot\gamma_{x_0,x_1}(t)\big\|^2. \tag{10}$$

This sidesteps the dynamic problem (9) entirely, preserving the simulation-free character of flow matching. The following examples illustrate the procedure for the harmonic Lagrangian of Example 2.2 and recover several existing flow matching constructions as special cases.

Example 2.3 (Harmonic flow)**.** For $\mathcal{L}_\omega$ of Example 2.2, substituting (4) into the action yields the cost

$$c_\omega(x_0, x_1) \;=\; \frac{\omega}{2\sin\omega}\big[\cos\omega\,(\|x_0\|^2 + \|x_1\|^2) - 2\,x_0\cdot x_1\big].$$

For any coupling $\pi \in \Pi(p_{\mathrm{init}}, p_{\mathrm{data}})$, the quadratic terms in c_ω integrate to constants determined by the marginals alone:

$$\int \|x_0\|^2\, d\pi(x_0,x_1) = \mathbb{E}_{x_0\sim p_{\mathrm{init}}}\|x_0\|^2 =: M_0, \qquad \int \|x_1\|^2\, d\pi(x_0,x_1) = \mathbb{E}_{x_1\sim p_{\mathrm{data}}}\|x_1\|^2 =: M_1,$$

which are finite under the standard second-moment assumption on p_{init} and p_{data}. The transport cost therefore decomposes as

$$\int c_\omega(x_0,x_1)\, d\pi(x_0,x_1) \;=\; \frac{\omega\cos\omega}{2\sin\omega}\,(M_0 + M_1) \;-\; \frac{\omega}{\sin\omega}\int x_0\cdot x_1\, d\pi(x_0,x_1),$$

where the first term does not depend on π. Minimizing over π is therefore equivalent to maximizing the cross-correlation $\int x_0\cdot x_1\, d\pi$, and by the polarization identity $x_0\cdot x_1 = \frac{1}{2}(\|x_0\|^2 + \|x_1\|^2 - \|x_0 - x_1\|^2)$, this is in turn equivalent to minimizing $\int \frac{1}{2}\|x_0 - x_1\|^2\, d\pi$. The static problem (8) with cost c_ω thus shares its minimizer with the quadratic-cost OT problem for every $\omega \in (0, \pi)$; we denote this common coupling by π^*_{OT}. The training objective (10) becomes

$$\min_{\theta}\; \mathbb{E}_{t,\,(x_0,x_1)\sim\pi^*_{\mathrm{OT}}} \big\| v_\theta\big(\gamma^{\omega}_{x_0,x_1}(t), t\big) - \dot\gamma^{\omega}_{x_0,x_1}(t)\big\|^2,$$

where the position $\gamma^{\omega}_{x_0,x_1}(t)$ and velocity $\dot\gamma^{\omega}_{x_0,x_1}(t)$ of the harmonic oscillator are given by Eqs. (4) and (5). This preserves the OT endpoint coupling while replacing the affine interpolation between paired endpoints with the curved harmonic trajectory. Fig. 3 visualizes how the resulting trajectories vary with $\omega \in \{0.001,\ 1,\ \pi/2\}$ on an 8-Gaussian-mixture source and a two-moons target. ■

Example 2.4 (OT-based flow matching as the $\omega \to 0$ limit)**.** Taking $\omega \to 0$ in Example 2.3, the harmonic trajectory reduces to the affine interpolation $(1-t)x_0 + tx_1$ and the cost reduces to $\frac{1}{2}\|x_1 - x_0\|^2$, so the training objective becomes

$$\min_{\theta} \; \mathbb{E}_{t,\,(x_0,x_1)\sim\pi^*_{\mathrm{OT}}} \big\| v_\theta\big((1-t)x_0 + tx_1,\, t\big) - (x_1 - x_0)\big\|^2,$$

recovering OT-based flow matching [29] (Fig. 3a; the small-ω case in Fig. 3b is visually indistinguishable). The affine construction of [19] is the conditional analogue of this limit, recovered in Sec. 3 by a different choice of conditioning variable. ■

Example 2.5 (Trigonometric interpolation at $\omega = \pi/2$)**.** At $\omega = \pi/2$, the harmonic trajectory reduces to the trigonometric interpolation

$$\gamma^{\pi/2}_{x_0,x_1}(t) \;=\; \cos\!\left(\frac{\pi t}{2}\right) x_0 + \sin\!\left(\frac{\pi t}{2}\right) x_1,$$

which connects to two existing constructions. First, it is the spatial interpolation underlying stochastic-interpolant constructions [1] (Fig. 3d and e show the two constructions side by side); here, however, the trajectory is fully deterministic once endpoints are fixed, with randomness entering only through endpoint sampling. Second, with $p_{\mathrm{init}} = \mathcal{N}(0, I)$, the trajectory reads $x_t = \sigma_t x_0 + \alpha_t x_1$ where $\sigma_t = \cos(\pi t/2)$ and $\alpha_t = \sin(\pi t/2)$ satisfy the variance-preserving normalization $\sigma_t^2 + \alpha_t^2 = 1$ used in diffusion-type paths [19, 14]. The harmonic Lagrangian thus provides a least-action interpretation of the trigonometric VP schedule at the level of marginal probability path; the full diffusion model also specifies a stochastic forward process, which lies outside the current deterministic framework. ■

3 Conditional Lagrangian Flow Matching

The least-action problem (9) provides a principled mechanism for selecting probability paths and velocity fields, but solving it directly can be challenging in practice. While the static–dynamic equivalence (Thm. 2.1) reduces the problem to a static optimal transport problem between p_{init} and p_{data}, the resulting OT problem may still be difficult to solve accurately, particularly in high-dimensional or sample-only regimes. A common strategy in flow matching is to represent a global probability path as a mixture of simpler conditional paths [19, 29]. We adapt this strategy to Lagrangian Flow Matching: each conditional path is selected by a least-action principle, and the Lagrangian is allowed to depend on the conditioning variable.

Let $z \sim q$ be a conditioning variable, and let $p_0(\cdot \mid z)$ and $p_1(\cdot \mid z)$ be conditional endpoint distributions whose mixtures recover the desired marginals:

$$\int p_0(\cdot \mid z)\, q(dz) = p_{\mathrm{init}}, \qquad \int p_1(\cdot \mid z)\, q(dz) = p_{\mathrm{data}}. \tag{11}$$

The conditional endpoints need not match p_{init} and p_{data} individually; only their mixtures must. For each z, let $\mathcal{L}_z(x, v, t)$ be a (possibly z-dependent) Lagrangian and let $S_z[\gamma] = \int_0^1 \mathcal{L}_z\big(\gamma(t), \dot\gamma(t), t\big)\, dt$ be the corresponding action. Then, the associated z-conditional least-action trajectory and cost are

$$\gamma^z_{x_0,x_1} \;:=\; \arg \min_{\substack{\gamma \in AC([0,1];\mathbb{R}^d) \\ \gamma(0)=x_0,\, \gamma(1)=x_1}} S_z[\gamma], \qquad c_z(x_0, x_1) \;:=\; S_z[\gamma^z_{x_0,x_1}]. \tag{12}$$

For each z, the conditional endpoint coupling is obtained by solving the z-conditional static optimal transport problem

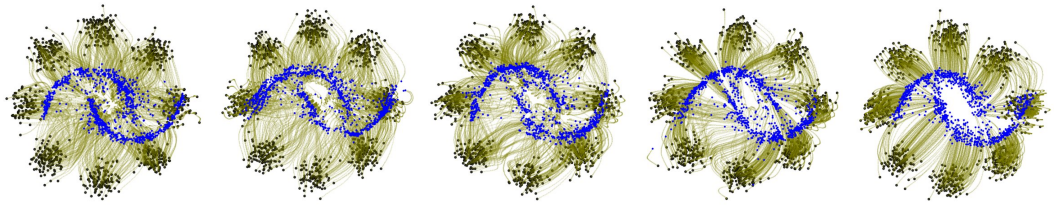
$$\pi^*_z \;\in\; \operatorname*{argmin}_{\pi \in \Pi(p_0(\cdot|z), p_1(\cdot|z))} \int_{\mathbb{R}^d \times \mathbb{R}^d} c_z(x_0, x_1)\, d\pi(x_0, x_1). \tag{13}$$

Pushing π^*_z forward along the least-action trajectories yields the conditional probability path and velocity field

$$p_t(\cdot \mid z) \;=\; \big(\gamma^z_{x_0,x_1}(t)\big)_{\#}\pi^*_z, \quad v_t(x \mid z) \;=\; \mathbb{E}_{(x_0,x_1)\sim\pi^*_z}\big[\dot\gamma^z_{x_0,x_1}(t) \,\big|\, \gamma^z_{x_0,x_1}(t) = x\big]. \tag{14}$$

Applying Thm. 2.1 to each z independently, $(p_t(\cdot \mid z), v_t(\cdot \mid z))$ is the optimal solution of the z-conditional dynamic least-action problem

$$\mathcal{A}_{\mathrm{cond}} \;:=\; \inf_{\{p_t(\cdot|z), v_t(\cdot|z)\}_z} \mathbb{E}_{z\sim q} \int_0^1 \!\!\int_{\mathbb{R}^d} \mathcal{L}_z\big(x, v_t(x \mid z), t\big)\, p_t(x \mid z)\, dx\, dt, \tag{15}$$



(a) $n = 1$ (b) $n = 10$ (c) $n = 25$ (d) $n = 50$ (e) $n = 100$

Figure 4: Mini-batch harmonic flow matching ($\omega = 1$) from an 8-Gaussian-mixture source (olive) to a double-moon target (blue), swept over OT batch size $n \in \{1, 10, 25, 50, 100\}$. Increasing n tightens the empirical coupling toward the OT solution: compare the many-to-many fan-out at $n = 1$ to the bundled, mode-to-region routing at $n = 100$. Trajectory curvature is set by ω and is preserved across all panels — the two axes (n for the coupling, $\mathcal{L}$ for the matched-pair trajectory) act independently.

with $p_t(\cdot|z)$ and $v_t(\cdot|z)$ subject to the continuity equation and endpoint constraints.

The least-action trajectories $\gamma^z_{x_0,x_1}$ provide the following training targets for the conditional Lagrangian flow matching objective

$$\min_\theta \; \mathbb{E}_{z\sim q}\, \mathbb{E}_{(x_0,x_1)\sim\pi^*_z}\, \mathbb{E}_{t\sim\mathcal{U}[0,1]} \big\| v_\theta(\gamma^z_{x_0,x_1}(t), t) - \dot\gamma^z_{x_0,x_1}(t) \big\|^2, \tag{16}$$

which generalizes the unconditional objective (10). The conditioning distribution q, conditional endpoints $p_0(\cdot|z)$ and $p_1(\cdot|z)$, and z-dependent Lagrangians $\mathcal{L}_z$ are design parameters that can be chosen for the problem at hand, while the simulation-free property is preserved provided the trajectories $\gamma^z_{x_0,x_1}$ admit closed-form expressions. The next proposition shows that the conditional formulation is consistent with the unconditional one when the same convex Lagrangian is used.

Proposition 3.1 (Conditional action upper-bounds the unconditional action)**.** *Assume $\mathcal{L}_z = \mathcal{L}$ for all z and that $\mathcal{L}(x, \cdot, t)$ is convex in its velocity argument. Then $\mathcal{A}_{\text{cond}} \geqslant \mathcal{A}_{\text{dyn}}$, where $\mathcal{A}_{\text{dyn}}$ is the unconditional least-action value of* (9)*. Moreover, equality is attained by the trivial conditioning $q = \delta_{z_0}$ with $p_0(\cdot \mid z_0) = p_{\text{init}}$ and $p_1(\cdot \mid z_0) = p_{\text{data}}$.*

The conditional formulation thus never undershoots the unconditional optimum and matches it in the trivial case. More details are deferred to Appendix D.2.

3.1 Mini-batch Conditional Lagrangian Flow Matching

A natural and useful instance of the conditional formulation is *mini-batch* conditional Lagrangian flow matching, in which the conditioning variable is a pair of finite batches drawn independently from p_{init} and p_{data}. This generalizes mini-batch optimal-transport flow matching [29, 24] from the kinetic Lagrangian to general Lagrangians, and is well-suited to the sample-only regime in which p_{init} and p_{data} are accessed only through samples.

Fix a batch size $n \geqslant 1$, and take the conditioning variable to be a pair of independent batches $z = (x_0^{(1:n)}, x_1^{(1:n)})$ drawn from $p_{\text{init}}^{\otimes n} \otimes p_{\text{data}}^{\otimes n}$, with conditional endpoint distributions given by the corresponding empirical distributions $p_0(\cdot \mid z) = \frac{1}{n}\sum_{i=1}^n \delta_{x_0^{(i)}}$ and $p_1(\cdot \mid z) = \frac{1}{n}\sum_{i=1}^n \delta_{x_1^{(i)}}$. The mixture conditions (11) hold by construction. We use a single, z-independent Lagrangian $\mathcal{L}$ of the form $K(v) - V(x)$, with associated trajectory γ_{x_0,x_1} and cost $c_{\mathcal{L}}(x_0, x_1)$ as in (7). With these choices, the conditional static problem (13) reduces to a discrete optimal transport problem between the two empirical batches:

$$\pi^*_z \;\in\; \operatorname*{argmin}_{\pi\in\Pi(p_0(\cdot|z),p_1(\cdot|z))} \sum_{i,j} \pi_{ij}\, c_{\mathcal{L}}(x_0^{(i)}, x_1^{(j)}), \tag{17}$$

which is a small linear assignment problem solvable by standard methods (e.g., the Hungarian algorithm or Sinkhorn iterations) at every gradient step. Substituting the matched pairs into (16) and writing the conditional expectation explicitly over batch indices gives the mini-batch conditional Lagrangian flow matching objective

$$\min_\theta \; \mathbb{E}_{z\sim p_{\text{init}}^{\otimes n}\otimes p_{\text{data}}^{\otimes n}}\, \mathbb{E}_{(i,j)\sim\pi^*_z}\, \mathbb{E}_{t\sim\mathcal{U}[0,1]} \big\| v_\theta\big(\gamma_{x_0^{(i)},x_1^{(j)}}(t), t\big) - \dot\gamma_{x_0^{(i)},x_1^{(j)}}(t) \big\|^2, \tag{18}$$

Table 1: Comparison of harmonic flow matching across four distribution pairs ($\mu \pm \sigma$ over five seeds) in terms of distribution fit (2-Wasserstein), deviation from the harmonic OT path at $\omega = 1$ (normalized path energy), and training time. Lower is better for both metrics; best is in **bold**.

Dataset →	$\mathcal{N} \to$ 8gaussians		moons↔8gaussians		$\mathcal{N} \to$ moons		$\mathcal{N} \to$ scurve		Avg. train time
Algorithm ↓ Metric →	W_2	$\text{NPE}_{\omega=1}$	W_2	$\text{NPE}_{\omega=1}$	W_2	$\text{NPE}_{\omega=1}$	W_2	$\text{NPE}_{\omega=1}$	($\times 10^3$ s)
OT-CFM	$0.505_{\pm 0.118}$	$0.087_{\pm 0.037}$	$0.336_{\pm 0.113}$	$0.187_{\pm 0.056}$	$0.197_{\pm 0.026}$	$0.138_{\pm 0.047}$	$0.976_{\pm 0.198}$	$0.051_{\pm 0.041}$	$\mathbf{0.827}_{\pm 0.037}$
OT-SI	$0.521_{\pm 0.092}$	$0.497_{\pm 0.042}$	$0.401_{\pm 0.154}$	$1.308_{\pm 0.148}$	$0.227_{\pm 0.036}$	$0.785_{\pm 0.117}$	$1.126_{\pm 0.183}$	$0.325_{\pm 0.041}$	$0.888_{\pm 0.095}$
OT-Harmonic, $\omega = 0.001$	$0.504_{\pm 0.119}$	$0.087_{\pm 0.037}$	$\mathbf{0.335}_{\pm 0.113}$	$0.186_{\pm 0.058}$	$0.197_{\pm 0.024}$	$0.136_{\pm 0.047}$	$\mathbf{0.975}_{\pm 0.199}$	$0.051_{\pm 0.041}$	$0.898_{\pm 0.080}$
OT-Harmonic, $\omega = 1$	$\mathbf{0.493}_{\pm 0.117}$	$\mathbf{0.025}_{\pm 0.030}$	$0.3725_{\pm 0.181}$	$\mathbf{0.046}_{\pm 0.035}$	$\mathbf{0.188}_{\pm 0.016}$	$\mathbf{0.036}_{\pm 0.029}$	$1.023_{\pm 0.241}$	$\mathbf{0.037}_{\pm 0.008}$	$0.845_{\pm 0.038}$
OT-Harmonic, $\omega = \pi/2$	$0.520_{\pm 0.092}$	$0.497_{\pm 0.045}$	$0.402_{\pm 0.155}$	$1.308_{\pm 0.149}$	$0.228_{\pm 0.037}$	$0.785_{\pm 0.118}$	$1.127_{\pm 0.186}$	$0.325_{\pm 0.041}$	$0.907_{\pm 0.115}$

where π_z^* from (17) is viewed as a distribution over index pairs $(i, j) \in \{1, \ldots, n\}^2$. Each gradient step samples a batch z, solves (17), and uses γ_{x_0,x_1} as the velocity target along matched pairs. As $n \to \infty$, the empirical marginals converge to p_{init} and p_{data}, and π_z^* converges to the global optimal coupling π^* under mild regularity. For finite n the coupling is approximate, but the objective remains simulation-free for any Lagrangian admitting closed-form least-action trajectories.[2]

Fig. 4 illustrates how the batch size n affects the trajectories in mini-batch harmonic flow matching. As n increases, the endpoint pairings gradually transition from independent pairings to OT pairing. Consequently, the resulting training trajectories increasingly reflect the global OT geometry. In the limiting case $\omega \to 0$, it recovers standard OT flow matching as $n \to \infty$, and recovers rectified or affine conditional flow matching when $n = 1$. Cases with $\omega \neq 0$ can therefore be viewed as natural extensions of these existing frameworks. See Appendix E for a detailed discussion with examples.

4 Experiments

We evaluate Lagrangian flow matching on three settings of increasing dimensionality: synthetic two-dimensional data, single-cell trajectory interpolation, and CIFAR-10 image generation. We instantiate three OT-Harmonic models at frequencies $\omega \in \{0.001, 1, \pi/2\}$ and compare against OT-CFM [29] and OT-SI [2], with an additional anisotropic variant (OT-Aniso, see Appendix F for details) on CIFAR-10. Baselines are retrained under our own procedure so that all methods share the same architecture, optimizer, training budget, and evaluation protocol for controlled comparison.

Three findings emerge. First, the OT-Harmonic family interpolates between the two baselines along ω: as $\omega \to 0$, harmonic geodesics degenerate to straight lines and recovers OT-CFM, while at $\omega = \pi/2$ the interpolant reduces to the cosine schedule underlying OT-SI. Second, $\omega = 0.001$ slightly improves on OT-CFM on Multiome and on CIFAR-10, suggesting that minor deviations from straight-line interpolation can be beneficial even in the limit where the harmonic flow approaches OT-CFM theoretically. Third, $\omega = 1$ offers the largest gains in the 2D setting, improving on both baselines in W_2 on three of four distribution pairs; in higher-dimensional settings it remains competitive but no longer dominates. We defer the experimental setup and additional simulation results to Appendix H.

Two-dimensional data. To assess how closely a learned flow recovers the harmonic optimal transport solution, we report the normalized path energy $\text{NPE}_\omega[\phi] = |\, K[\phi]/C_\omega(\pi_0, \pi_1) - 1 \,|$, the relative deviation of the flow's kinetic energy $K[\phi]$ from the harmonic-OT reference cost $C_\omega(\pi_0, \pi_1)$ at frequency ω (closed-form derivation in Appendix G). $\text{NPE}_\omega = 0$ iff ϕ is the harmonic-OT flow at ω.

Table 1 summarizes the results. At $\omega = 1$ the learned flow tracks the harmonic-OT reference path closely (low $\text{NPE}_{\omega=1}$) while improving on OT-CFM in W_2 on $\mathcal{N} \to$ 8gaussians and $\mathcal{N} \to$ moons. At $\omega = 0.001$ OT-Harmonic reproduces OT-CFM to within

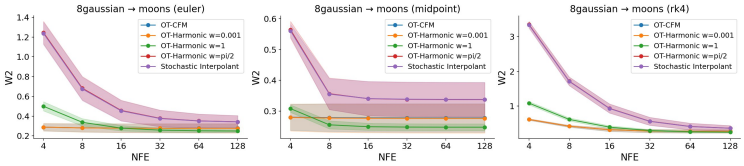


Figure 5: Sample quality vs. inference budget on $\mathcal{N} \to$ moons across three ODE solvers (Euler, midpoint, RK4). OT-Harmonic reduce the 2-Wasserstein distance for a fixed NFE during inference.

[2]In practice, the OT batch size and the training batch size need not be the same. For instance, when the OT batch size n is too small, one may aggregate the outputs of several OT batches to form a larger training batch.

seed variation, confirming the linear-interpolant limit predicted by the $\omega \to 0$ analysis. We further examine the inference-time sample quality in Fig. 5 (also Fig. 8 in Appendix H): OT-Harmonic requires more function evaluations as ω grows, consistent with ω controlling the path curvature.

Single-cell interpolation. We next evaluate on single-cell trajectory interpolation, following the leave-one-out protocol of [29]: the model interpolates the cell distribution at a held-out timepoint t given snapshots at $[0, t-1]$ and $[t+1, T]$. Table 2 reports the 1-Wasserstein distance to the held-out distribution on three datasets: embryoid body [21], CITE-seq, and Multiome [8]. OT-CFM and OT-Harmonic at $\omega = 0.001$ are the two strongest methods and remain within 1.5×10^{-3} on all three datasets, consistent with the $\omega \to 0$ limit in which the harmonic flow recovers OT-CFM. Increasing ω degrades performance monotonically: $\omega = 1$ falls behind both leaders by 4–8% on every dataset, and $\omega = \pi/2$ collapses to the OT-SI regime.

Table 2: Single-cell trajectory interpolation across datasets, averaged over leave-one-out held-out timepoints. Bold indicates methods within the standard deviation of the lowest mean.

Algorithm ↓ Dataset →	Cite	EB	Multi
OT-CFM	$\mathbf{0.8991 \pm 0.0340}$	$\mathbf{0.9519 \pm 0.0405}$	$\mathbf{1.0721 \pm 0.0587}$
OT-SI	1.2424 ± 0.0343	1.3378 ± 0.1381	1.4856 ± 0.0698
OT-Harmonic, $\omega = 0.001$	$\mathbf{0.8985 \pm 0.0374}$	$\mathbf{0.9523 \pm 0.0404}$	$\mathbf{1.0708 \pm 0.0574}$
OT-Harmonic, $\omega = 1$	0.9654 ± 0.0298	0.9909 ± 0.0567	1.1562 ± 0.0765
OT-Harmonic, $\omega = \pi/2$	1.2373 ± 0.0226	1.3352 ± 0.1334	1.4938 ± 0.0742

CIFAR-10. Table 3 reports FID and NFE on CIFAR-10, and Fig. 6 shows uncurated samples from OT-CFM and OT-Harmonic at $\omega = 0.001$; additional sample grids are in Fig. 9. Under adaptive integration, FID rises monotonically with ω across the harmonic family, recovering OT-CFM in the small-ω limit and approaching OT-SI as $\omega \to \pi/2$. At $\omega = 0.001$, OT-Harmonic reaches FID $= 3.681$, slightly improving on our OT-CFM baseline (FID $= 3.706$) at the same adaptive NFE budget; the two sample grids in Fig. 6 are visually indistinguishable, consistent with the $\omega \to 0$ correspondence between the two methods. OT-Harmonic at $\omega = 0.001$ also attains the lowest FID at 1000-step Euler integration (3.849 vs. 3.901), while at the low-budget 100-step setting OT-SI attains the lowest FID (4.474); we read this as an inference-time effect of the cosine schedule rather than a training-time advantage, since OT-SI trails all harmonic and OT-CFM variants once enough integration steps are available. Our OT-CFM number is above the FID $= 3.577$ reported in [29], reflecting our shorter 400,000-step training budget, which we apply uniformly across all variants for a controlled comparison. The anisotropic variant OT-Aniso achieves FID $= 4.078$, no better than OT-SI: when CIFAR-10 pixel coordinates are approximately independent, the rotation recovered from a small fit batch carries little geometric signal beyond an isotropic schedule.

Table 3: FID score and number of function evaluations (NFE) for different ODE solvers: fixed-step Euler integration with 100 and 1000 steps and adaptive integration (DOPRI5). We have run OT-CFM, OT-SI following our training procedure for $400,000$ steps. The four last rows report the results of our proposed methods OT-Harmonic and its anisotropic variant (OT-Aniso).

NFE / sample →	100	1000	Adaptive	
Algorithm ↓	FID	FID	FID	NFE
OT-CFM	4.730	3.901	3.706	**133.12**
OT-SI	**4.474**	4.077	4.003	163.84
OT-Harmonic, $\omega = 0.001$	4.656	**3.849**	**3.681**	**133.12**
OT-Harmonic, $\omega = 1$	4.741	3.963	3.890	143.36
OT-Harmonic, $\omega = \pi/2$	4.584	4.184	4.116	163.85
OT-Aniso	4.821	4.168	4.078	153.60

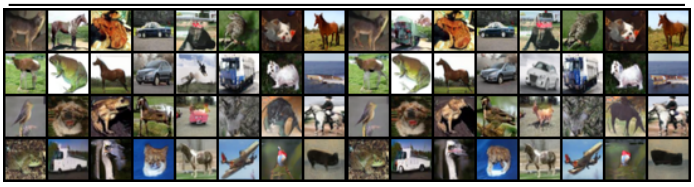


Figure 6: CIFAR-10 images generated by OT-CFM (left) and OT-Harmonic (right) with frequency $\omega = 0.001$.

5 Conclusion

We have proposed Lagrangian flow matching, a physics-based framework that selects flow matching probability paths and velocity fields by minimizing a Lagrangian action subject to the continuity equation and prescribed endpoints. The static–dynamic equivalence of Lagrangian optimal transport reduces this problem to a static OT problem with a closed-form transport cost, yielding a family of simulation-free training objectives that recover OT-based and conditional flow matching as the kinetic

special case and admit new constructions, such as harmonic flows, in the non-kinetic case. Several directions remain open. The choice of Lagrangian provides an inductive bias on the learned dynamics whose interaction with data geometry is not yet well understood; in particular, anisotropic or learned potentials V may align trajectories with low-dimensional data structure. Extending the framework to stochastic dynamics, in the spirit of Schrödinger bridges, is a natural next step.

Acknowledgments

This work was supported in part by the AWS Build on Trainium program and the Theta EdgeCloud computing resources. The authors gratefully acknowledge these providers for the computing infrastructure that enabled the experiments in this paper.

References

[1] M. S. Albergo, N. M. Boffi, and E. Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *Journal of Machine Learning Research*, 26(209):1–80, 2025.

[2] M. S. Albergo and E. Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations (ICLR)*, 2023.

[3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*. Birkhäuser, Basel, 2005.

[4] V. I. Arnold, K. Vogtmann, and A. Weinstein. *Mathematical Methods of Classical Mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer, New York, 2 edition, 1989.

[5] M. Balcerak, T. Amiranashvili, A. Terpin, S. Shit, L. Bogensperger, S. Kaltenbach, P. Koumoutsakos, and B. Menze. Energy matching: Unifying flow matching and energy-based models for generative modeling. *arXiv preprint arXiv:2504.10612*, 2025.

[6] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

[7] P. Bernard and B. Buffoni. Optimal mass transportation and Mather theory. *Journal of the European Mathematical Society*, 9(1):85–121, 2007.

[8] D. Burkhardt, J. Bloom, R. Cannoodt, M. D. Luecken, S. Krishnaswamy, C. Lance, A. O. Pisco, and F. J. Theis. Multimodal single-cell integration across time, individuals, and batches. In *NeurIPS 2022 Competition Track*, 2022.

[9] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

[10] Y. Chen, T. T. Georgiou, and M. Pavon. On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *Journal of Optimization Theory and Applications*, 169(2):671–691, 2016.

[11] V. De Bortoli, J. Thornton, J. Heng, and A. Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. In *Advances in Neural Information Processing Systems*, volume 34, pages 17695–17709, 2021.

[12] Y. Du and I. Mordatch. Implicit generation and modeling with energy based models. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

[13] W. Grathwohl, R. T. Q. Chen, J. Bettencourt, and D. Duvenaud. Scalable reversible generative models with free-form continuous dynamics. In *International Conference on Learning Representations (ICLR)*, 2019.

[14] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020.

[15] T. Koshizuka and I. Sato. Neural Lagrangian Schrödinger bridge: Diffusion modeling for population dynamics. *arXiv preprint arXiv:2204.04853*, 2022.

[16] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1–2):83–97, 1955.

[17] Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, and F. J. Huang. A tutorial on energy-based learning. In G. Bakir, T. Hofman, B. Schölkopf, A. Smola, and B. Taskar, editors, *Predicting Structured Data*. MIT Press, Cambridge, MA, 2006.

[18] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014.

[19] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations (ICLR)*, 2023.

[20] X. Liu, C. Gong, and Q. Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations (ICLR)*, 2023.

[21] K. R. Moon, D. van Dijk, Z. Wang, S. Gigante, D. B. Burkhardt, W. S. Chen, K. Yim, A. van den Elzen, M. J. Hirn, R. R. Coifman, et al. Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology*, 37(12):1482–1492, 2019.

[22] K. Neklyudov, R. Brekelmans, D. Severo, and A. Makhzani. Action matching: Learning stochastic dynamics from samples. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, volume 202 of *Proceedings of Machine Learning Research*, pages 25858–25889. PMLR, 2023.

[23] K. Neklyudov, R. Brekelmans, A. Tong, L. Atanackovic, Q. Liu, and A. Makhzani. A computational framework for solving Wasserstein Lagrangian flows. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, 2024.

[24] A.-A. Pooladian, H. Ben-Hamu, C. Domingo-Enrich, B. Amos, Y. Lipman, and R. T. Q. Chen. Multisample flow matching: Straightening flows with minibatch couplings. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, volume 202 of *Proceedings of Machine Learning Research*, pages 28100–28127. PMLR, 2023.

[25] A.-A. Pooladian, C. Domingo-Enrich, R. T. Q. Chen, and B. Amos. Neural optimal transport with Lagrangian costs. In *Proceedings of the 40th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2024.

[26] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*, volume 37 of *Proceedings of Machine Learning Research*, pages 2256–2265. PMLR, 2015.

[27] Y. Song, C. Durkan, I. Murray, and S. Ermon. Maximum likelihood training of score-based diffusion models. In *Advances in Neural Information Processing Systems*, volume 34, pages 1415–1428, 2021.

[28] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2021.

[29] A. Tong, K. Fatras, N. Malkin, G. Huguet, Y. Zhang, J. Rector-Brooks, G. Wolf, and Y. Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, 2024. Expert Certification.

[30] C. Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, Berlin, Heidelberg, 2009.

A Related Work

Lagrangian flow matching sits at the intersection of two lines of work: flow- and diffusion-based generative models built on prescribed probability paths, and variational or action-based approaches to learning transport dynamics. We discuss connections to representative works in each line.

A.1 Flow- and Diffusion-based Generative Models

Continuous Normalizing Flows and Neural ODEs [9, 13]. Continuous normalizing flows model the data distribution as the pushforward of a simple base distribution under the flow of a neural ODE $\dot{x} = v_\theta(x, t)$ [9, 13]. The induced density evolves according to the continuity equation, and likelihood-based training maximizes the change-of-variables log-likelihood of observed data, which requires solving the ODE forward and the adjoint backward at every gradient step. This formulation is expressive enough to represent a broad class of continuous probability paths [27], but the repeated ODE solves make training computationally expensive and difficult to scale. Flow matching, including our framework, retains the same deterministic ODE model at inference time but replaces likelihood-based training with a regression objective against a prescribed velocity field, which removes ODE solves from the training loop. Our contribution lies upstream of this regression step: rather than prescribing the probability path and velocity field by hand, we select them as the least-action solution of a Lagrangian, providing a principled mechanism for shaping the ODE trajectories that a continuous normalizing flow would otherwise be asked to learn directly from likelihoods.

Diffusion Models [26, 14, 28]. Diffusion models specify a forward stochastic process that progressively corrupts data into a tractable reference distribution, and learn the corresponding reverse process by score matching or denoising regression [26, 14]. The continuous-time formulation of [28] unifies these constructions through a stochastic differential equation, with sampling carried out by a reverse-time SDE or its associated probability-flow ODE. The probability path is implicitly prescribed by the forward noising schedule, and standard choices such as the variance-preserving and variance-exploding processes induce specific families of marginal paths between data and noise. Our framework relates to the deterministic side of this picture: the trigonometric path of Example 2.5 matches the marginal of the variance-preserving diffusion path with a cosine schedule [19, 14], and arises in our framework as the harmonic-oscillator least-action trajectory at $\omega = \pi/2$. This provides a least-action interpretation of the variance-preserving schedule at the level of the marginal probability path, although the full diffusion model additionally specifies a stochastic forward process that lies outside the deterministic continuity-equation framework developed here. Stochastic extensions, in the spirit of Schrödinger bridges (Appendix C), are a natural direction for connecting the two frameworks more closely.

Conditional, Rectified, and OT-based Flow Matching [19, 20, 29]. Flow matching trains a neural velocity field by regression against a target velocity associated with a prescribed probability path between p_{init} and p_{data}, avoiding the expensive ODE solves required by likelihood-based training of continuous normalizing flows [9, 13]. The framework is determined by the choice of probability path and its associated velocity field. [19] introduce conditional flow matching with affine Gaussian paths, in which conditioning on a target sample yields a closed-form linear interpolation and a tractable per-sample regression objective. [20] develop the rectified flow construction, which interpolates linearly between independently sampled endpoints and iteratively straightens the resulting trajectories. OT-CFM [29] replaces the independent endpoint coupling by a mini-batch optimal transport coupling under squared-Euclidean cost, producing straight-line trajectories along the matched pairs and improving sample efficiency at inference. These constructions share a common geometric structure: endpoints are matched by a coupling and connected by straight-line trajectories. Our framework extends this design space by replacing the straight-line trajectory with the least-action trajectory of a general Lagrangian, while preserving the simulation-free regression structure. The kinetic Lagrangian of Example 2.4 recovers OT-CFM as a special case, and the affine construction of [19] arises as its conditional analogue with $z = x_1$ and $p_{\text{init}} = \mathcal{N}(0, I)$. Non-kinetic Lagrangians yield curved least-action trajectories outside the affine and OT-CFM design space, with the harmonic family of Example 2.3 as a concrete instance.

Stochastic Interpolants [2, 1]. Stochastic interpolants construct a probability path by prescribing a pointwise interpolation $x_t = I(t, x_0, x_1) + \gamma(t)\, z$ between coupled endpoints (x_0, x_1), with an optional latent noise z governed by a schedule $\gamma(t)$. The interpolant I and schedule γ are chosen

by hand, typically as trigonometric or affine functions, and the velocity field is then defined as the conditional expectation of $\partial_t x_t$ given x_t. This framework unifies flow- and diffusion-based generative models under a common interpolation viewpoint and accommodates several specific path constructions, including affine and trigonometric variance-preserving interpolants, by direct specification of I and γ. Our framework differs in how the interpolant itself is selected. Rather than prescribing I directly, we obtain it as the least-action trajectory of a Lagrangian $\mathcal{L}(x, \dot{x}, t)$, with the endpoint coupling determined by the induced optimal transport problem. The trigonometric interpolant of Example 2.5 appears in the stochastic-interpolant framework as a specific hand-picked choice of I and γ; in our framework it instead naturally arises as the harmonic-oscillator least-action trajectory at $\omega = \pi/2$, embedded in a continuous one-parameter family $\{\gamma^{\omega}_{x_0,x_1}\}_{\omega\in(0,\pi)}$ that smoothly deforms the free-particle straight line into curved trajectories. More general Lagrangians produce interpolants outside the stochastic-interpolant design space, including paths whose curvature and endpoint coupling jointly reflect a prescribed potential. In this sense, the two frameworks are complementary: stochastic interpolants offer flexibility through the direct specification of I and γ together with stochastic forward processes, whereas Lagrangian flow matching provides a principled mechanism for selecting deterministic interpolants from a physical least-action principle, with stochastic extensions left to future work.

A.2 Action-based Methods for Learning Transport Dynamics

Several recent works have explored connections between optimal transport, least-action principles, and learning-based models for transport dynamics [22, 23, 25, 15]. These works share the use of variational or action-based principles to learn or infer transport dynamics. Our work follows this broad direction but preserves the standard flow-matching learning structure: once the training targets are constructed, the neural network simply learns to approximate the velocity field. The distinguishing feature is that the prescribed interpolation used in standard flow matching is replaced by one induced from a Lagrangian: the Lagrangian defines endpoint costs, the induced optimal transport problem selects the endpoint pairing, and the least-action principle determines the trajectory between paired samples. We discuss the connections to each of the four works above in turn.

Action Matching [22]. Action Matching learns continuous-time dynamics from samples of temporal marginal distributions $\{p_{t_k}\}_{k=1}^{K}$ along an absolutely continuous probability path. The method parameterizes a scalar action potential $s_\theta(x, t)$ and identifies its spatial gradient ∇s_θ with the velocity field that transports the prescribed marginals through the continuity equation $\partial_t p_t + \nabla \cdot (p_t \nabla s_\theta) = 0$, fitting θ through a variational objective on the potential. Two features distinguish this setup from ours. First, the probability path is prescribed: Action Matching observes intermediate marginals and recovers a velocity field consistent with them, whereas in Lagrangian flow matching only the endpoints p_{init} and p_{data} are given and the Lagrangian determines the entire path. Second, our least-action trajectories admit closed-form expressions in (x_0, x_1, t) and serve as explicit velocity targets in a standard flow-matching regression, in contrast to a variational objective on a scalar potential.

Wasserstein Lagrangian Flows [23]. Wasserstein Lagrangian Flows formulate action minimization directly on the space of probability densities: a density path $\{p_t\}_{t\in[0,1]}$ is selected by minimizing a Lagrangian action functional $\int_0^1 \mathcal{L}(p_t, \dot{p}_t, t)\, dt$ with marginal constraints at observed time points. Appropriate choices of kinetic and potential terms recover a wide range of transport problems as special cases, including standard optimal transport, Schrödinger bridges, unbalanced transport, and physically constrained transport. The corresponding computational problem is an infinite-dimensional optimization over density paths, solved through a Hamiltonian dual that introduces a co-state variable and reduces to a min–max saddle-point problem. Our framework is less general at the density-space level: we work with a separable classical Lagrangian $\mathcal{L}(x, v, t) = K(v) - V(x)$ at the sample-trajectory level. In exchange, the static–dynamic equivalence (Thm. 2.1) converts the action into a closed-form endpoint cost and explicit least-action trajectories between matched endpoints, so training reduces to the usual flow-matching regression form without min–max optimization or density-space dual variables.

Neural Optimal Transport with Lagrangian Costs [25]. Neural Optimal Transport with Lagrangian Costs studies the computational problem of optimal transport when the ground cost is itself defined by a least-action principle: $c_{\mathcal{L}}(x_0, x_1) = \inf_\gamma \int_0^1 \mathcal{L}(\gamma(t), \dot{\gamma}(t), t)\, dt$ subject to $\gamma(0) = x_0$ and $\gamma(1) = x_1$, allowing the resulting transport map and least-action paths to reflect obstacles, non-Euclidean

geometries, or other prior structure. The method parameterizes both the transport map and the least-action paths as neural networks and trains them jointly. This overlaps with the static component of our framework: the cost $c_{\mathcal{L}}$ in our Eq. (7) coincides with theirs in the time-independent case. The focus is different, however. Their work develops methods for *computing* Lagrangian optimal transport maps as the end product, with paths reported as a geometric byproduct. We instead use the optimal coupling and least-action trajectories as an *intermediate step* in constructing conditional probability paths and velocity targets for flow matching, with the simulation-free training objective in Eq. (16) as the end product. The static–dynamic equivalence (Thm. 2.1) lets us bypass the need for an explicit transport map altogether.

Neural Lagrangian Schrödinger Bridges [15]. Neural Lagrangian Schrödinger Bridges also use Lagrangian ideas to learn transport dynamics, but model stochastic rather than deterministic dynamics through a neural SDE $dX_t = b_\theta(X_t, t)\, dt + \sigma\, dW_t$ whose drift encodes a Lagrangian action with kinetic and potential terms. Training is formulated as a Schrödinger-bridge problem with marginal constraints at observed time points: parameters are fit by simulating the learned SDE and matching the resulting marginals to empirical samples, which requires forward integration of the SDE inside the training loop and is therefore not simulation-free in the flow-matching sense. Our framework differs along three axes: the underlying dynamics are deterministic and governed by the continuity equation, with a stochastic extension along the lines of Appendix C left to future work; the static–dynamic equivalence (Thm. 2.1) converts the Lagrangian action into a closed-form per-pair training target; and training reduces to a standard simulation-free regression objective on matched endpoint pairs, preserving the scalability that motivates flow matching.

B Connection to Energy-based Methods

We briefly discuss the relation between Lagrangian flow matching and energy-based methods. Classical energy-based models (EBMs) assign a scalar energy $E_\theta(x)$ to each state and define a distribution by favoring low-energy configurations [17, 12]; more recent work [5] unifies energy-based modeling and flow matching by learning a single scalar field that plays both energy and drift roles.

Our connection to energy-based modeling is structural rather than methodological. EBMs assign a preference to individual states; Lagrangians assign a preference to entire trajectories,

$$S[\gamma] \;=\; \int_0^1 \mathcal{L}\big(\gamma(t), \dot\gamma(t), t\big)\, dt,$$

lifting the same scalar-function-as-preference idea from the state space $\mathbb{R}^d$ to the path space $\mathrm{AC}([0,1];\mathbb{R}^d)$. In the separable case $\mathcal{L}(x, v, t) = K(v) - V(x)$, the kinetic term $K(v)$ penalizes rapid motion and the potential term $V(x)$ shapes which regions of state space are preferred along the path. From this viewpoint, the static–dynamic equivalence (Thm. 2.1) is the bridge that pushes a path-level preference down to a state-level transport problem: the optimal coupling and least-action trajectories between matched endpoints are jointly selected by $\mathcal{L}$.

This perspective suggests a natural extension of our framework: rather than prescribing the potential V, one could parameterize and learn it from data through a family of Lagrangians

$$\mathcal{L}_\theta(x, v, t) \;=\; K(v) - V_\theta(x, t),$$

where V_θ is a learnable potential. The induced least-action problem would then adapt the path-space action itself within the flow-matching framework, combining the simulation-free training of flow matching with the energy-learning viewpoint of EBMs. We leave a full treatment to future work.

C Connection to Schrödinger Bridges

The relation between Schrödinger bridges and optimal transport has been extensively explored [18, 10, 11, 29]. In particular, Schrödinger bridges can be viewed as a stochastic, entropy-regularized counterpart of optimal transport: in the Brownian reference case the dynamic Schrödinger problem involves stochastic transport with diffusion, whereas the optimal transport limit is governed by deterministic transport dynamics.

Given endpoint distributions p_0 and p_1, the Schrödinger bridge problem seeks a stochastic process whose initial and terminal marginals match these endpoints while remaining close, in relative entropy,

to a reference process. If P^{ref} denotes the path measure of a reference process, such as a Wiener process initialized from p_0, the problem can be written as

$$\inf_{P\,:\,(e_0)_\# P = p_0,\ (e_1)_\# P = p_1} \mathrm{KL}\big(P \,\|\, P^{\mathrm{ref}}\big)\,,$$

where e_t denotes the evaluation map at time t. Lagrangian flow matching, by contrast, is based on deterministic flow dynamics described by the continuity equation, and minimizes an action functional over probability paths and velocity fields,

$$\inf_{(p,v)} \int_0^1 \int_{\mathbb{R}^d} \mathcal{L}\big(x, v(x,t), t\big)\, p_t(x)\, dx\, dt,$$

subject to endpoint constraints $p_{t=0} = p_0$ and $p_{t=1} = p_1$. The two formulations select transport dynamics by different principles: a Schrödinger bridge minimizes a relative entropy on path measures, whereas Lagrangian flow matching minimizes a least-action functional on density–velocity pairs.

This comparison suggests a stochastic extension of Lagrangian flow matching. In place of the continuity equation, one could impose a Fokker–Planck equation associated with stochastic dynamics driven by drift and diffusion. Such an extension would replace deterministic transport by a stochastic counterpart, moving Lagrangian flow matching closer to Schrödinger-bridge and diffusion-based models.

D Proofs

This appendix collects the proofs of the two theoretical results in the main text. Appendix D.1 proves the static–dynamic equivalence (Theorem 2.1) underlying the Lagrangian flow matching framework; Appendix D.2 proves the conditional action bound (Proposition 3.1) underlying the conditional formulation in Sec. 3. Both proofs follow the standard optimal-transport playbook: the forward direction is established by constructing an admissible candidate from a static optimum and bounding its action via Jensen's inequality, while the reverse direction lifts an admissible density–velocity pair to a path measure via the superposition principle of [3].

D.1 Proof of Theorem 2.1.

The proof proceeds in two steps. We first construct an admissible density–velocity pair from an optimal static coupling and show that its action is bounded by $\mathcal{A}_{\mathrm{stat}}$, giving $\mathcal{A}_{\mathrm{dyn}} \leqslant \mathcal{A}_{\mathrm{stat}}$. We then show the reverse inequality by lifting any admissible (p, v) to a path measure via the superposition principle and bounding its action below by $\mathcal{A}_{\mathrm{stat}}$. The construction of the forward direction is encapsulated in the following lemma.

Lemma D.1. *Let π^* be an optimal plan of* (8)*, and let*

$$\eta^* := (\gamma_{x_0,x_1})_\# \pi^*.$$

Define

$$p^*(\cdot, t) := (e_t)_\# \eta^*,$$
$$v^*(x,t) := \mathbb{E}_{\gamma\sim\eta^*}\big[\dot\gamma(t) \mid \gamma(t) = x\big].$$

Then $(p^*, v^*) \in \mathfrak{D}$.

Proof. First, since $\pi^* \in \Pi(p_{\mathrm{init}}, p_{\mathrm{data}})$, we have

$$p^*(\cdot, 0) = (e_0)_\# \eta^* = p_{\mathrm{init}}, \qquad p^*(\cdot, 1) = (e_1)_\# \eta^* = p_{\mathrm{data}}.$$

It remains to verify the continuity equation in the sense of distributions. Let $\varphi \in C_c^\infty(\mathbb{R}^d \times (0,1))$. Since $p^*(\cdot, t) = (e_t)_\# \eta^*$, we have

$$\int_{\mathbb{R}^d} \varphi(x,t) p^*(x,t)\, dx = \int_{AC([0,1];\mathbb{R}^d)} \varphi(\gamma(t), t)\, d\eta^*(\gamma).$$

Hence

$$\int_0^1\int_{\mathbb{R}^d} \partial_t\varphi(x,t)p^*(x,t)\,dx\,dt = \int_0^1\int_{AC([0,1];\mathbb{R}^d)} \partial_t\varphi(\gamma(t),t)\,d\eta^*(\gamma)\,dt.$$

Since η^* is concentrated on absolutely continuous curves, the map $t \mapsto \varphi(\gamma(t),t)$ is absolutely continuous for η^*-a.e. γ, and

$$\frac{d}{dt}\varphi(\gamma(t),t) = \partial_t\varphi(\gamma(t),t) + \nabla\varphi(\gamma(t),t)\cdot\dot\gamma(t)$$

for a.e. $t \in (0,1)$. Therefore,

$$\int_0^1 \frac{d}{dt}\varphi(\gamma(t),t)\,dt = \varphi(\gamma(1),1) - \varphi(\gamma(0),0).$$

Since φ has compact support in $(0,1)$ in time, the right-hand side vanishes. Integrating over η^*, we obtain

$$\int_0^1\int_{AC([0,1];\mathbb{R}^d)} \partial_t\varphi(\gamma(t),t)\,d\eta^*(\gamma)\,dt = -\int_0^1\int_{AC([0,1];\mathbb{R}^d)} \nabla\varphi(\gamma(t),t)\cdot\dot\gamma(t)\,d\eta^*(\gamma)\,dt.$$

By definition,

$$p^*(\cdot,t) = (e_t)_\#\eta^*,$$

so for any measurable vector field $F : \mathbb{R}^d \to \mathbb{R}^d$,

$$\int_{AC([0,1];\mathbb{R}^d)} F(\gamma(t))\cdot\dot\gamma(t)\,d\eta^*(\gamma) = \int_{\mathbb{R}^d} F(x)\cdot\mathbb{E}_{\gamma\sim\eta^*}\big[\dot\gamma(t)\mid\gamma(t)=x\big]\,p^*(x,t)\,dx.$$

Taking

$$F(x) = \nabla\varphi(x,t)$$

and using the definition

$$v^*(x,t) = \mathbb{E}_{\gamma\sim\eta^*}\big[\dot\gamma(t)\mid\gamma(t)=x\big],$$

we obtain

$$\int_{AC([0,1];\mathbb{R}^d)} \nabla\varphi(\gamma(t),t)\cdot\dot\gamma(t)\,d\eta^*(\gamma) = \int_{\mathbb{R}^d} \nabla\varphi(x,t)\cdot v^*(x,t)p^*(x,t)\,dx.$$

Therefore

$$\int_0^1\int_{\mathbb{R}^d} \Big(\partial_t\varphi(x,t) + \nabla\varphi(x,t)\cdot v^*(x,t)\Big)p^*(x,t)\,dx\,dt = 0,$$

which proves

$$\partial_t p^* + \operatorname{div}(v^*p^*) = 0$$

in the sense of distributions. Hence $(p^*,v^*) \in \mathfrak{D}$. □

Proof of Theorem 2.1. We first prove $\mathcal{A}_{\mathrm{dyn}} \leqslant \mathcal{A}_{\mathrm{stat}}$.

Let π^* be an optimal plan of (8). Since the minimizer γ_{x_0,x_1} of (6) is unique and the map is measurable, we can define the pushforward probability measure on the path space:

$$\eta^* := (\gamma_{x_0,x_1})_\#\pi^*.$$

For each $t \in [0,1]$, define

$$p^*(\cdot,t) := (e_t)_\#\eta^*.$$

Then

$$p^*(\cdot,0) = p_{\mathrm{init}}, \qquad p^*(\cdot,1) = p_{\mathrm{data}}.$$

Since η^* is concentrated on absolutely continuous curves, $\dot\gamma(t)$ exists for η^*-a.e. γ and a.e. t. We define the Eulerian velocity field v^* by

$$v^*(x,t) := \mathbb{E}_{\gamma\sim\eta^*}\big[\dot\gamma(t)\mid\gamma(t)=x\big].$$

By Lemma D.1, the pair (p^*,v^*) satisfies

$$\partial_t p^* + \operatorname{div}(v^*p^*) = 0$$

in the sense of distributions, and therefore $(p^*, v^*) \in \mathfrak{D}$.

We now estimate the action of (p^*, v^*). Since $\mathcal{L}(x, \cdot, t)$ is convex, Jensen's inequality yields

$$\mathcal{L}(x, v^*(x,t), t) = \mathcal{L}\big(x, \mathbb{E}_{\gamma\sim\eta^*}[\dot\gamma(t) \mid \gamma(t) = x], t\big) \leqslant \mathbb{E}_{\gamma\sim\eta^*}[\mathcal{L}(x, \dot\gamma(t), t) \mid \gamma(t) = x]$$

Multiplying by $p^*(x,t)$ and integrating in x, we obtain

$$\int_{\mathbb{R}^d} \mathcal{L}(x, v^*(x,t), t)\, p^*(x,t)\, dx \leqslant \int_{\mathbb{R}^d} \mathbb{E}_{\gamma\sim\eta^*}[\mathcal{L}(x, \dot\gamma(t), t) \mid \gamma(t) = x]\, p^*(x,t)\, dx.$$

By the defining property of conditional expectation,

$$\int_{\mathbb{R}^d} \mathbb{E}_{\gamma\sim\eta^*}[\mathcal{L}(x, \dot\gamma(t), t) \mid \gamma(t) = x]\, p^*(x,t)\, dx = \int_{AC([0,1];\mathbb{R}^d)} \mathcal{L}(\gamma(t), \dot\gamma(t), t)\, d\eta^*(\gamma).$$

Therefore,

$$\int_{\mathbb{R}^d} \mathcal{L}(x, v^*(x,t), t)\, p^*(x,t)\, dx \leqslant \int_{AC([0,1];\mathbb{R}^d)} \mathcal{L}(\gamma(t), \dot\gamma(t), t)\, d\eta^*(\gamma).$$

Integrating over $t \in [0,1]$, it follows that

$$\int_0^1 \int_{\mathbb{R}^d} \mathcal{L}(x, v^*(x,t), t)\, p^*(x,t)\, dx\, dt \leqslant \int_{AC([0,1];\mathbb{R}^d)} \left(\int_0^1 \mathcal{L}(\gamma(t), \dot\gamma(t), t)\, dt \right) d\eta^*(\gamma).$$

By construction of η^*, for π^*-a.e. (x_0, x_1), γ_{x_0,x_1} minimizes (6), so

$$\int_0^1 \mathcal{L}(\gamma_{x_0,x_1}(t), \dot\gamma_{x_0,x_1}(t), t)\, dt = c_{\mathcal{L}}(x_0, x_1).$$

Therefore

$$\int_{AC([0,1];\mathbb{R}^d)} \left(\int_0^1 \mathcal{L}(\gamma(t), \dot\gamma(t), t)\, dt \right) d\eta^*(\gamma) = \int_{\mathbb{R}^d\times\mathbb{R}^d} c_{\mathcal{L}}(x_0, x_1)\, d\pi^*(x_0, x_1).$$

Since π^* is optimal for (8),

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} c_{\mathcal{L}}(x_0, x_1)\, d\pi^*(x_0, x_1) = \mathcal{A}_{\text{stat}}.$$

Thus

$$\int_0^1 \int_{\mathbb{R}^d} \mathcal{L}(x, v^*(x,t), t)\, p^*(x,t)\, dx\, dt \leqslant \mathcal{A}_{\text{stat}},$$

and hence $\mathcal{A}_{\text{dyn}} \leqslant \mathcal{A}_{\text{stat}}$.

We next prove the reverse inequality $\mathcal{A}_{\text{dyn}} \geqslant \mathcal{A}_{\text{stat}}$.

Let $(p, v) \in \mathfrak{D}$ have finite action. Since V is controlled by the kinetic term in the sense stated above, finite action implies finite kinetic energy. Therefore, using the positive definiteness of K,

$$\mathbb{E}_{(x,t)\sim p}\|v(x,t)\|^2 \leqslant C\mathbb{E}_{(x,t)\sim p} K(v(x,t)) < \infty.$$

By [3, Theorem 8.2.1], there exists a probability measure η concentrated on $AC([0,1];\mathbb{R}^d)$ such that

$$(e_t)_\#\eta = p(\cdot, t) \quad \text{for all } t \in [0,1], \qquad \dot\gamma(t) = v(\gamma(t), t) \quad \eta\text{-a.e. } \gamma$$

Define $\pi := (e_0, e_1)_\#\eta \in \Pi(p_{\text{init}}, p_{\text{data}})$. Then

$$\int_0^1 \int_{\mathbb{R}^d} \mathcal{L}(x, v(x,t), t) p(x,t)\, dx\, dt = \int_{AC([0,1];\mathbb{R}^d)} \left(\int_0^1 \mathcal{L}(\gamma(t), \dot\gamma(t), t)\, dt \right) d\eta(\gamma).$$

By definition of the least-action cost,

$$\int_0^1 \mathcal{L}(\gamma(t), \dot\gamma(t), t)\, dt \geqslant c_{\mathcal{L}}(\gamma(0), \gamma(1)).$$

Therefore

$$\int_0^1 \int_{\mathbb{R}^d} \mathcal{L}(x, v(x,t), t) p(x,t)\, dx\, dt \geqslant \int_{\mathbb{R}^d\times\mathbb{R}^d} c_{\mathcal{L}}(x_0, x_1)\, d\pi(x_0, x_1) \geqslant \mathcal{A}_{\text{stat}}.$$

Taking the infimum over all admissible pairs $(p, v) \in \mathfrak{D}$ gives

$$\mathcal{A}_{\text{dyn}} \geqslant \mathcal{A}_{\text{stat}}.$$

This completes the proof.

□

D.2 Proof of Proposition 3.1

Proof. The proof has two steps. We first show that the marginal pair (p_t, v_t) induced by any admissible conditional family is itself admissible for the unconditional problem; the inequality then follows from Jensen applied to the convexity of $\mathcal{L}(x, \cdot, t)$.

Let $\{p_t(\cdot \mid z), v_t(\cdot \mid z)\}_z$ be any admissible conditional family, and let (p_t, v_t) be the induced marginal path and velocity field defined by

$$p_t(x) = \int p_t(x \mid z)\, q(dz), \qquad v_t(x) = \int v_t(x \mid z) \frac{p_t(x \mid z)}{p_t(x)}\, q(dz). \tag{19}$$

Integrating the continuity equations for $p_t(\cdot|z)$ over z, we obtain

$$\partial_t p_t(x) + \nabla \cdot \big(v_t(x) p_t(x)\big) = 0. \tag{20}$$

Moreover, the endpoint mixture conditions (11) imply

$$p_0 = p_{\text{init}}, \qquad p_1 = p_{\text{data}}.$$

Thus, the marginal pair $(p_t, v_t) \in \mathfrak{D}$ is admissible for the unconditional Lagrangian flow matching problem; see Section 2.1.

Since $\mathcal{L}_z = \mathcal{L}$ and $\mathcal{L}(x, \cdot, t)$ is convex, Jensen's inequality gives, for $p_t(x) > 0$,

$$\mathcal{L}\big(x, v_t(x), t\big) \leqslant \int \mathcal{L}\big(x, v_t(x \mid z), t\big) \frac{p_t(x \mid z)}{p_t(x)}\, q(dz).$$

Multiplying by $p_t(x)$ and integrating over x and t, we obtain

$$\int_0^1 \int_{\mathbb{R}^d} \mathcal{L}\big(x, v_t(x), t\big) p_t(x)\, dx\, dt \leqslant \mathbb{E}_{z\sim q} \int_0^1 \int_{\mathbb{R}^d} \mathcal{L}\big(x, v_t(x \mid z), t\big) p_t(x \mid z)\, dx\, dt.$$

Since (p_t, v_t) is admissible for the unconditional least-action problem,

$$\mathcal{A}_{\text{dyn}} \leqslant \int_0^1 \int_{\mathbb{R}^d} \mathcal{L}\big(x, v_t(x), t\big) p_t(x)\, dx\, dt.$$

Combining the two inequalities and then taking the infimum over admissible conditional families yields

$$\mathcal{A}_{\text{dyn}} \leqslant \mathcal{A}_{\text{cond}}.$$

Finally, if the unconditional problem admits an optimizer (p_t^*, v_t^*), then equality is attained by the trivial conditioning structure, namely by taking z to be constant and setting

$$p_t(\cdot \mid z) = p_t^*, \qquad v_t(\cdot \mid z) = v_t^*.$$

This proves the claim. □

E Mini-batch Conditional Harmonic Flow Matching

The following examples illustrate how the Lagrangian $\mathcal{L}$ and the batch size n shape the resulting algorithm: the harmonic Lagrangian preserves the OT coupling but curves the training trajectories, mini-batch OT-CFM is recovered as the $\omega \to 0$ kinetic limit, and the batch size $n = 1$ case collapses to independent-coupling flow matching.

Example E.1 (Mini-batch harmonic flow)**.** The harmonic Lagrangian $\mathcal{L}_\omega$ of Example 2.2 yields the harmonic trajectory (4) between matched endpoints, with cost

$$c_\omega(x_0^{(i)}, x_1^{(j)}) \;=\; \frac{\omega}{2\sin\omega}\big[\cos\omega\,(\|x_0^{(i)}\|^2 + \|x_1^{(j)}\|^2) - 2\, x_0^{(i)} \cdot x_1^{(j)}\big].$$

The empirical marginals fix $\sum_i \pi_{ij} \|x_0^{(i)}\|^2$ and $\sum_j \pi_{ij} \|x_1^{(j)}\|^2$ for any feasible π, so

$$\sum_{i,j} \pi_{ij}\, c_\omega(x_0^{(i)}, x_1^{(j)}) \;=\; -\frac{\omega}{\sin\omega} \sum_{i,j} \pi_{ij}\, x_0^{(i)} \cdot x_1^{(j)} \;+\; C(z),$$

with $C(z)$ independent of π. By polarization, this is equivalent to minimizing the quadratic cost $\sum_{i,j} \pi_{ij} \frac{1}{2}\|x_0^{(i)} - x_1^{(j)}\|^2$, so the discrete coupling (17) remains the quadratic-cost mini-batch OT — only the training trajectory in (18) changes. The curvature of the matched-pair trajectories visible across all panels of Fig. 4 is the harmonic signature: it is set by ω and is independent of n. Additionally, panels 4b–4e of Fig. 4 show the intermediate regime, where increasing n tightens the empirical coupling toward π^* while ω continues to control trajectory curvature. ■

Example E.2 (Mini-batch OT-CFM as the $\omega \to 0$ limit)**.** Taking the frequency to limit $\omega \to 0$ in Example E.1 recovers the kinetic Lagrangian $\mathcal{L} = \frac{1}{2}\|v\|^2$ of Example 2.1, with quadratic cost $c_{\mathcal{L}}(x_0, x_1) = \frac{1}{2}\|x_0 - x_1\|^2$ and affine trajectory $\gamma_{x_0,x_1}(t) = (1-t)x_0 + tx_1$. Both (17) and (18) then reduce to mini-batch OT-CFM [29, 24]. ■

Example E.3 (Independent-coupling flow matching as the $n = 1$ case)**.** The other degenerate limit is $n = 1$, where the discrete OT problem is trivial: the conditional coupling reduces to the independent product $\delta_{x_0^{(1)}} \otimes \delta_{x_1^{(1)}}$, and marginalizing over z recovers the unconditional objective (10) with the independent coupling $p_{\text{init}} \otimes p_{\text{data}}$. Fig. 4a illustrates this regime: every source mode emits trajectories spanning both target moons, the many-to-many routing characteristic of independent coupling. Additionally, with the kinetic Lagrangian, this further reduces to the straight-line conditional-flow-matching objective of [19]. ■

Mini-batch harmonic flow matching thus reuses the OT-CFM solver while replacing straight-line targets by curved least-action targets controlled by ω. We exploit this in Sec. 4, where we use the same mini-batch OT solver as OT-CFM and vary only ω (and the anisotropic potential A in Appendix F) to study the effect of the Lagrangian on the learned dynamics.

F Anisotropic Harmonic Flow Matching

This appendix develops the anisotropic harmonic case in detail and reports synthetic experiments. The construction generalizes the isotropic harmonic Lagrangian of Example 2.2 by replacing the scalar frequency ω with a symmetric positive-definite matrix $A \in \mathbb{R}^{d\times d}$, allowing different curvatures along different directions in $\mathbb{R}^d$.

Lagrangian, trajectory, and cost Let $A \in \mathbb{R}^{d\times d}$ be symmetric positive-definite with spectral decomposition $A = Q\Lambda Q^\top$ and $\Lambda = \operatorname{diag}(\omega_1^2, \dots, \omega_d^2)$, $\omega_k \in (0, \pi)$. Define the anisotropic harmonic Lagrangian

$$\mathcal{L}_A(x, v) \;=\; \tfrac{1}{2}\|v\|^2 - \tfrac{1}{2}\, x^\top A x. \tag{21}$$

The Euler–Lagrange equation $\ddot{\gamma} + A\gamma = 0$ decouples under the change of variables $\tilde{\gamma} = Q^\top \gamma$ into d scalar oscillators, one per eigendirection. Solving each scalar problem as in Example 2.2 and transforming back gives the closed-form least-action trajectory

$$\gamma^A_{x_0,x_1}(t) \;=\; \frac{\sin\big((1-t)\sqrt{A}\big)}{\sin\sqrt{A}}\, x_0 + \frac{\sin\big(t\sqrt{A}\big)}{\sin\sqrt{A}}\, x_1, \tag{22}$$

where the matrix functions are defined by $f(A) = Q\, f(\Lambda)\, Q^\top$ applied entrywise to the eigenvalues. The velocity is

$$\dot{\gamma}^A_{x_0,x_1}(t) \;=\; -\frac{\sqrt{A}\cos\big((1-t)\sqrt{A}\big)}{\sin\sqrt{A}}\, x_0 + \frac{\sqrt{A}\cos\big(t\sqrt{A}\big)}{\sin\sqrt{A}}\, x_1. \tag{23}$$

Both reduce to (4) when $A = \omega^2 I$, and to the affine interpolation as $A \to 0$.

Cost. Substituting (22)–(23) into the action and using the per-mode calculation from Example 2.3 gives

$$c_A(x_0, x_1) \;=\; \tfrac{1}{2}\, x_0^\top \Phi(A)\, x_0 + \tfrac{1}{2}\, x_1^\top \Phi(A)\, x_1 - x_0^\top \Psi(A)\, x_1, \tag{24}$$

where

$$\Phi(A) \;:=\; \sqrt{A}\cot\sqrt{A}, \qquad \Psi(A) \;:=\; \sqrt{A}\,\csc\sqrt{A}, \tag{25}$$

both symmetric positive-definite for the eigenvalue range $\omega_k \in (0, \pi)$. The isotropic limit $A = \omega^2 I$ recovers $\Phi(A) = \omega\cot\omega \cdot I$ and $\Psi(A) = (\omega/\sin\omega)\, I$, matching the prefactors in Example 2.3.

Optimal coupling: a generalized OT problem A salient feature of the anisotropic case is that the polarization argument of Example 2.3 no longer reduces the static problem to quadratic-cost OT. For any $\pi \in \Pi(p_{\text{init}}, p_{\text{data}})$,

$$\int x_0^\top \Phi(A) x_0 \, d\pi = \int x_0^\top \Phi(A) x_0 \, p_{\text{init}}(dx_0) \;=\; \operatorname{tr}\big(\Phi(A)\,\Sigma_0\big) + \mu_0^\top \Phi(A) \mu_0,$$

where μ_0 and Σ_0 are the mean and covariance of p_{init}, and analogously for x_1 under p_{data}. Both quadratic terms in (24) are therefore π-independent, so

$$\int c_A(x_0, x_1)\, d\pi \;=\; -\int x_0^\top \Psi(A)\, x_1 \, d\pi \;+\; \text{const.} \tag{26}$$

Minimizing over π is equivalent to maximizing the *generalized cross-correlation* $\int x_0^\top \Psi(A) x_1 \, d\pi$. By the polarization identity in the metric induced by $\Psi(A)$, $x_0^\top \Psi(A) x_1 = \frac{1}{2}\big(x_0^\top \Psi(A) x_0 + x_1^\top \Psi(A) x_1 - (x_0 - x_1)^\top \Psi(A)(x_0 - x_1)\big)$, this is in turn equivalent to the weighted-quadratic OT problem

$$\pi_A^* \;\in\; \underset{\pi \in \Pi(p_{\text{init}}, p_{\text{data}})}{\operatorname{argmin}} \int \tfrac{1}{2}\,(x_0 - x_1)^\top \Psi(A)\,(x_0 - x_1)\, d\pi(x_0, x_1). \tag{27}$$

Equivalently, π_A^* is the standard quadratic OT plan between the pushforwards of p_{init} and p_{data} under the linear map $x \mapsto \Psi(A)^{1/2} x$. Unlike the isotropic case, π_A^* generally depends on A when $\Psi(A)$ is not a scalar multiple of identity, so anisotropic harmonic flow matching does not in general reuse the OT-CFM coupling. The mini-batch approximation (17) carries over by replacing $\frac{1}{2}\|x_0 - x_1\|^2$ with $\frac{1}{2}(x_0 - x_1)^\top \Psi(A)(x_0 - x_1)$, and is solved by the same linear-assignment routines.

Choice of A via PCA. We parameterize A from a principal component analysis of the data. Let $\Sigma_{\text{data}} = U\,\Lambda_{\text{data}}\,U^\top$ with $\Lambda_{\text{data}} = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$ and $\lambda_1 \geqslant \cdots \geqslant \lambda_d > 0$. We align the eigenframe of A with the PCA frame U and assign per-mode frequencies as a monotonically decreasing function of the principal variances:

$$A \;=\; U \operatorname{diag}\big(\omega_1^2, \ldots, \omega_d^2\big)\, U^\top, \qquad \omega_k \;=\; \omega_{\max} \left(\frac{\lambda_d}{\lambda_k}\right)^{\alpha}, \tag{28}$$

with hyperparameters $\omega_{\max} \in (0, \pi)$ and $\alpha \geqslant 0$. This yields small ω_k along high-variance principal directions — where data already provide a strong signal and the trajectory should remain close to the affine interpolant — and larger ω_k along low-variance directions, where the harmonic potential more aggressively contracts deviations from the geodesic. The eigenvalue constraint $\omega_k \in (0, \pi)$ is automatically satisfied by the choice $\omega_{\max} < \pi$. The exponent α controls anisotropy: $\alpha = 0$ recovers the isotropic harmonic case $A = \omega_{\max}^2 I$ of Example 2.3, while $\alpha = \frac{1}{2}$ corresponds to $A \propto \Sigma_{\text{data}}^{-1}$ rescaled to fit the spectral cap, recovering the covariance-aligned parameterization discussed informally above. In practice, we estimate U and $\{\lambda_k\}$ from a held-out subset of p_{data} and freeze A during training; if the ambient dimension is large, we apply this construction on the top-r principal subspace and set $\omega_k = \omega_{\max}$ for the orthogonal complement, so that A acts only on the directions where data anisotropy is statistically meaningful.

Synthetic experiments. Fig. 7 visualizes the OT-Aniso flow on a 2D toy benchmark in which eight Gaussian source modes (olive) surround a two-moons target (blue). Olive curves show learned sample trajectories integrated forward from each source cluster to the target. The anisotropic harmonic potential bends each trajectory along the data-subspace directions of largest variance, producing the characteristic outward-then-inward sweep visible at every source mode. We further evaluate OT-Aniso on CIFAR-10 image generation in Table 3 and Fig. 9. The anisotropic variant OT-Aniso achieves FID $= 4.078$ at adaptive integration, no better than OT-SI and worse than every harmonic and OT-CFM variant. We attribute this to the high-dimensional and largely uncorrelated structure of CIFAR-10 pixel data: when most coordinates are approximately independent, the rotation recovered from a small fit batch carries little geometric signal beyond what an isotropic schedule already provides, and the per-direction frequency assignment confers no meaningful advantage. Learning the potential A directly from data, rather than fitting it from a single batch of empirical variances, is a natural direction for future work.

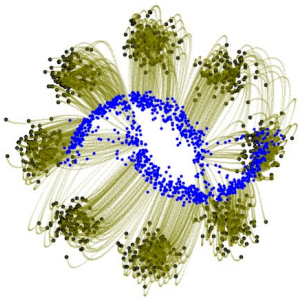


Figure 7: OT-Aniso flow on the 8-Gaussians→two-moons benchmark.

G Derivation of the normalized path energy

This appendix supplies the closed forms behind the diagnostic NPE_ω used in Sec. 4: the harmonic geodesic between two endpoints, its kinetic energy in closed form, the resulting harmonic-OT cost $C_\omega(\pi_0, \pi_1)$, and a decomposition of $\mathrm{NPE}_\omega - 1$ into a coupling term and a path term that isolate distinct sources of deviation.

Harmonic geodesics. A harmonic geodesic at frequency ω is a smooth path $\phi : [0,1] \to \mathbb{R}^d$ satisfying $\ddot{\phi}_t + \omega^2 \phi_t = 0$. The general solution $\phi_t = a\cos(\omega t) + b\sin(\omega t)$, specialized to the boundary conditions $\phi_0 = x_0$ and $\phi_1 = x_1$, gives, for $\omega \in (0, \pi)$,

$$\phi_t^\star(x_0, x_1) \;=\; \cos(\omega t)\, x_0 + \sin(\omega t)\, B, \qquad B \;:=\; \frac{x_1 - \cos(\omega)\, x_0}{\sin(\omega)}. \tag{29}$$

Differentiating yields $\dot{\phi}_t^\star = -\omega \sin(\omega t)\, x_0 + \omega \cos(\omega t)\, B$.

Per-pair kinetic energy. The kinetic energy along $\phi^\star$ is $k_\omega(x_0, x_1) = \int_0^1 \frac{1}{2} \|\dot{\phi}_t^\star\|^2 \, dt$. Expanding the square gives

$$\|\dot{\phi}_t^\star\|^2 = \omega^2 \Big[\|x_0\|^2 \sin^2(\omega t) + \|B\|^2 \cos^2(\omega t) - 2\,(x_0 \cdot B)\, \sin(\omega t) \cos(\omega t) \Big]. \tag{30}$$

Integrating each term on $[0,1]$ produces the trigonometric integrals

$$I_{ss}(\omega) = \tfrac{1}{2} - \tfrac{\sin(2\omega)}{4\omega}, \qquad I_{cc}(\omega) = \tfrac{1}{2} + \tfrac{\sin(2\omega)}{4\omega}, \qquad I_{sc}(\omega) = \tfrac{\sin^2(\omega)}{2\omega}, \tag{31}$$

yielding the closed form

$$k_\omega(x_0, x_1) \;=\; \tfrac{\omega^2}{2} \Big[\|x_0\|^2\, I_{ss}(\omega) + \|B\|^2\, I_{cc}(\omega) - 2\,(x_0 \cdot B)\, I_{sc}(\omega) \Big]. \tag{32}$$

Equation (32) can be evaluated directly from (x_0, x_1) via B; we verify it numerically against Simpson-rule integration of $\|\dot{\phi}^\star\|^2$ to $\sim 10^{-15}$ relative error.

Limiting cases. Two limits provide useful sanity checks. As $\omega \to 0$, $B \to x_1 - x_0$, $I_{ss} \to 0$, $I_{cc} \to 1$, and $I_{sc} \to 0$, so

$$k_\omega(x_0, x_1) \;\xrightarrow{\omega \to 0}\; \tfrac{1}{2} \|x_1 - x_0\|^2, \tag{33}$$

recovering the linear-interpolant kinetic energy: harmonic geodesics degenerate to straight lines, and $C_\omega \to W_2^2/2$. As $\omega \to \pi^-$, $\sin(\omega) \to 0$ and the B-term diverges unless $x_1 = -x_0$, reflecting the loss of geodesic uniqueness at the period boundary; (32) is therefore restricted to $\omega \in (0, \pi)$.

Harmonic-OT cost. The harmonic transport cost between π_0 and π_1 is the infimum of expected per-pair kinetic energy over couplings,

$$C_\omega(\pi_0, \pi_1) \;=\; \inf_{\pi \in \Pi(\pi_0, \pi_1)} \mathbb{E}_{(x_0, x_1) \sim \pi}\, k_\omega(x_0, x_1). \tag{34}$$

Substituting (32) and using $B = (x_1 - \cos(\omega)\, x_0)/\sin(\omega)$, the cross-term $(x_0 \cdot B) = (\langle x_0, x_1 \rangle - \cos(\omega)\|x_0\|^2)/\sin(\omega)$ is the only coupling-dependent quantity (the marginal moments $\mathbb{E}\|x_0\|^2, \mathbb{E}\|x_1\|^2$ are fixed). Hence the optimal coupling for C_ω is the same coupling that maximizes $\mathbb{E}\langle x_0, x_1 \rangle$, which is the standard W_2 coupling on centered measures. We estimate C_ω on minibatches by solving the linear assignment problem with cost matrix $M_{ij} = k_\omega(x_0^{(i)}, x_1^{(j)})$ via the Hungarian algorithm [16].

Decomposition of NPE_ω. Let $\hat{\pi}_\theta$ denote the coupling induced by the learned flow (each x_0 paired with the endpoint $\phi_1(x_0)$ produced by ODE integration of v_θ), and $\pi^\star$ the harmonic-OT coupling. Define the model-coupling action $\bar{k}_\theta = \mathbb{E}_{(x_0, x_1) \sim \hat{\pi}_\theta}\, k_\omega(x_0, x_1)$. Then

$$K[\phi] - C_\omega(\pi_0, \pi_1) \;=\; \underbrace{\bar{k}_\theta - C_\omega(\pi_0, \pi_1)}_{\text{coupling excess}} \;+\; \underbrace{K[\phi] - \bar{k}_\theta}_{\text{path excess}}. \tag{35}$$

The coupling excess is non-negative by definition of the OT infimum and vanishes iff $\hat{\pi}_\theta = \pi^\star$; the path excess is signed and vanishes iff each model trajectory traces the harmonic geodesic between its endpoints. Both vanish iff ϕ realizes the harmonic-OT geodesic flow at frequency ω, in which case $\mathrm{NPE}_\omega = 0$. The decomposition lets us attribute observed NPE deviations to either an incorrect pairing (e.g. the independent coupling (x_0, z) used by vanilla harmonic flow) or non-geodesic trajectories (e.g. rectified flow).

Estimation. We integrate the model ODE $\dot{\phi}_t = v_\theta(\phi_t, t)$ with a fixed-step RK4 scheme at $N = 200$ steps, recording $\frac{1}{2}\|v_\theta(\phi_t, t)\|^2$ on the same grid, and compute $K[\phi]$ from these samples by composite Simpson's rule. The harmonic-OT cost C_ω in (34) is estimated by solving the linear assignment problem with cost matrix $M_{ij} = k_\omega(x_0^{(i)}, x_1^{(j)})$ on a fixed minibatch of size $n = 512$, drawn independently of the RK4 evaluation batch. The coupling and path components of (35) are computed on the RK4 batch using the model-induced pairs $(x_0,\ \phi_1(x_0))$. We report all NPE-based diagnostics at the reference frequency $\omega = 1$, and aggregate as mean $\pm$ standard deviation across five independent training runs (different seeds for data sampling, model initialization, and optimization).

H Experimental Setup

This appendix gives the experimental details deferred from Sec. 4: training protocols, evaluation procedures, and supplementary results for the three tracks (2D synthetic, single-cell trajectory interpolation, and CIFAR-10 image generation). Common settings shared across tracks are stated once in Appendix H.1 and not repeated.

H.1 Shared protocol

Framework and architectures. All experiments build on the conditional flow matching framework of [29]: model architectures, optimizer settings, training schedules, dataset preprocessing, and evaluation harnesses follow that codebase, with the harmonic conditional flow matchers swapping in for the OT-CFM matcher where appropriate. The 2D and single-cell tracks use a time-conditioned MLP; the CIFAR-10 track uses the U-Net configuration of [29]. We refer to [29] for hyperparameter values; only deviations or additions are listed below.

Methods. We compare six conditional flow matching variants. All share architecture, optimizer, training-step count, and minibatch size within a given track; differences arise solely from the conditional probability path and source–target coupling.

- **OT-CFM** [29]: linear interpolant $x_t = (1-t)\,x_0 + t\,x_1$ with the exact-OT minibatch coupling on the squared-Euclidean cost.
- **OT-SI** [2]: variance-preserving (cosine-schedule) interpolant with the exact-OT coupling on the squared-Euclidean cost.
- **OT-Harmonic** ($\omega \in \{10^{-3},\ 1,\ \pi/2\}$): the isotropic harmonic conditional flow of Sec. 2.1 and 3 with frequency ω, coupled by exact minibatch OT on the harmonic action cost c_ω (Example E.1).
- **OT-Aniso** (CIFAR-10 only): the anisotropic harmonic flow of Appendix F, coupled by exact minibatch OT on the anisotropic action cost c_A.

The $\omega = 10^{-3}$ setting serves as a numerical proxy for the $\omega \to 0$ limit of the harmonic family, which recovers the affine interpolant and hence OT-CFM up to seed variation; we report both for completeness.

Training and seeding. Optimal-transport couplings are computed per-minibatch by the Hungarian algorithm [16]. For OT-CFM and OT-SI the cost is the squared-Euclidean distance. For isotropic OT-Harmonic the cost is the harmonic action c_ω, which differs from squared-Euclidean by a π-independent constant under fixed empirical marginals and so shares the same minimizer (Example E.1). For OT-Aniso the cost is the anisotropic action c_A, which does not reduce to squared-Euclidean and yields a genuinely different coupling. We use independent random seeds for data sampling, parameter initialization, and OT-coupling batches; results are reported as mean $\pm$ standard deviation across five independent retrains.

Evaluation metrics. The 2-Wasserstein distance W_2 between generated and target samples is estimated by linear assignment on a held-out batch of $N_{\text{eval}} = 2{,}048$ samples. The normalized path energy NPE_ω is computed as in Appendix G on a sub-batch of $n = 512$ source samples drawn from the N_{eval} batch and $n = 512$ fresh target samples: model trajectories are integrated from the source sub-batch by fixed-step RK4 at 200 steps, the kinetic-energy integrand $\frac{1}{2}\|v_\theta(\phi_t, t)\|^2$ is sampled on the same grid and integrated by composite Simpson's rule, and the harmonic-OT cost C_ω between

source and fresh target sub-batches is estimated by Hungarian assignment on the 512×512 pairwise harmonic-action matrix.

H.2 Two-dimensional synthetic data

Distributions. We use four sources and three targets, all in $\mathbb{R}^2$: $\mathcal{N}(0, I)$ as the canonical noise source; the *8gaussians* mixture (eight isotropic Gaussians equally spaced on a radius-5 circle, $\sigma = 0.5$); the *moons* dataset (two interlocking half-circles, additive noise 0.1); and a 2D projection of the standard *S-curve* manifold (using coordinates 0 and 2, standardized then rescaled by 7 for visual contrast). The four pairs reported in Table 1 are $\mathcal{N} \to$ 8gaussians, moons $\leftrightarrow$ 8gaussians, $\mathcal{N} \to$ moons, and $\mathcal{N} \to$ S-curve. Each configuration is trained for $20{,}000$ Adam steps at minibatch size 256.

Inference and the NFE plots. Table 1 in the main text reports distribution-fit and path-energy diagnostics at a fixed inference budget. To probe the inference-cost / sample-quality trade-off directly, we additionally sweep $\mathrm{NFE} \in \{4, 8, 16, 32, 64, 128\}$ across three fixed-step ODE integrators of increasing order (EULER, MIDPOINT, RK4). Since the higher-order solvers require more vector-field evaluations per integration step, the actual number of integration steps for a given NFE budget is NFE/k with $k = 1, 2, 4$ for Euler, midpoint, and RK4 respectively; this normalization plots all three curves on a common compute budget rather than a common step count. Fig. 8 shows the resulting curves for all five source→target settings. Each panel reports mean W_2 across five training seeds at $N_{\mathrm{eval}} = 2{,}048$, with shaded ± 1 standard-deviation bands. The figure makes two observations from Table 1 visible across the full inference budget: (i) OT-CFM and OT-Harmonic at $\omega = 10^{-3}$ overlap in every panel, confirming the affine-limit prediction numerically and serving as a sanity check on the implementation; and (ii) OT-Harmonic at $\omega = 1$ matches or improves on this shared baseline at every NFE, with the largest gap in the low-NFE regime where step error from non-straight trajectories dominates.

H.3 Single-cell trajectory interpolation

We follow the leave-one-out protocol of [29] on three datasets: *Embryoid body (EB)* [21], *CITE-seq*, and *Multiome* [8]. For each dataset we hold out one intermediate timepoint t at a time, train on the union of remaining timepoints, and evaluate the 1-Wasserstein distance between the model-interpolated distribution at t and the held-out empirical distribution. Table 4 reports the per-seed scalar (mean W_1 across held-out timepoints), averaged as mean $\pm$ standard deviation across five training seeds $\{42, 43, 44, 45, 46\}$.

Methods. We compare the five non-anisotropic methods of Appendix H.1: OT-CFM, OT-SI, and OT-Harmonic at $\omega \in \{10^{-3}, 1, \pi/2\}$. We do not run OT-Aniso on this track: PCA-based anisotropy fitting requires a sufficiently rich fit batch, which the per-timepoint sample sizes do not provide.

Architecture and training. The velocity field is parameterized by an MLP with width 64 and time-conditioned input. We train with AdamW at learning rate 10^{-3}, weight decay 10^{-5}, batch size 128, and noise scale $\sigma = 0.1$. The training budget is matched to the runner protocol of [29]: each epoch yields one minibatch per trajectory segment over the 80% training split, giving $\mathrm{steps/epoch} = \min_t \lfloor 0.8\,|X_t|/\mathrm{batch} \rfloor$ optimizer steps; we run for $1{,}000$ epochs.

Evaluation. At test time we integrate from the earliest surviving timepoint to the renumbered position of the held-out one with fixed-step Euler at 100 steps, then compute W_1 via exact linear assignment between $1{,}000$ generated samples and $1{,}000$ samples from the held-out distribution.

Preprocessing. All three datasets are obtained from the public Mendeley release of [21]. We use the top-5 principal components of each dataset, standardized to zero mean and unit standard deviation per coordinate. The CITE-seq and Multiome data were originally released as part of the Open Problems Multimodal Single-Cell Integration competition. For EB, we use the same PC dimensionality ($r = 5$) as [29], but the resulting embedding is not identical to theirs due to a preprocessing difference; absolute W_1 values on EB are therefore not directly comparable to numbers reported in that work, while the relative ordering between methods remains a meaningful comparison since all five methods are evaluated on the same embedding.

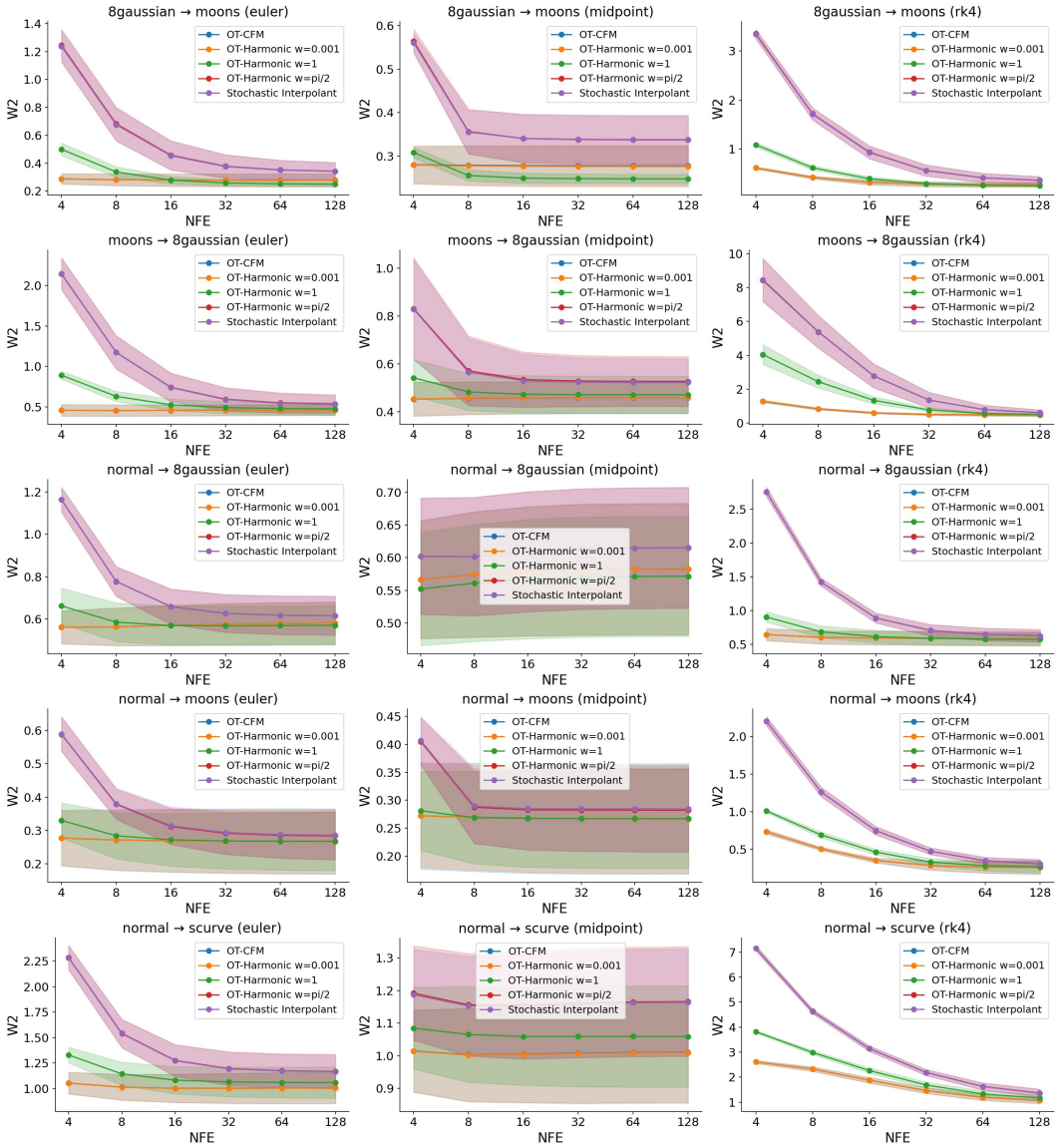


Figure 8: Sample quality vs. inference budget. We plot the 2-Wasserstein distance W_2 between generated and target samples against the number of function evaluations (NFE) at inference, across five source→target settings (rows) and three ODE integrators (columns). OT-CFM and OT-Harmonic at $\omega = 10^{-3}$ are visually indistinguishable, consistent with the harmonic interpolant degenerating to the affine one as $\omega \to 0$. OT-Harmonic at $\omega = 1$ matches or improves on this shared baseline at every NFE budget, with the largest gains in the low-NFE regime where step error dominates. Shaded bands show ± 1 standard deviation across five training seeds.

Time-Stratified 1-Wasserstein Loss Across Held-Out Timepoints. Table 4 reports leave-one-out 1-Wasserstein distances at each held-out timepoint across the three single-cell datasets. The harmonic interpolant in the small-ω regime ($\omega = 0.001$) is competitive with OT-CFM throughout: the two methods are within 3×10^{-3} of each other on every CITE and EB timepoint, and OT-Harmonic ($\omega = 0.001$) attains the lowest error on both CITE timepoints, on EB at $t = 2$, and on both Multiome timepoints. This is consistent with the theoretical limit in which the harmonic flow recovers the OT-CFM trajectory as $\omega \to 0$. Increasing ω degrades performance monotonically: $\omega = 1$ remains within roughly 5–15% of the leaders but already loses on most timepoints, and $\omega = \pi/2$ collapses to the OT-SI regime, with the two methods producing essentially indistinguishable errors across all seven columns (e.g., 1.1748 vs. 1.1852 on EB-$t = 1$; 1.8278 vs. 1.8139 on Multiome-$t = 2$). The

Table 4: Single-cell trajectory interpolation across three datasets, evaluated at each leave-one-out held-out timepoint. Entries are 1-Wasserstein distance to the held-out distribution; bold indicates the lowest mean per column.

	CITE		EB			Multiome	
Algorithm	$t = 1$	$t = 2$	$t = 1$	$t = 2$	$t = 3$	$t = 1$	$t = 2$
OT-CFM	0.9262 ± 0.035	0.8720 ± 0.035	$\mathbf{0.9504 \pm 0.045}$	0.8865 ± 0.045	$\mathbf{1.0187 \pm 0.140}$	0.8574 ± 0.060	1.2868 ± 0.080
OT-SI	1.1196 ± 0.035	1.3651 ± 0.035	1.1852 ± 0.045	1.1934 ± 0.045	1.6348 ± 0.140	1.1573 ± 0.060	1.8139 ± 0.080
OT-Harmonic, $\omega = 0.001$	$\mathbf{0.9261 \pm 0.038}$	$\mathbf{0.8709 \pm 0.038}$	0.9510 ± 0.045	$\mathbf{0.8846 \pm 0.045}$	1.0213 ± 0.140	$\mathbf{0.8562 \pm 0.060}$	$\mathbf{1.2855 \pm 0.080}$
OT-Harmonic, $\omega = 1$	0.9338 ± 0.030	0.9969 ± 0.030	0.9829 ± 0.060	0.8951 ± 0.060	1.0946 ± 0.140	0.9017 ± 0.080	1.4106 ± 0.080
OT-Harmonic, $\omega = \pi/2$	1.1232 ± 0.025	1.3514 ± 0.025	1.1748 ± 0.045	1.1524 ± 0.045	1.6785 ± 0.140	1.1599 ± 0.080	1.8278 ± 0.080

gap between the harmonic and OT-CFM/OT-SI extremes widens with the prediction horizon: on EB, all methods are tightly clustered at $t = 2$, but by $t = 3$ the high-ω variants incur roughly 60% more error than the low-ω ones, suggesting that the oscillatory component of the harmonic flow primarily hurts long-range extrapolation rather than near-boundary interpolation.

H.4 CIFAR-10 image generation

All CIFAR-10 runs share a single training pipeline that selects the appropriate flow matcher via a model flag. Architecture, optimizer, and schedule follow [29] unchanged: a U-Net with 128 base channels, channel multipliers $[1, 2, 2, 2]$, two residual blocks per resolution, four attention heads with 64 channels each, attention at the 16×16 resolution, dropout 0.1; trained for $400{,}000$ steps with Adam at learning rate 2×10^{-4}, $5{,}000$-step linear warmup, gradient-norm clipping at 1.0, batch size 128, EMA on model weights with decay 0.9999, and image normalization to $[-1, 1]$. For Fig. 9 we sample trajectories from a fixed noise seed $x_0 \sim \mathcal{N}(0, I)$, integrated forward to $t = 1$ with DOPRI5; intermediate columns are uniformly sampled from this trajectory and rescaled from $[-1, 1]$ to $[0, 1]$ for display. The six panels of Fig. 9 show samples from each variant at the end of training. OT-CFM (Fig. 9a) and OT-Harmonic at $\omega = 0.001$ (Fig. 9b) produce visually indistinguishable samples, consistent with the $\omega \to 0$ limit in which the harmonic flow recovers OT-CFM. As ω increases, samples grow progressively noisier: $\omega = 1$ (Fig. 9c) shows mild loss of detail, and $\omega = \pi/2$ (Fig. 9d) is qualitatively similar to OT-SI (Fig. 9e), again matching the theoretical correspondence at the upper limit. OT-Aniso (Fig. 9f) yields samples comparable to OT-SI rather than to OT-CFM, suggesting that the data-fit anisotropic potential does not exploit useful low-dimensional structure on this benchmark.

Anisotropic variant. The OT-Aniso row fits an anisotropic harmonic potential $A = R^\top \mathrm{diag}(\omega_1^2, \ldots, \omega_d^2)\, R$ on the flat image dimension $d = 3 \times 32 \times 32 = 3{,}072$ from a single fit batch of 10 training batches ($1{,}280$ images at batch size 128). Since the fit batch ($N = 1{,}280$) is smaller than d, R is computed by thin SVD on the centered fit batch: the $1{,}280$ data-subspace eigenvectors are taken from the right-singular vectors, and the remaining $d - N = 1{,}792$ null-space directions are completed by orthogonal random vectors projected against the data subspace. Per-direction frequencies are assigned linearly: data-subspace directions receive ω_k uniformly spaced in $[\omega_{\min}, \omega_{\max}]$ with the highest-variance PC at $\omega_{\min} = 0.8$ and the lowest-variance data PC at $\omega_{\max} = 1.6$, while the $1{,}792$ null-space directions all receive $\omega_{\max}$ since no variance evidence exists for them. The eigenvalue constraint $\omega_k \in (0, \pi)$ holds automatically since $\omega_{\max} = 1.6 < \pi$. The OT coupling is solved on the anisotropic action cost c_A rather than squared-Euclidean. A is precomputed once at the start of training and frozen thereafter.

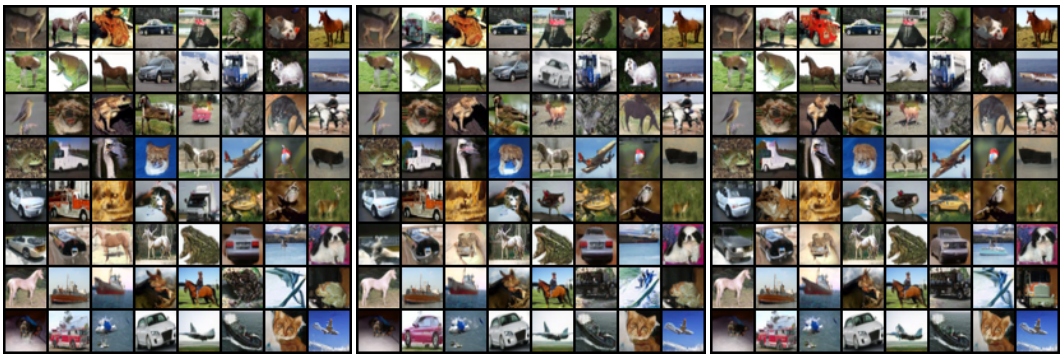
H.5 Compute and reproducibility

All experiments were run on NVIDIA H100 GPUs (80 GB). Code, configurations, and pretrained checkpoints will be released at the project URL upon publication.

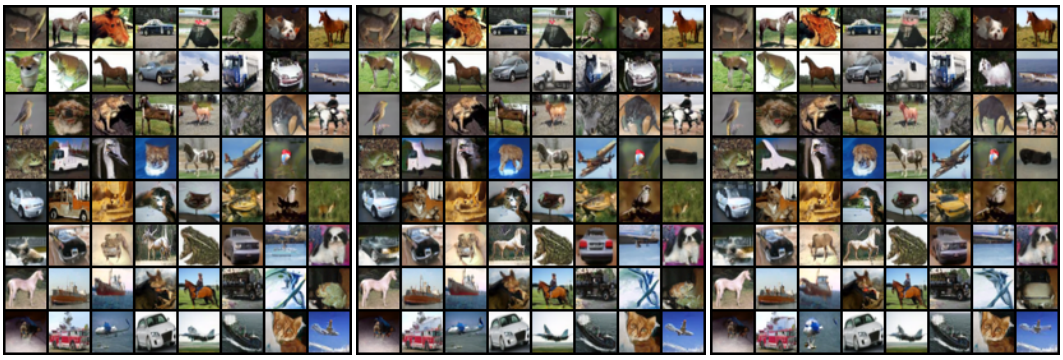
Limitations

We discuss several limitations of the present work.

Restricted Lagrangian class. Our results cover Lagrangians of the canonical form $\mathcal{L}(x, v, t) = K(v) - V(x)$, where K is a positive-definite quadratic form and V is sufficiently regular and



(a) OT-CFM (b) OT-Harmonic, $\omega = 0.001$ (c) OT-Harmonic, $\omega = 1$



(d) OT-Harmonic, $\omega = \pi/2$ (e) OT-SI (f) OT-Aniso

Figure 9: Uncurated CIFAR-10 samples from each flow-matching variant after 400,000 training steps, generated with EMA weights and the DOPRI5 adaptive solver from a fixed noise seed $x_0 \sim \mathcal{N}(0, I)$. The harmonic family interpolates between OT-CFM ($\omega \to 0$) and OT-SI ($\omega = \pi/2$) as ω increases. OT-Aniso uses the data-fit anisotropic harmonic potential described in Appendix H.4.

controlled by the kinetic term. This setting is sufficient to derive the static–dynamic equivalence. However, a simulation-free method requires the least-action trajectories to admit closed-form expressions. Thus, the approach does not extend trivially to general Lagrangians; in that case, one would typically need to solve the Euler–Lagrange boundary-value problem numerically, thereby sacrificing the simulation-free benefits of the current framework. Developing efficient batched solvers for computing these least-action trajectories is therefore a natural direction for future work.

The potential is prescribed, not learned. Throughout the paper we treat the potential V (or the matrix A in the anisotropic case) as a hyperparameter rather than a learnable component. The harmonic frequency ω is set by hand, and the anisotropic potential A is fit once from a small batch of empirical variances and frozen for the remainder of training. Our CIFAR-10 results (Table 3) suggest that this fit-once-and-freeze strategy is the dominant limitation of the anisotropic variant on high-dimensional natural images: when the data covariance carries little geometric signal beyond a near-isotropic spectrum, the prescribed A provides no meaningful inductive bias. Jointly learning V with the velocity field, in the spirit of energy-based models (Appendix B), is a natural extension that we leave to future work.

No single Lagrangian dominates. The harmonic family is not uniformly dominant over its endpoints: $\omega \to 0$ recovers OT-CFM and is the strongest setting on CIFAR-10, while $\omega = 1$ improves on both baselines on 2D synthetic targets and Multiome single-cell interpolation. This is a structural feature of the framework rather than a defect — a Lagrangian encodes an inductive bias, and no fixed bias is optimal across all data geometries — but it does mean that the choice of Lagrangian is itself a modeling decision, analogous to the choice of kernel in kernel methods or architecture in deep learning. A principled procedure for selecting ω (or, more generally, the potential V) from data,

perhaps by joint learning as discussed in Appendix B, would strengthen the practical applicability of the framework.

Empirical scope. Our experiments span 2D synthetic targets, low-dimensional single-cell embeddings, and CIFAR-10 at resolution 32×32. We have not evaluated on higher-resolution natural images, text-to-image generation, video, or molecular tasks, and the comparative behavior of different Lagrangians at larger scales remains open. The CIFAR-10 training budget of $400{,}000$ steps is shorter than the run reported in [29] ($500{,}000$ steps); we use this budget uniformly across methods for a controlled comparison within our compute budget, but absolute FID values are correspondingly higher than the longer-run baseline.

Deterministic dynamics only. The framework as developed selects probability paths through the deterministic continuity equation, which makes it a flow-matching construction rather than a diffusion or Schrödinger-bridge construction. The trigonometric VP path of Example 2.5 matches the marginal of the corresponding diffusion model but not the full stochastic forward process, and stochastic Lagrangian extensions — replacing the continuity equation with a Fokker–Planck equation, in the spirit of Appendix C — are left to future work.