Path-independent Flow Matching for Multi-parameter Generative Dynamics

Francisco Téllez[*] Université de Montréal, Mila
AmirHossein Zamani[*] Concordia University, Mila
Philippe Martin[*] Université de Montréal, Mila
Shuang Ni Université de Montréal, Mila
Guy Wolf Université de Montréal, Mila
Eugene Belilovsky Concordia University, Mila
Sina Sanjari[†] Royal Military College of Canada
Yanlei Zhang[†] Queen's University, Mila
Correspondance: zhang.yanlei@queensu.ca

Abstract

Flow Matching is a powerful framework for learning transport maps between probability distributions. Yet its standard single-parameter formulation is not designed to capture multi-parameter variations where the resulting transport should be path-independent. Path independence is crucial because it ensures that transformations depend only on the initial and target distributions, not on the specific path. In this work, we introduce *Path-independent Flow Matching (PiFM)*, a method for learning vector fields whose induced flows yield path-independent transport between distributions. We show that PiFM generalizes Flow Matching to higher-dimensional parameter domains while enforcing structural conditions that ensure consistency of composed transformations. In addition, we show that, under suitable assumptions, PiFM approximates the Wasserstein barycenter, linking the framework to a notion of distributional interpolation. To enable practical training, we propose a tractable, simulation-free objective that regresses onto multi-parameter conditional probability paths. We showcase empirically that PiFM outperforms other approaches on both synthetic and real world data in interpolating path-independent trajectories and generating desired out of distribution samples.

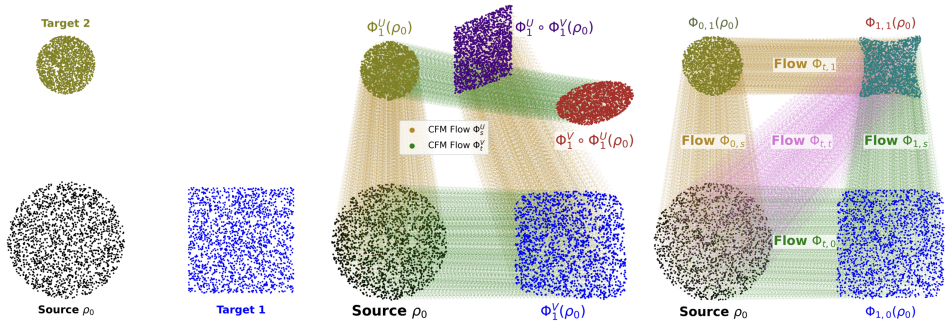
1 Introduction

Learning transformations between probability distributions is a central problem in modern machine learning, with applications spanning generative modeling, domain adaptation, trajectory inference, and scientific discovery. In these settings, one seeks not only to match target distributions, but also to understand how distributions relate through structured, continuous transformations such as the flows of vector fields. Recent neural approaches achieve this by reducing population-level objectives to sample-wise computations, enabling scalable training via stochastic optimization. This paradigm underlies VAEs (Kingma & Welling, 2022; Diederik & Max, 2019), GANs (Goodfellow et al., 2014; S & Durgadevi, 2021), diffusion models (Ho et al., 2020; Song et al., 2021), and continuous-time transport methods (Chen et al., 2019; Grathwohl et al., 2018; Lipman et al., 2023; Tong et al., 2023, 2024), and has led to growing interest in modeling distributions through dynamical systems.

A broad class of methods, including continuous normalizing flows (Chen et al., 2019; Grathwohl et al., 2018), diffusion and score-based models (Ho et al., 2020; Song et al., 2021; Koehler et al., 2022; Berner et al., 2022) , and Schrödinger bridges (Léonard, 2013; Bortoli et al., 2023; Liu et al., 2023),

[*]Equal contribution
[†]Co-senior authors

40th Conference on Neural Information Processing Systems (NeurIPS 2026).



(a) Data (b) CFM (c) PiFM (ours)

Figure 1: Learned trajectories for multi-marginal transport between a source (unit disc) and two targets (square, small disc). Traditional CFMs do not commute by composing flows of two separate processes, whereas PiFM (ours) is able to model commutative multi-process trajectories.

can be interpreted as learning time-varying transport dynamics that push a source distribution toward a target. Among these, Flow Matching (Lipman et al., 2023; Liu, 2022) provides a particularly simple and simulation-free training principle by directly regressing neural vector fields toward reference conditional velocities induced by a prescribed probability path. Conditional Flow Matching (Tong et al., 2023, 2024) and related extensions further generalize this idea, unifying several continuous-time generative frameworks by allowing transport between a broader family of source and target distributions.

Despite their success, existing flow matching methods are inherently limited to a single time parameter, implicitly assuming that all variations can be ordered along a single temporal axis. In many real-world settings, however, data are influenced by multiple independent processes that evolve concurrently and should compose in a path-independent manner. This arises naturally in settings such as biological perturbations acting on time-evolving cellular populations in single-cell genomics (Lotfollahi et al., 2023; Jiang et al., 2025), compositional attributes like shape and color in images (Wang et al., 2024), and modular transformations in physical systems (Neary & Topcu, 2023). Naively composing independently trained flows generally leads to non-commutative dynamics, where the final distribution depends on the order of application, limiting interpretability and downstream geometric reasoning as tasks such as analogy completion or attribute extrapolation become ill-defined when transformations fail to commute.

To address this limitation, we introduce *Path-independent Flow Matching (PiFM)*, a framework for learning multi-parameter generative dynamics driven by vector fields with a path-independent transport of distributions property. PiFM models probability paths over multiple temporal domains and enforces path independence through structural conditions on the vector fields, echoing classical commutativity and integrability criteria for flows (Lee, 2003; Warner, 1983; Abraham et al., 1988). This yields a well-defined notion of composing transformations, enabling consistent interpolation, extrapolation, and analogy-like operations in distribution space. We further establish an empirical connection between PiFM and Wasserstein barycenters (Agueh & Carlier, 2011; Bigot & Klein, 2017), with additional conditions for exact matching. Beyond the associated Wasserstein simplex parameter space, we show that our model offers extrapolation capabilities, extending this framework as a natural generalization of multi-marginal generative modeling (Cao et al., 2019; Albergo & et al., 2023; Beier et al., 2022; Pass, 2015). As a result, PiFM is shown to support OOD inference while avoiding the need for full manipulation of densities in distribution space as in methods such as Atanackovic et al. (2025); Haviv et al. (2025). Lastly, we evaluate PiFM through several computational experiments ranging from image translation and adaptive target generation to synthetic datasets for demonstration purposes.

Our contributions are summarized as follows:

1. **Path-independent Flow Matching (PiFM):** We introduce PiFM, a multi-parameter Flow Matching framework driven by multiple vector fields. We establish the existence of path-independent flows and derive the sufficient integrability conditions ensuring consistency.

2. **Conditional-to-Marginal Flow Consistency:** We show how conditional probability paths induce coherent marginal dynamics in the multi-parameter setting. This leads to a training objective that

combines vector field regression with a path-independence regularizer, which we further show becomes unnecessary in the affine case.
3. **Connection to Wasserstein Barycenters:** We establish a connection between PiFM and Wasserstein barycenters, enabling both interpolation and extrapolation of distributions.
4. **Empirical Validation:** We demonstrate the effectiveness of PiFM on synthetic and image translation tasks. PiFM outperforms prior methods such as Meta Flow Matching (Atanackovic et al., 2025) and Curly Flow Matching (Petrović et al., 2025) in low-dimensional domain shift settings, and achieves improved compositionality and consistency in high-dimensional multi-attribute image translation. We further validate PiFM on a dense single-cell reprogramming dataset (Schiebinger et al., 2019), treating experimental day and pluripotency score as two interacting flows to test path-independent generation in a high-dimensional, multi-stage biological process.

The paper is organized as follows. Section 2 reviews the necessary background on flow matching. Section 3 introduces the PiFM framework, its theoretical properties, and its connection to Wasserstein barycenters. Section 4 presents experimental results in both low-dimensional synthetic domain-shift settings and high-dimensional image translation tasks and single-cell application.

2 Background: Flow Matching and Minibatch Optimal Transport

The primary objective in distribution matching is to recover a transport map $\psi : \mathbb{R}^d \to \mathbb{R}^d$ that transports a source density $q(a)$ to a target density $q(b)$. This push-forward requirement, denoted $q(b) = [\psi]_{\#}(q(a))$, underpins both optimal transport and modern generative modeling. In the empirical setting, where we lack analytical access to densities and observe only finite datasets $\{x_0^i\}_{i=1}^N \sim q(a)$ and $\{x_1^j\}_{j=1}^M \sim q(b)$, the challenge lies in approximating ψ from discrete samples. A powerful approach to model ψ is through dynamical systems, specifically ordinary differential equations (ODEs), which motivate the Flow Matching approach.

Flow Matching (Liu, 2022; Lipman et al., 2023; Tong et al., 2023; Albergo & Vanden-Eijnden, 2023) provides a framework for learning a dynamical system by establishing a correspondence between a probability density path $p : [0, 1] \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ and a vector field $v : [0, 1] \times \mathbb{R}^d \to \mathbb{R}^d$. We say that p_t is generated by v_t if the pair (v, p) satisfies the following *continuity equation*

$$\frac{\partial p_t(x)}{\partial t} + \nabla \cdot (p_t(x)v_t(x)) = 0, \tag{1}$$

subject to the boundary condition $p_0 = q(a)$ and $p_1 = q(b)$.

While solving for the marginal vector field directly is intractable, *Conditional Flow Matching* (CFM) circumvents this by exploiting the linearity of (1) and considering the marginal densities $p_t(x)$ as a mixture of simpler conditional densities $p_t(x|z)$ conditioned on a latent variable z. In the CFM setting, the conditional densities $p_t(x|z)$ are induced by a conditional vector field $u_t(x|z)$. Consider a latent variable z with corresponding density q. Then, $p_t(x) = \int p_t(x|z)q(z)\,dz$, and by defining the unconditional vector field $u_t(x)$ by $u_t(x) := \mathbb{E}_{q(z)}\left[\frac{u_t(x|z)p_t(x|z)}{p_t(x)}\right]$, (Lipman et al., 2023) showed that when the conditional vector field $u_t(x|z)$ generates the conditional probability path $p_t(x|z)$, the unconditional vector field $u_t(x)$ generates the marginal probability path $p_t(x)$.

For CFM, a vector field $v^\theta(t, x)$ is trained to approximate the intractable unconditional vector field by regressing onto the conditional fields. The CFM learning objective is to find θ that minimizes the loss function $\mathcal{L}_{\mathrm{CFM}}(\theta)$ given by

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{t\sim U(0,1), z\sim q,\, x\sim p_t(\cdot|z)}\left[\|v_\theta(t, x) - u_t(x|z)\|^2\right].$$

This formulation enables simulation-free training of continuous normalizing flows by sampling $t \sim \mathcal{U}(0, 1)$, $z \sim q(z)$, and $x \sim p_t(\cdot|z)$, while relying on efficient sampling from $q(z)$.

The density $q(z)$ can be specified in different ways. In independent CFM (I-CFM), $q(z) = q(a)q(b)$, while in optimal transport CFM (OT-CFM), $q(z) = \pi(a, b)$, where π is an optimal transport coupling between $q(a)$ and $q(b)$. Training under I-CFM uses independently sampled pairs (a, b), whereas OT-CFM requires sampling from π, typically approximated via mini-batch optimal transport due to the cubic time and quadratic memory cost of full OT (Cuturi, 2013). Despite being approximate, mini-batch OT preserves the transport geometry in expectation and performs well empirically (Fatras et al., 2021); the exact OT-CFM objective is recovered when the batch size equals the full dataset.

3 Multi-parameter Flow Matching and Path Independence

We begin by introducing a multi-parameter flow matching formulation. To simplify our notations, we restrict our exposition to the case of two independent parameters, denoted by t and s (see Section A.1 for details on the more general case). Then, we introduce the PiFM framework, which extends CFM to a multi-parameter setting. We present the modeling choices and assumptions that form the foundation of PiFM, and show how these elements collectively ensure coherent marginal generation and path-independent transport of distributions. Finally, we describe how to train PiFM using a deep learning approach and establish a connection to the Wasserstein Barycenter. Proofs are included in the Appendix A.2 and Appendix A.3.

3.1 Multi-parameter Flows and Path Independence

Suppose we have two continuously differentiable vector fields $u, v : [0,1]^2 \times \mathbb{R}^d \to \mathbb{R}^d$ depending on parameters t and s with associated ODEs:

$$\frac{\partial \phi}{\partial t}(t,s) = u_{t,s}(\phi(t,s)), \;\; \forall s \in [0,1] \;\; ; \;\; \frac{\partial \psi}{\partial s}(t,s) = v_{t,s}(\psi(t,s)), \;\; \forall t \in [0,1]. \tag{2}$$

With initial conditions $\phi_{0,0}(x) = x$ and $\psi_{0,0}(x) = x$, both equations are well-defined at $(t,s) = (0,0)$, yielding the functions $\phi_{t,0}$ and $\psi_{0,s}$. These solutions can then be used to define consistent initial conditions along the axes, namely $\phi_{0,s}(x) = \psi_{0,s}(x)$ and $\psi_{t,0}(x) = \phi_{t,0}(x)$, which in turn determine the full solutions $\phi_{t,s}$ and $\psi_{t,s}$. Note that this construction proceeds by solving ODEs sequentially, existence and uniqueness at each stage are guaranteed by standard ODE theory (see, e.g., (Wiggins, 2003)). Associated with these solutions, we define the flow maps $\Phi, \Psi : [0,1]^2 \times \mathbb{R}^d \to \mathbb{R}^d$ by $\Phi_{t,s}(x) = \phi_{t,s}(x)$ and $\Psi_{t,s}(x) = \psi_{t,s}(x)$.

Analogous to the one-parameter setting, the vector fields $u_{t,s}$ and $v_{t,s}$ generate densities $p_{t,s}$ and $q_{t,s}$, respectively, provided that the following continuity equations hold for all $t, s \in [0,1]$:

$$\frac{\partial p_{t,s}(x)}{\partial t} + \nabla \cdot \big(p_{t,s}(x)\, u_{t,s}(x)\big) = 0 \qquad \frac{\partial q_{t,s}(x)}{\partial s} + \nabla \cdot \big(q_{t,s}(x)\, v_{t,s}(x)\big) = 0. \tag{3}$$

Similarly to the definition of initial conditions for ODEs, we prescribe boundary conditions along the coordinate axes. Given an initial density ρ_1 and target densities ρ_2 and ρ_3, we define the boundary conditions by setting $p_{0,s} = (\Psi_{0,s})_{\#}\rho_1$, $p_{1,s} = (\Psi_{1,s})_{\#}\rho_2$, $q_{t,0} = (\Phi_{t,0})_{\#}\rho_1$ and $q_{t,1} = (\Phi_{t,1})_{\#}\rho_3$. With these constructions, the continuity equations yield consistent transport of $p_{t,s}$ along the flow $\Phi_{t,s}$ and of $q_{t,s}$ along the flow $\Psi_{t,s}$.

We say the transport flow is distributionally path-independent (or commutative) if

$$(\Phi_{t,s})_{\#} \circ (\Psi_{0,s})_{\#} p_{0,0} = (\Psi_{t,s})_{\#} \circ (\Phi_{t,0})_{\#} p_{0,0}, \quad \forall t, s \in [0,1]. \tag{4}$$

Intuitively, (4) states that moving first along the t-direction while fixing $s = 0$ using the flow $\Phi_{t,s}$, and then along the s-direction using $\Psi_{t,s}$, yields the same result as performing the operations in the reverse order. The implication this has for path independence is that if we consider a parametrized path $\gamma(l) : [0,1] \to [0,1]^2$, with $\gamma(l) = (\gamma^{(1)}(l), \gamma^{(2)}(l)) = (t(l), s(l))$, then (4) implies that the order in which we apply the transport maps $\Phi_{\#}$ and $\Psi_{\#}$ does not matter: we always arrive at the same distribution at the point (t,s) along the trajectory prescribed by γ. This is formalized via

$$(\Phi_{\gamma(l)})_{\#} \circ (\Psi_{0,\gamma^{(2)}(l)})_{\#} p_{0,0} = (\Psi_{\gamma(l)})_{\#} \circ (\Phi_{\gamma^{(1)}(l),0})_{\#} p_{0,0}.$$

To preserve distributional path-independence, it is sufficient if there exists a single probability density path that satisfies continuity equations (3), that is, when $p_{t,s}$ and $q_{t,s}$ coincide for all t and s in $[0,1]$. We formally state this as a lemma.

Lemma 3.1 (Path-independent transport of distributions)**.** *Consider $u_{t,s}$ and $v_{t,s}$ vector fields. If there exists a unique generated probability density path $p_{t,s}$ satisfying* (3)*, then the associated flows $\Phi_{t,s}$ and $\Psi_{t,s}$ satisfy* (4)*.*

Lemma 3.1 provides a sufficient condition for path-independent transport of distributions. However, other sufficient conditions for this property exist. One way this can occur is when the system defined by (2) admits a unique joint trajectory $\phi_{t,s}$. This implies pointwise path independence, while for our formulation we only need distributional path independence, which is a weaker property. A necessary and sufficient condition for a unique solution $\phi_{t,s}$ to exist is provided in Proposition 19.29 of (Lee, 2003) which uses the partial differential equation described in (5).

Proposition 3.2 ((Pointwise) path-independence equation)**.** *Suppose that the vector fields $u_{t,s}$ and $v_{t,s}$ are satisfy*

$$\partial_s u_{t,s} - \partial_t v_{t,s} = [u_{t,s}, v_{t,s}], \quad \forall x \in \mathbb{R}^d \tag{5}$$

where $\partial_s u_{t,s} := \partial u_{t,s}(x)/\partial s$, $\partial_t v_{t,s} := \partial v_{t,s}(x)/\partial t$, and $[\cdot,\cdot]$ is the Lie Bracket, $[u_{t,s}, v_{t,s}] := (\nabla_x v_{t,s}(x))\, u_{t,s}(x) - (\nabla_x u_{t,s}(x))\, v_{t,s}(x)$. Then there exists a unique flow $\Phi_{t,s}$ which satisfies (4)*.*

The preceding proposition can be read as follows, pointwise path-independence can be characterized by the integrability equation (5). When this equation is satisfied, existence of a joint flow $\Phi_{t,s}$ of $u_{t,s}$ and $v_{t,s}$ is guaranteed. This function satisfies (4), so in this case distributional path-independence is also guaranteed.

3.2 Consistent and Path-independent Transport of Distributions in PiFM

We consider data distributions $q(a)$, $q(b)$ and $q(c)$, and the joint density $q(z)$ with $z = (a, b, c)$. In contrast to the traditional flow matching framework, we consider the probability density path $p : [0,1]^2 \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$ of two-parameter with boundry conditions $p_{0,0}(x) = q(a)$, $p_{1,0}(x) = q(b)$ and $p_{0,1}(x) = q(c)$. Analogous to CFM, we condition on z yielding,

$$p_{t,s}(x) = \int p_{t,s}(x|z) q(z) dz, \tag{6}$$

where the conditional density path $p_{t,s}(x|z)$ is induced by some conditional vector fields $u_{t,s}(x|z)$ and $v_{t,s}(x|z)$. We seek two vector fields $u_{t,s}, v_{t,s} : [0,1]^2 \times \mathbb{R}^d \to \mathbb{R}^d$ that generate $p_{t,s}$ by satisfying the continuity equations for all $t, s \in [0,1]$

$$\frac{\partial p_{t,s}(x)}{\partial t} + \nabla \cdot \big(p_{t,s}(x)\, u_{t,s}(x)\big) = 0, \qquad \frac{\partial p_{t,s}(x)}{\partial s} + \nabla \cdot \big(p_{t,s}(x)\, v_{t,s}(x)\big) = 0. \tag{7}$$

If the two equations above hold, then the corresponding flows satisfy the distributional path-independence property. Given conditional vector fields $u_{t,s}(x|z)$ and $v_{t,s}(x|z)$, we model the unconditional vector fields by

$$u_{t,s}(x) = \mathbb{E}_{q(z)}\left[\frac{u_{t,s}(x \mid z) p_{t,s}(x \mid z)}{p_{t,s}(x)}\right], \quad v_{t,s}(x) = \mathbb{E}_{q(z)}\left[\frac{v_{t,s}(x \mid z) p_{t,s}(x \mid z)}{p_{t,s}(x)}\right]. \tag{8}$$

We note that the precise forms of the conditional vector fields $u_{t,s}(x|z)$ and $v_{t,s}(x|z)$ are not prescribed a priori and left as a user-defined choice.

Theorem 3.3. *Let the conditional vector fields $u_{t,s}(x|z)$ and $v_{t,s}(x|z)$ generate the unique conditional probability density path $p_{t,s}(x|z)$ satisfying the analogous equations* (7)*. Then, the vector fields $u_{t,s}(x)$ and $v_{t,s}(x)$ in* (8) *generate the unique probability density path $p_{t,s}(x)$ satisfying* (7)*.*

Note that Theorem 3.3 is a statement about how $p_{t,s}$ is consistently transported using vector fields $u_{t,s}$ and $v_{t,s}$, it is not a statement about pointwise or distributional-path-independence. Additionally, Theorem 3.3 can be viewed as a generalization of the conditional-to-marginal generation result in CFM, extending the framework to accommodate multiple parameter settings and additional target distributions. As a corollary to Theorem 3.3 and Lemma 3.1, we have the following result: the flows of the unconditional vector fields have the distributional path-independence property.

Corollary 3.4 (Path-independent transport of distributions in PiFM)**.** *Suppose that the vector fields $u_{t,s}(x|z)$ and $v_{t,s}(x|z)$ generate the unique conditional probability density path $p_{t,s}(x|z)$ satisfying the analogous equations* (7)*. Then, the associated flows $\Phi_{t,s}$ and $\Psi_{t,s}$ of the vector fields $u_{t,s}(x)$ and $v_{t,s}(x)$ given in* (8) *satisfy* (4)*.*

3.3 Training via Path-independent Flow Matching

Now we discuss a neural approach for training a PiFM model. Given two parameterized vector fields $u^\theta_{t,s}$ and $v^\theta_{t,s}$, PiFM learns the dynamics by regressing these fields onto the corresponding conditional vector fields. We define the flow-matching regression loss as

$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{(t,s)\sim U[0,1]^2, z\sim q,\, x\sim p_{t,s}(\cdot|z)}\Big[\|u^\theta_{t,s}(x) - u_{t,s}(x \mid z)\|^2 + \|v^\theta_{t,s}(x) - v_{t,s}(x \mid z)\|^2\Big]$$

For conditional vector fields that do not automatically satisfy the pointwise path-independence condition in (5), we consider additionally penalizing violations of this condition, this is particularly

useful in settings such as integrating PiFM with Curly Flow Matching (Petrović et al., 2025); see Section 4.1.2. The path-independence regularizer is

$$\mathcal{L}_{\text{Pi}}(\theta) = \mathbb{E}\Big[\Big\|\partial_s u^\theta_{t,s}(x \mid z) - \partial_t v^\theta_{t,s}(x \mid z) - [u^\theta_{t,s}(x \mid z), v^\theta_{t,s}(x \mid z)]\Big\|^2\Big],$$

with the expectation computed over (t, s), z and x as in the Flow Matching regression loss. Given a regularization weight $\lambda \geq 0$, the PiFM objective is $\mathcal{L}_{\text{PiFM}}(\theta; \lambda) = \mathcal{L}_{\text{FM}}(\theta) + \lambda\mathcal{L}_{\text{Pi}}(\theta)$. Training proceeds analogously to CFM. We sample $(t, s) \sim \mathcal{U}[0, 1]^2$ and $z = (a, b, c) \sim q(z)$, where $q(z)$ may be specified using a variety of couplings such as the coupling that results from using OT couplings in (a, b) and (a, c), multimarginal OT coupling in (a, b, c) or independent coupling which arises from independently sampling in each of the three distributions separately. We then sample $x \sim p_{t,s}(\cdot \mid z)$ and update θ by minimizing $\mathcal{L}_{\text{PiFM}}(\theta; \lambda)$. The full training procedure is summarized in Algorithm 1.

One specific choice of conditional vector fields and conditional probability path is given by

$$u_{t,s}(x \mid z) = b - a, \quad v_{t,s}(x \mid z) = c - a, \quad p_{t,s}(x \mid z) = \mathcal{N}\big(x \mid \mu_{t,s}(z), \sigma^2\big), \tag{9}$$

for fixed $\sigma > 0$. This generalizes the affine interpolation commonly used in CFM to a multi-parameter setting by introducing a two-dimensional interpolation of the form $\mu_{t,s}(z) = a+(b-a)t+(c-a)s$. This choice of conditional vector fields and conditional probability density paths is guaranteed to induce unconditional vector fields that transport the marginal probability density path $p_{t,s}$ in a consistent and path-independent manner by satisfying both Theorem 3.3 and Corollary 3.4. We emphasize that $\mathcal{L}_{Pi}$ can promote both pointwise and distributional path-independence. It directly affects pointwise path-independence through the integrability condition (5), while distributional path-independence follows from the pointwise property. Therefore, the path-independence regularizer is unnecessary when the assumptions of Corollary 3.4 hold, but may be beneficial when they are violated, as demonstrated in Section 4.1.2.

Algorithm 1 Path-Independent Flow Matching (PiFM)

1: **Input:** Efficiently samplable $q(a, b, c)$, $p_{t,s}(x|(a, b, c))$, computable conditional vector fields $u_{t,s}(x|(a, b, c))$ and $v_{t,s}(x|(a, b, c))$, initial two-headed network (u^θ, v^θ) with a single backbone, learning rate $\eta > 0$ and regularization $\lambda > 0$.
2: **while** training **do**
3: Sample $(a, b, c) \sim q(a, b, c)$, $(t, s) \sim \mathcal{U}([0, 1]^2)$, $x \sim p_{t,s}(x|(a, b, c))$
4: Compute loss $\mathcal{L}_{\text{PiFM}}(\theta; \lambda)$
5: Update $\theta \leftarrow \theta - \eta\nabla_\theta\mathcal{L}_{\text{PiFM}}(\theta)$
6: **end while**
7: **return** u_θ, v_θ

3.4 Relation to Wasserstein Barycenter

In the previous subsections, we discussed PiFM, its properties, and its training. Now, we show that PiFM empirically approximates the Wasserstein Barycenter between three distributions. Additionally, under suitable conditions, we provide theoretical guarantees that PiFM matches the Wasserstein barycenter, demonstrating that PiFM recovers a geometrically meaningful notion that is globally consistent with interpolation between distributions.

Let $\lambda := (\lambda_1, \ldots, \lambda_K)$ with $\lambda_j \geq 0$ and $\sum_{j=1}^K \lambda_j = 1$ for some $K \in \mathbb{N}$. The Wasserstein barycenter ρ_λ of probability measures $\rho_1, \ldots, \rho_K \in \mathcal{P}_2(\mathbb{R}^d)$ with finite second moment is a distribution that minimizes the multi-marginal optimal transport cost $J(\gamma, \lambda)$ given below

$$J(\gamma, \lambda) = \int \sum_{j=1}^K \lambda_j \left\|T_\lambda(x_1, \ldots, x_K) - x_j\right\|^2 d\gamma,$$

over all couplings γ of $\rho_1, \ldots, \rho_K$ with $T_\lambda(x_1, \ldots, x_K) = \sum_{j=1}^K \lambda_j x_j$.

One popular algorithm for computing the Wasserstein barycenter is the free support computation algorithm (Cuturi & Doucet, 2014), available in the POT package (Flamary et al., 2021). Empirically, we observe in Figure 2a. that distributions $\rho^{\text{PiFM}}_{t,s}$ computed from our method coincide with the Wasserstein barycenter $WB_{1-t-s,t,s}(\rho_0, \rho_1, \rho_2)$ obtained with the free support algorithm. In addition, we empirically observe that our framework extrapolates beyond the simplex constraint on λ in Figure 2b. Specifically, PiFM and free-support generated distributions coincide in the extrapolation region.

To support our claim that PiFM can approximate the Wasserstein barycenter, we provide additional experimental results in Supplementary Figure 7 on more complex distributions. In addition, using the multi-marginal optimal transport formulation and the notion of admissible families of deformations from (Boissard et al., 2013), Theorem 3.5 gives sufficient conditions under which the PiFM-generated distribution $\hat{\rho}_{t,s}$ is the minimizer of the Wasserstein barycenter objective. From the discussion in section 3.2 note that $\hat{\rho}_{t,s} = (\Phi_{t,s})_{\#} \circ (\Psi_{0,s})_{\#}\rho_0 = (\Psi_{t,s})_{\#} \circ (\Phi_{t,0})_{\#}\rho_0 =: \Gamma_{t,s}(\rho_0)$.

Theorem 3.5. *Consider $\rho_0, \rho_1, \rho_2 \in \mathcal{P}(\mathbb{R}^d)$ with $\lambda = (1-t-s, t, s)$. Suppose that the optimal transport maps $T_1 : \rho_0 \to \rho_1$, $T_2 : \rho_0 \to \rho_2$ belong to an admissible family of deformations, and the conditional paths satisfy $z_{t,s}(a) = (1-t-s)a + tT_1(a) + sT_2(a)$. Then, under $\sigma \to 0$, PiFM approximates the Wasserstein barycenter between (ρ_0, ρ_1, ρ_2), i.e , $\hat{\rho}_{t,s} = \Gamma_{t,s}(\rho_0)$ satisfies $\hat{\rho}_{t,s} = WB_\lambda(\rho_0, \rho_1, \rho_2)$. Additionally, the optimal transport cost is*

$$\int \left[(1-t-s) \left\| z_{t,s}(a) - a \right\|^2 + t \left\| z_{t,s}(a) - T_1(a) \right\|^2 + s \left\| z_{t,s}(a) - T_2(a) \right\|^2 \right] d\rho_0(a).$$

Additional results relating PiFM to the Wasserstein barycenter are presented in Appendix A.3 and Appendix B. Since computing the Wasserstein barycenter is computationally expensive, the results in this section highlight the potential of PiFM as a neural network-based alternative for approximating Wasserstein barycenters. Section B.1 provides an initial exploration of the computational time differences, while a more in-depth analysis is left for future work.

4 Experiments

We present empirical results for low- and high-dimensional data settings. In the low-dimensional setting, we demonstrate that PiFM outperforms Meta Flow Matching (MFM) in (Atanackovic et al., 2025) on a toy dataset under domain shift and show how PiFM can be incorporated with Curly Flow Matching Petrović et al. (2025). In the high-dimensional case, we evaluate PiFM's performance on unsupervised image translation and demonstrate its superiority over traditional CFM in data extrapolation. We also evaluate on a real-world single-cell RNA-seq reprogramming dataset (Schiebinger et al., 2019), where PiFM models cellular states while jointly predicting experimental day and pluripotency score, testing whether the learned dynamics capture both chronological progression and fate acquisition toward successful reprogramming. In the following experiments we use OT couplings between (a, b) and between (a, c) for training PiFM.

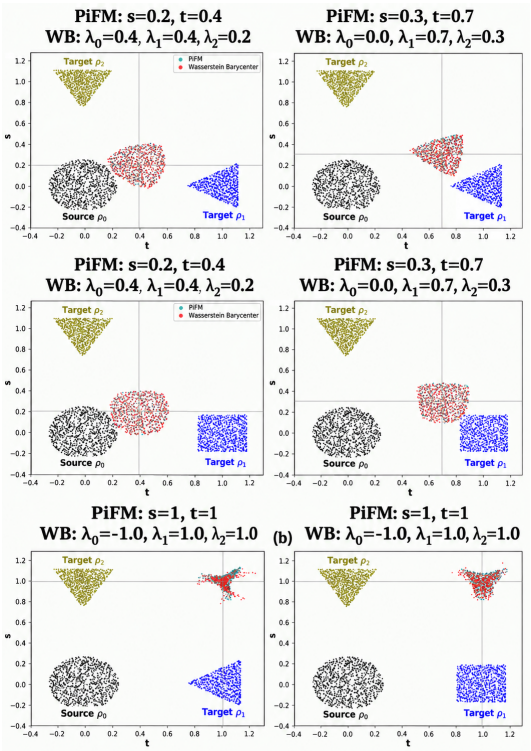


Figure 2: Comparison between output distributions from PiFM model (ours) and Wasserstein Barycenters for a) within and b) outside the probability simplex . The predicted distributions from the PiFM model correspond to the Wasserstein barycenters.

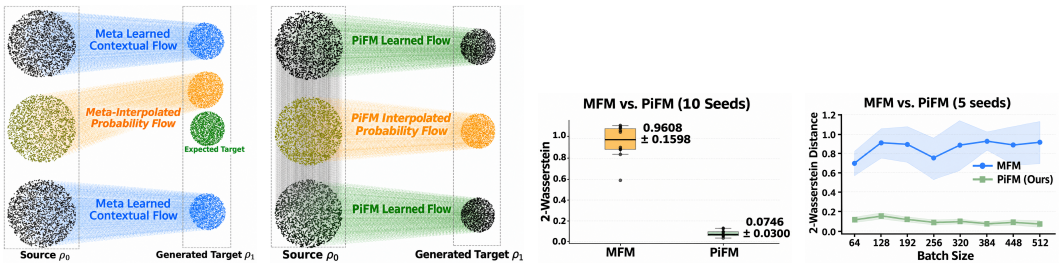
4.1 Low Dimensional Setting

To assess PiFM's ability to infer samples under out-of-distribution conditions, we use synthetic data in which the distribution is shifted spatially in one direction and changes shape and shifts simultaneously in another direction.

4.1.1 PiFM for Inferring Unseen Population on Shifted Domain

Toy dataset. We construct unit discs with centers at (0, 0) and (0, 5) as training source distributions and small discs $(r = 0.5)$ with centers at (0, 5) and (5, 5) as training target distributions. Given the unit disc with center at (0, 2.5) as the unseen distribution in the source, the goal is to infer the small disc with center at (5, 2.5) as expected target.

Comparison. Figure 3a shows that MFM has the correct flow (in blue) and target samples in the training set (top and bottom row), but misses the target on the test set (middle row). On the contrary,



(a) MFM (b) PiFM (ours) (c) Single Batch-size (d) Across Batch-size

Figure 3: Comparison of Meta Flow Matching (MFM) and PiFM on a toy dataset, including qualitative trajectories and quantitative evaluation. Panels (a) and (b) show that MFM generalizes poorly to unseen source embeddings, while PiFM maintains consistent trajectories. Panels (c) and (d) report the 2-Wasserstein distance between ground truth and generated samples, showing that PiFM achieves more accurate and consistent transport.

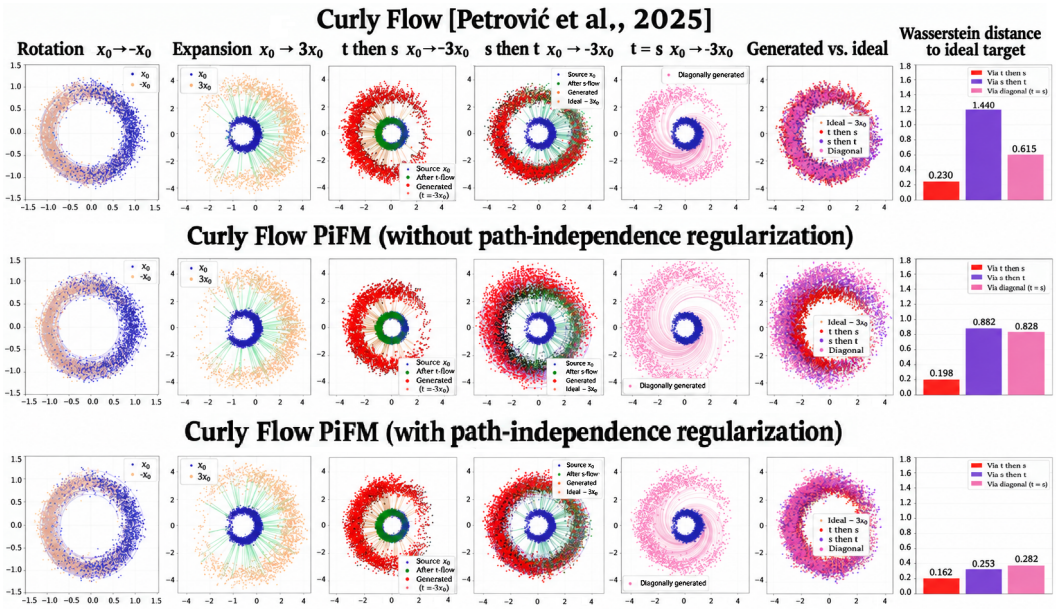


Figure 4: Illustration of PiFM in the Curly Flow Matching setting (Petrović et al., 2025) together with a quantitative evaluation of path dependence. Curly-FM and unregularized PiFM produce path-dependent dynamics and inconsistent final distributions across integration strategies. In contrast, PiFM with path-independence regularization yields consistent transport and the lowest value in Wasserstein-2 distance to the target distribution.

Figure 3b shows that PiFM predicts the correct flow and target from the unseen source distribution (middle row) by inferring the source distributions as samples on a "base" flow over the vertical direction. Moreover, we systematically evaluate the 2-Wasserstein distance between the predicted and expected target distributions over multiple runs of varying seeds. For the same number of batches (= 512) and training epochs (= 3000) for both MFM and PiFM over 10 random seeds, Figure 3c shows that the mean and standard deviation of 2-Wasserstein distance are much smaller for PiFM than for MFM, indicating the PiFM is more robust and accurate in this task. For same number of training epochs (= 2000) with varying batch sizes, PiFM consistently shows better performance (Figure 3d).

4.1.2 PiFM for Inferring Desirable Path-independent Distributions

Toy dataset. We evaluate PiFM in the Curly Flow Matching setting (Petrović et al., 2025), where the goal is to compose a rotation flow ($x_0 \to -x_0$) and an expansion flow ($x_0 \to 3x_0$) to produce a joint transformation $x_0 \to -3x_0$. A distributionally path-independent model should yield the same target distribution regardless of the integration strategy used at inference time.

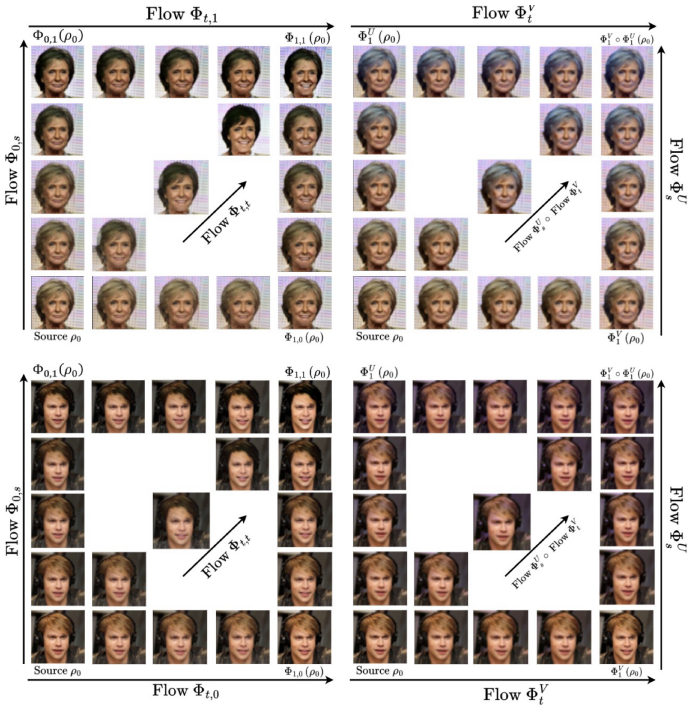
Comparison. As shown in Figure 4, the original Curly-FM model, which trains the two vector fields separately, fails to achieve commutativity: sequential and diagonal integration paths produce inconsistent results, revealing path-dependent dynamics. PiFM without regularization also fails to recover commutativity despite joint training. In contrast, PiFM with path-independence regularization produces consistent outputs across all paths, with generated distributions closely matching the target

distribution. This is further confirmed quantitatively in Figure 4, where Curly-FM and unregularized PiFM exhibit large variations in Wasserstein-2 distance across integration strategies. PiFM with path-independence regularization achieves the lowest values, demonstrating that enforcing commutativity during training is key to reliable path-independent generation.

4.2 High-Dimensional Setting: PiFM for Unsupervised Image-to-Image Translation

We train and evaluate on the CelebA (Liu et al., 2015) dataset using the `Smiling` and `Black_Hair` attributes. We employ a shared U-Net backbone with two per-pixel MLP heads implemented as 1×1 convolutions with SiLU activations, each predicting the vector field associated with one attribute flow. A full description of the preprocessing and architecture is provided in Appendix B.

Figure 5 shows qualitative results for female (panel (a) top and (b) top) and male (panel (a) bottom and (b) bottom) examples across the three different flow integration strategies: $t \to s$, $s \to t$ and diagonal. In the PiFM results, the A-field (`Smiling`) produces localized mouth/cheek edits consistent with smiling, while the B-field (`Black_Hair`) darkens and re-textures the hair region. Note that the identity (global face shape, eyes) is largely preserved. The diagonal integration often produces smoother simultaneous edits to both attributes. In contrast, CFM often fails to modify the source image along either one or both flow axes (e.g., CFM cannot generate smiling faces in Figure 5). The consistent inference of attributes in PiFM across samples and integration paths reiterates that PiFM has strong empirical path-independence properties at the distribution level. Likewise, individually each sample shows that across different integration paths approximately the same picture is obtained at the end, which indicates at least approximately pointwise path-independence is attained. Additional comparisons using FID are presented in Appendix B.2.3. We find FID poorly suited for evaluating commutativity as in Figure 5, since baseline methods may generate images close in pixel space without achieving the desired attribute transformations.



(a) **PiFM (ours)** (b) **CFM**

Figure 5: We present an image-to-image translation comparison of PiFM on the CelebA (Liu et al., 2015) dataset under three different flow integration strategies. PiFM successfully infers the missing attribute, whereas CFM fails to produce the desired transformation.

4.3 High-Dimensional Multi-Step Setting: PiFM for Single-Cell Reprogramming

We further evaluate PiFM on a real-world single-cell RNA-seq reprogramming dataset (Schiebinger et al., 2019). Data source and preprocessing details are provided in Appendix B.3.1. Cells are represented in 50-dimension PCA space, and the two biological axes are experimental day and pluripotency score. This setting provides a high-dimensional, multi-step biological test case: experimental day captures chronological reprogramming, while pluripotency score reflects acquisition of the target iPSC-like fate. Thus, the task evaluates whether PiFM can jointly model temporal dynamics and fate acquisition, rather than treating reprogramming as a single ordered trajectory.

As shown in Figure 6, PiFM-inferred trajectories evolve coherently along both biological axes, capturing temporal reprogramming at fixed pluripotency and fate acquisition at fixed experimental day. Moreover, when transporting day 0 low-pluripotency cells to day 6 high-pluripotency targets, different integration orders yield overlapping trajectories and generated distributions that align with

the real target population, supporting distributional path-independent generation. Biologically, this suggests that PiFM can model reprogramming as a multi-step and heterogeneous process, where cells observed at the same experimental day may occupy different intermediate states and only a subset progresses toward the pluripotent iPSC-like fate. We evaluate target prediction quality and commutativity in Appendix B.3.2.

5 Discussion

Conclusion We introduced Path-Independent Flow Matching (PiFM), a multi-parameter Flow Matching that enforces consistent path-independent transport of distributions. We also make a connection between PiFM and Wasserstein barycenters. We showed that empirically, PiFM outperforms existing approaches on synthetic and image translation tasks, particularly in order-invariant and out-of-distribution settings. We further demonstrated applications to Curly Flow Matching and single-cell RNA-seq trajectory inference.

Limitations While our theory provides conditions for path-independent transport at the distribution level, it does not guarantee path independence for individual samples. Moreover, the connection to the Wasserstein barycenter is only studied in the $n = 2$ setting; further investigation is needed to establish the result in this work for the more general case.

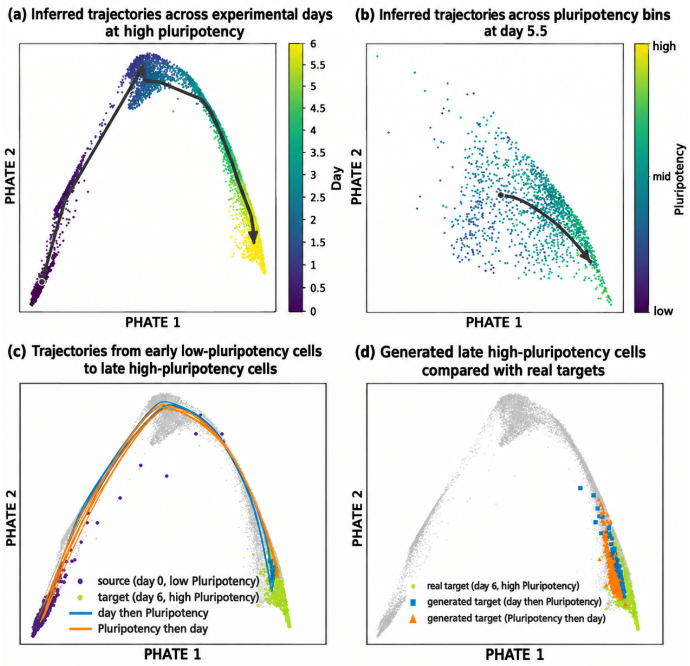


Figure 6: PiFM-inferred single-cell reprogramming dynamics across experimental day and pluripotency progression, visualized by PHATE Moon et al. (2019). (a) and (b) show trajectories along each biological axis, varying experimental day at fixed high pluripotency and varying pluripotency bins at fixed experimental day. (c) and (d) show path-ordered generation from day 0 low-pluripotency cells to day 6 high-pluripotency targets, where the two integration orders produce overlapping trajectories and generated distributions that match the real target population.

Future Work Promising future directions include extending PiFM to more general geometries, as in Chen & Lipman (2024), and further applying it to single-cell perturbation data and trajectory inference tasks. Additionally, we will focus on validating PiFM as a Wasserstein barycenter generative framework beyond the setting discussed here.

Aknowledgements We thank Professor Paul François for his mentorship and support of Philippe in his capacity as a co-supervisor during the course of this project. This research was supported in part by Canada CIFAR AI Chair [G.W.], NSERC Discovery Grant 03267 [G.W.], NSF grant DMS-2327211 [G.W.], FRQNT–NSERC grant 2023-NOVA-329125 [G.W.; E.B.], Open Philanthropy [E.B.; Y.Z.] and NSERC Discovery grant 05439 [S.S.]. The authors are solely responsible for the content of this work, and the views expressed in it do not necessarily reflect those of the funding agencies. This research was enabled in part by compute resources provided by Mila, Université de Montréal, Compute Canada, and Calcul Québec. We sincerely thank the anonymous reviewers and Area Chairs for their constructive feedback and efforts throughout the review process, which helped us improve the paper.

Impact statement The authors acknowledge that this work may have broader societal implications; however, no implications are identified as requiring further specific discussion in the context of this manuscript.

Code availability Our source code is publicly available at github.com/YanleiZhang01/PiFM.

References

Abraham, R., Marsden, J. E., and Ratiu, T. *Manifolds, Tensor Analysis, and Applications*. Springer New York, NY, 2 edition, 1988.

Agueh, M. and Carlier, G. Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924, 2011.

Albergo, M. S. and et al. Multimarginal generative modeling with stochastic interpolants. *arXiv preprint arXiv:2310.03695*, 2023.

Albergo, M. S. and Vanden-Eijnden, E. Building normalizing flows with stochastic interpolants, 2023. URL https://arxiv.org/abs/2209.15571.

Atanackovic, L., Zhang, X., Amos, B., Blanchette, M., Lee, L. J., Bengio, Y., Tong, A., and Neklyudov, K. Meta flow matching: Integrating vector fields on the wasserstein manifold, 2025. URL https://arxiv.org/abs/2408.14608.

Beier, F., von Lindheim, J., Neumayer, S., and Steidl, G. Unbalanced multi-marginal optimal transport. *Journal of Mathematical Imaging and Vision*, 65(3):394–413, October 2022. ISSN 1573-7683. doi: 10.1007/s10851-022-01126-7. URL http://dx.doi.org/10.1007/s10851-022-01126-7.

Berner, J., Richter, L., and Ullrich, K. An optimal control perspective on diffusion-based generative modeling. In *NeurIPS 2022 Workshop on Score-Based Methods*, 2022. URL https://openreview.net/forum?id=jlgs_E-0UdM.

Bigot, J. and Klein, T. Characterization of barycenters in the wasserstein space by averaging optimal transport maps, 2017. URL https://arxiv.org/abs/1212.2562.

Boissard, E., Gouic, T. L., and Loubes, J.-M. Distribution's template estimate with wasserstein metrics, 2013. URL https://arxiv.org/abs/1111.5927.

Bortoli, V. D., Thornton, J., Heng, J., and Doucet, A. Diffusion schrödinger bridge with applications to score-based generative modeling, 2023. URL https://arxiv.org/abs/2106.01357.

Cao, J., Mo, L., Zhang, Y., Jia, K., Shen, C., and Tan, M. Multi-marginal wasserstein gan. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Chen, R. T. and Lipman, Y. Flow matching on general geometries. In *12th International Conference on Learning Representations, ICLR 2024 ; Conference date: 07-05-2024 Through 11-05-2024*, January 2024.

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. Neural ordinary differential equations, 2019. URL https://arxiv.org/abs/1806.07366.

Cuturi, M. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 2292–2300, 2013.

Cuturi, M. and Doucet, A. Fast computation of wasserstein barycenters. In *Proceedings of Machine Learning Research*, volume 32, pp. 685–693. Journal of Machine Learning Research, 2014.

Diederik, P. K. and Max, W. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, November 2019. ISSN 1935-8245. doi: 10.1561/2200000056. URL http://dx.doi.org/10.1561/2200000056.

Fatras, K., Sejourne, T., Flamary, R., and Courty, N. Unbalanced minibatch optimal transport; applications to domain adaptation. In Meila, M. and Zhang, T. (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 3186–3197. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/fatras21a.html.

Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., et al. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021. URL http://jmlr.org/papers/v22/20-451.html.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. Generative adversarial networks, 2014. URL https://arxiv.org/abs/1406.2661.

Grathwohl, W., Chen, R. T. Q., Bettencourt, J., Sutskever, I., and Duvenaud, D. Ffjord: Free-form continuous dynamics for scalable reversible generative models, 2018. URL https://arxiv.org/abs/1810.01367.

Haviv, D., Pooladian, A.-A., Pe'er, D., and Amos, B. Wasserstein flow matching: Generative modeling over families of distributions, 2024. *URL https://arxiv. org/abs/2411.00698*, 2025.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models, 2020. URL https://arxiv.org/abs/2006.11239.

Jiang, L., Dalgarno, C., Papalexi, E., Mascio, I., Wessels, H.-H., Yun, H., Iremadze, N., Lithwick-Yanai, G., Lipson, D., and Satija, R. Systematic reconstruction of molecular pathway signatures using scalable single-cell perturbation screens. *Nature Cell Biology*, 27:505–517, 02 2025. doi: 10.1038/s41556-025-01622-z.

Kingma, D. P. and Welling, M. Auto-encoding variational bayes, 2022. URL https://arxiv.org/abs/1312.6114.

Koehler, F., Heckett, A., and Risteski, A. Statistical efficiency of score matching: The view from isoperimetry, 2022. URL https://arxiv.org/abs/2210.00726.

Lee, J. M. Introduction to smooth manifolds. In *Introduction to smooth manifolds*, pp. 1–29. Springer, 2003.

Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling, 2023. URL https://arxiv.org/abs/2210.02747.

Liu, G.-H., Vahdat, A., Huang, D.-A., Theodorou, E. A., Nie, W., and Anandkumar, A. I^2sb: Image-to-image schrödinger bridge, 2023. URL https://arxiv.org/abs/2302.05872.

Liu, Q. Rectified flow: A marginal preserving approach to optimal transport, 2022. URL https://arxiv.org/abs/2209.14577.

Liu, Z., Luo, P., Wang, X., and Tang, X. Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pp. 3730–3738, 2015.

Lotfollahi, M., Klimovskaia Susmelj, A., De Donno, C., Hetzel, L., Ji, Y., Ibarra, I. L., Srivatsan, S. R., Naghipourfar, M., Daza, R. M., Martin, B., et al. Predicting cellular responses to complex perturbations in high-throughput screens. *Molecular Systems Biology*, pp. e11517, 2023.

Léonard, C. A survey of the schrödinger problem and some of its connections with optimal transport, 2013. URL https://arxiv.org/abs/1308.0215.

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

Neary, C. and Topcu, U. Compositional learning of dynamical system models using port-hamiltonian neural networks, 2023. URL https://arxiv.org/abs/2212.00893.

Pass, B. Multi-marginal optimal transport: Theory and applications. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1771–1790, 2015. doi: 10.1051/m2an/2015020. URL https://www.numdam.org/articles/10.1051/m2an/2015020/.

Petrović, K., Atanackovic, L., Moro, V., Kapuśniak, K., İsmail İlkan Ceylan, Bronstein, M., Bose, A. J., and Tong, A. Curly flow matching for learning non-gradient field dynamics, 2025. URL https://arxiv.org/abs/2510.26645.

S, K. and Durgadevi, M. Generative adversarial network (gan): a general review on different variants of gan and applications. In *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, pp. 1–8, 2021. doi: 10.1109/ICCES51350.2021.9489160.

Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., Lee, L., Chen, J., Brumbaugh, J., Rigollet, P., Hochedlinger, K., Jaenisch, R., Regev, A., and Lander, E. S. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943.e22, 2019. doi: 10.1016/j.cell.2019.01.006.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. Score-based generative modeling through stochastic differential equations, 2021. URL https://arxiv.org/abs/2011.13456.

Tong, A., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Fatras, K., Wolf, G., and Bengio, Y. Improving and generalizing flow-based generative models with minibatch optimal transport, 2023.

Tong, A., Malkin, N., Fatras, K., Atanackovic, L., Zhang, Y., Huguet, G., Wolf, G., and Bengio, Y. Simulation-free schrödinger bridges via score and flow matching, 2024. URL https://arxiv.org/abs/2307.03672.

Wang, Y., Li, Z., Zhang, Z., Zhang, W., Xie, B., Liu, X., Zeng, W., and Jin, X. Scene graph disentanglement and composition for generalizable complex image generation, 2024. URL https://arxiv.org/abs/2410.00447.

Warner, F. *Foundations of Differentiable Manifolds and Lie Groups*. Graduate Texts in Mathematics. Springer, 1983. ISBN 9780387908946. URL https://books.google.ca/books?id=iaeUqc2yQVQC.

Wiggins, S. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, volume 2 of *Texts in Applied Mathematics*. Springer, New York, 2003. ISBN 978-0-387-00177-7.

A Theoretical Details and Proofs

A.1 The general multi-parameter case

PiFM generalizes to the case of multiple parameters $t_1, \ldots, t_n$. In this setting, we consider a source data distribution $q(a)$ and target data distributions $q(b_1), \ldots, q(b_n)$. We condition on the joint distribution $q(z)$, where $z = (a, b_1, \ldots, b_n)$, with a marginalization,

$$p_{t_1,\ldots,t_n}(x) = \int p_{t_1,\ldots,t_n}(x \mid z) q(z)\, dz. \tag{10}$$

The continuously differentiable vector fields $u^{(i)}_{t_1,\ldots,t_n}$ are associated with the ODEs

$$\frac{\partial \phi^{(i)}(x)_{t_1,\ldots,t_n}}{\partial t_i} = u^{(i)}_{t_1,\ldots,t_n}\big(\phi_{t_1,\ldots,t_n}(x)\big) \tag{11}$$

Each vector field $u^{(i)}_{t_1,\ldots,t_n}$ governs the evolution from the source distribution $q(a)$ to the target distribution $q(b_i)$ and has an associated flow $\Phi_{t_1,\ldots,t_n} : [0,1]^n \times \mathbb{R}^d \to \mathbb{R}^d$, $\Phi_{t_1,\ldots,t_n}(x) = \phi_{t_1,\ldots,t_n}(x)$. We define

$$u^{(i)}_{t_1,\ldots,t_n}(x) = \mathbb{E}_{q(z)}\left[\frac{u^{(i)}_{t_1,\ldots,t_n}(x|z)\, p_{t_1,\ldots,t_n}(x|z)}{p_{t_1,\ldots,t_n}(x)}\right] \tag{12}$$

for all $i = 1, \ldots, n$.

Here, $u^{(i)}_{t_1,\ldots,t_n}(x|z)$ are the conditional vector fields. We consider the n continuity equations

$$\frac{\partial p_{t_1,\ldots,t_n}(x)}{\partial t_i} + \nabla \cdot \big(p_{t_1,\ldots,t_n}(x)\, u^{(i)}_{t_1,\ldots,t_n}(x)\big) = 0, \tag{13}$$

which hold for all $t_j \in [0,1]$ with $j \neq i$. Additionally, the path-independence property (4) is replaced by a family of equalities corresponding to integrating along each parameter direction sequentially, resulting in $n!$ equivalent expressions. Finally, (5) is replaced by $\binom{n}{2}$ conditions encoding pairwise consistency between vector fields. For all $i, j = 1, \ldots, n$, $i \neq j$, we obtain

$$\partial_{t_j} u^{(i)}_{t_1,\ldots,t_n}(x) - \partial_{t_i} u^{(j)}_{t_1,\ldots,t_n}(x) = [u^{(i)}_{t_1,\ldots,t_n}(x), u^{(j)}_{t_1,\ldots,t_n}(x)] \tag{14}$$

The generalized version of the affine probability path and conditional vector fields is as follows. We take

$$u^{(i)}_{t_1,\ldots,t_n}(x|z) = b_i - a \quad , \forall i = 1, ..., n$$

and

$$p_{t_1,\ldots,t_n}(x|z) = \mathcal{N}(x|\mu_{t_1,\ldots,t_n}(z), \sigma^2)$$

with constant $\sigma > 0$ and $\mu_{t_1,\ldots,t_n}(z) = a + \sum_{i=1}^n (b_i - a) t_i$. As discussed in the main part of the paper, the choice of conditional vector fields and probability paths is a user defined choice as long as there is consistent and path-independent distribution transport.

A.2 Proofs of theorems

Lemma A.1 (General form Lemma 3.1)**.** *Consider $u^{(i)}_{t_1,\ldots,t_n}(x)$ vector fields. If there exists a unique generated probability density path $p_{t_1,\ldots,t_n}(x)$ satisfying* (13)*, then the associated flows $\Phi^{(i)}_{t_1,\ldots,t_n}(x)$ satisfy the path-independence condition.*

Proof. Suppose that the vector fields $u^{(i)}_T(x)$ generate $p_T(x)$,

$$\frac{\partial p_T(x)}{\partial t_i} + \nabla \cdot \big(p_T(x)\, u^{(i)}_T(x)\big) = 0, \qquad \forall i = 1, \ldots, n.$$

There are $n!$ possible orderings in which one can successively transport along each coordinate, starting from $(0, \ldots, 0)$ and reaching $T = (t_1, \ldots, t_n)$. Without loss of generality, it suffices to consider the ordering $t_1 \to t_2 \to \cdots \to t_n$, as all other cases follow analogously.

We proceed iteratively. First, we transport using the flow $\Phi^{(1)}_{t_1,0,\dots,0}$ which is associated to the vector field $u^{(1)}_T$, obtaining $(\Phi^{(1)}_{t_1,0,\dots,0})_{\#}p_{0,\dots,0}$. Next, we transport using $\Phi^{(2)}_{t_1,t_2,0,\dots,0}$, yielding $(\Phi^{(2)}_{t_1,t_2,0,\dots,0})_{\#} \circ (\Phi^{(1)}_{t_1,0,\dots,0})_{\#}\, p_{0,\dots,0}$. Continuing in this manner, we arrive at the expression

$$(\Phi^{(n)}_{t_1,\dots,t_n})_{\#} \circ \cdots \circ (\Phi^{(2)}_{t_1,t_2,0,\dots,0})_{\#} \circ (\Phi^{(1)}_{t_1,0,\dots,0})_{\#}\, p_{0,\dots,0}.$$

Since the vector fields $u^{(i)}_T$ jointly generate $p_T(x)$, it follows that at $T = (t_1, \dots, t_n)$ we have

$$(\Phi^{(n)}_{t_1,\dots,t_n})_{\#} \circ \cdots \circ (\Phi^{(2)}_{t_1,t_2,0,\dots,0})_{\#} \circ (\Phi^{(1)}_{t_1,0,\dots,0})_{\#}\, p_{0,\dots,0} = p_{t_1,\dots,t_n}.$$

Repeating the same argument for the remaining $n! - 1$ orderings, and noting that each construction yields $p_{t_1,\dots,t_n}$, we conclude that all $n!$ resulting expressions coincide. □

Proposition A.2 (General form of Proposition 3.2)**.** *Suppose that the vector fields $u^{(i)}_{t_1,\dots,t_n}(x)$ satisfy the integrability condition* (14) *Then there exists a unique flow $\Phi_{t_1,\dots,t_n}$ which satisfies the path-independence property.*

Proof. Suppose that the vector fields $u^{(i)}_T$ satisfy the integrability condition

$$\partial_{t_i} u^{(i)}_{t_1,\dots,t_n} - \partial_{t_j} u^{(j)}_{t_1,\dots,t_n} = [u^{(i)}_{t_1,\dots,t_n}, u^{(j)}_{t_1,\dots,t_n}]$$

By Proposition 19.29 in (Lee, 2003), there exists a unique solution $\phi_{t_1,\dots,t_n}$ to the system

$$\frac{\partial \phi_{t_1,\dots,t_n}(x)}{\partial t_i} = u^{(i)}_{t_1,\dots,t_n}\big(\phi_{t_1,\dots,t_n}(x)\big), \qquad \forall i = 1, \dots, n,$$

In this case, the map $\Phi : [0,1]^n \times \mathbb{R}^d \to \mathbb{R}^d$, defined by

$$\Phi_{t_1,\dots,t_n}(x) = \phi_{t_1,\dots,t_n}(x),$$

is the unique flow associated with the family of vector fields $u^{(i)}_T$. It follows that transporting along the flows $\Phi^{(i)}$, which in this case is Φ, in any order yields the same result. □

Theorem A.3 (General form of Theorem 3.3)**.** *Let the conditional vector fields $u^{(i)}_{t_1,\dots,t_n}(x|z)$ generate the unique conditional probability density path $p_{t_1,\dots,t_n}(x|z)$ satisfying the analogous equations* (13)*. Then, the vector fields $u^{(i)}_{t_1,\dots,t_n}(x)$ in* (12) *generate the unique probability density path $p_{t_1,\dots,t_n}(x)$ satisfying* (13)*.*

Proof. Suppose that the conditional vector fields $u^{(i)}_{t_1,\dots,t_n}(x|z)$ generate $p_{t_1,\dots,t_n}(x|z)$. That is,

$$\frac{\partial p_{t_1,\dots,t_n}(x|z)}{\partial t_i} + \nabla \cdot (p_{t_1,\dots,t_n}(x|z) u^{(i)}_{t_1,\dots,t_n}(x|z)) = 0 \quad , \forall i = 1, ..., n.$$

Then,

$$\begin{aligned}
-\nabla \cdot (p_{t_1,\dots,t_n}(x) u^{(i)}_{t_1,\dots,t_n}(x)) &= -\nabla \cdot \left(p_{t_1,\dots,t_n}(x) \int u^{(i)}_{t_1,\dots,t_n}(x|z) \frac{p_{t_1,\dots,t_n}(x|z) q(z)}{p_{t_1,\dots,t_n}(x)} dz \right) \\
&= -\nabla \cdot \left(\int u^{(i)}_{t_1,\dots,t_n}(x|z) p_{t_1,\dots,t_n}(x|z) q(z) dz \right) \\
&= -\int \nabla \cdot \left(u^{(i)}_{t_1,\dots,t_n}(x|z) p_{t_1,\dots,t_n}(x|z) q(z) \right) dz \\
&= -\int \nabla \cdot \left(u^{(i)}_{t_1,\dots,t_n}(x|z) p_{t_1,\dots,t_n}(x|z) \right) q(z) dz \\
&= \int \frac{\partial p_{t_1,\dots,t_n}(x|z)}{\partial t_i} q(z) dz \\
&= \frac{\partial}{\partial t_i} \int p_{t_1,\dots,t_n}(x|z) q(z) dz \\
&= \frac{\partial p_{t_1,\dots,t_n}(x)}{\partial t_i}.
\end{aligned}$$

Therefore, the vector fields $u^{(i)}_{t_1,\dots,t_n}$ generate $p_{t_1,\dots,t_n}(x)$. □

Corollary A.4 (General form of Corollary 3.4)**.** *Suppose that the vector fields $u_{t_1,\dots,t_n}(x|z)$ generate the unique conditional probability density path $p_{t_1,\dots,t_n}(x|z)$ satisfying the analogous equations* (13)*. Then, the associated flows $\Phi^{(i)}_{t_1,\dots,t_n}$ of the vector fields $u^{(i)}_{t_1,\dots,t_n}(x)$ given in* (12) *satisfy the path-independence property.*

Proof. Suppose that the conditional vector fields $u^{(i)}_{t_1,\dots,t_n}(x \mid z)$ generate $p_{t_1,\dots,t_n}(x \mid z)$. By Theorem 3.3, it follows that the vector fields $u^{(i)}_{t_1,\dots,t_n}(x)$ generate $p_{t_1,\dots,t_n}(x)$. Finally, by Lemma 3.1, transporting along the flows $\Phi^{(i)}_{t_1,\dots,t_n}$ in any order yields the same result. □

A.3 Discussion about the relation between PiFM and Wasserstein Barycenter

We hereafter discuss Theorem 3.5 and the connection between PiFM and Wasserstein Barycenter.

Lemma A.5 (Lemma 3.1 of Boissard et al. (2013))**.** *Let $\Pi(\mu_1,\dots,\mu_K)$ be the set of probability measures on $(\mathbb{R}^d)^K$ with marginals $\mu_1,\dots,\mu_K$, respectively, and*

$$T(x_1,\dots,x_K) = \sum_{1\le j\le K} \lambda_j x_j$$

with $\lambda_j \ge 0$ such that $\sum_{1\le j\le K} \lambda_j = 1$. A probability measure ν is a barycenter of $\mu_1,\dots,\mu_K$ with weights $(\lambda_j)_{1\le j\le K}$ if and only if $\nu = T\#\gamma$ where $\gamma \in \Pi(\mu_1,\dots,\mu_K)$ minimizes the multi marginal Optimal Transport:

$$J(\gamma,\lambda) = \int \sum_{j=1}^{K} \lambda_j \left\| T(x_1,\dots,x_K) - x_j \right\|^2 d\gamma(x_1,\dots,x_K) \tag{15}$$

By choosing T as $T(a,b,c) = a+t(b-a)+s(c-a) = (1-t-s)a+tb+sc$ and $\lambda = (1-t-s,t,s)$ we have:

$$J(\gamma,\lambda) = \int \left[(1-t-s)\|T(a,b,c)-a\|^2+t\|T(a,b,c)-b\|^2+s\|T(a,b,c)-c\|^2\right] d\rho_0(a)\rho_1(b)\rho_2(c)$$

With Optimal Coupling such that $b = T_1(a)$ and $c = T_2(a)$, T only depends on a and we have $T(a,b,c) = z_{t,s}(a)$ and the associated multi marginal transport is defined as

$$J^{\text{PiFM}}(\rho_0,\lambda) := \int \left[(1-t-s)\|z_{t,s}(a)-a\|^2 + t\|z_{t,s}(a)-T_1(a)\|^2 + s\|z_{t,s}(a)-T_2(a)\|^2\right] d\rho_0(a) \tag{16}$$

Without specific assumptions, $J^{\text{PiFM}}(\rho_0,\lambda)$ is not guaranteed to be a minimizer of (15). In the following, we give sufficient conditions referred to as the admissible family of deformations under which $J^{\text{PiFM}}(\rho_0,\lambda)$ minimizes (15).

Theorem A.6 (Theorem 4.1 of Boissard et al. (2013))**.** *Suppose that $(T_i)_{1\le i\le K}$ is an admissible family of deformations on a domain $\Omega \subset \mathbb{R}^n$, and let $\mu \in \mathcal{P}_2(\Omega)$ be absolutely continuous with respect to the Lebesgue measure. Let $\mu_j = (T_j)_{\#}\mu$, $\mu = (\mu_j)_{1\le j\le K}$ and $\lambda = (\lambda_j)_{1\le j\le K}$. Then,*

$$WB(\mu,\lambda) = \left(\sum_{j=1}^{K} \lambda_j T_j\right)_{\#} \mu \tag{17}$$

The notion of an admissible family of deformations is defined in Boissard et al. (2013) (Definition 4.2) as follows:

1. there exists j_0 such that $T_{j_0} = \text{Id}$
2. for each $1 \le j \le K$: T_j is one-to-one and onto,
3. the composition $T_i \circ T_j^{-1}$ of two maps in this class remains an optimal one for all $1 \le i,j \le K$

We now prove Theorem 3.5.

Proof of Theorem 3.5. Let's first recall that with optimal coupling such that $b = T_1(a)$ and $c = T_2(a)$ and in the limit of $\sigma \to 0$, we have $T(a, b, c) = z_{t,s}(a) = a + (1-t)T_1(a) + (1-s)T_2(b)$. Therefore, by construction:

$$\hat{\rho}_{t,s} := \Gamma_{t,s}(\rho_0) \tag{18}$$

$$:= ((1-t-s)\mathrm{Id} + tT_1 + sT_2))_{\#}\rho_0 \tag{19}$$

where we recall that $\Gamma_{t,s}(\rho_0) = (\Phi_{t,s})_{\#} \circ (\Psi_{0,s})_{\#}\rho_0 = (\Psi_{t,s})_{\#} \circ (\Phi_{t,0})_{\#}\rho_0$. Equation (19) can be rewritten as

$$\hat{\rho}_{t,s} = \left(\sum_{j=1}^{K} \lambda_j T_j \right)_{\#} \rho_0$$

Assuming that T_1 and T_2 belong to the admissible family of deformation, using Theorem A.6 we get that $\hat{\rho}_{t,s}$ is exactly the Wasserstein Barycenter, i.e PiFM generates the Wasserstein barycenter of ρ_0, ρ_1, ρ_2 with $\lambda = (1-t-s, t, s)$, this concludes the proof.

Additionally, using (16) and applying Lemma A.5, we get:

$$\begin{aligned} J^{\text{PiFM}}(\rho_0, \lambda) &= \int \left[(1-t-s) \left|\left| z_{t,s}(a) - a \right|\right|^2 + t \left|\left| z_{t,s}(a) - T_1(a) \right|\right|^2 + s \left|\left| z_{t,s}(a) - T_2(a) \right|\right|^2 \right] d\rho_0(a) \\ &= \min_{\gamma} J(\gamma, \lambda) \end{aligned}$$

□

A.4 Additional results relating PiFM to the Wasserstein Barycenter

We note that not only the model predicts the Wasserstein barycenter within the simplex defined by ρ_0, ρ_1 and ρ_2, but we also empirically observe that for toy distributions outside of this simplex, $\rho_{t,s}^{\text{PiFM}}$ coincides with the horizontal barycenter $\rho_s^U(t)$ (resp. vertical barycenter $\rho_t^V(s)$) after applying $\Gamma_{0,s}$ (resp. $\Gamma_{t,0}$) to the initial and target distributions as shown in Supplementary Figure 12 and in Supplementary Figure 13. Below, we discuss the Gaussian case with proof related to horizontal and vertical Wasserstein barycenter.

Theorem A.7. *Let $m_0, m_1, m_2 \in \mathbb{R}^d$. Suppose that there exists a symmetric positive semidefinite matrix $\Sigma \in \mathbb{R}^{d \times d}$ such that*

$$\rho(0) = \mathcal{N}(m_0, \Sigma), \qquad \rho(1) = \mathcal{N}(m_1, \Sigma), \qquad \rho(2) = \mathcal{N}(m_2, \Sigma).$$

Define the conditional vector fields

$$\hat{u}_{t,s}(x) := u_{t,s}(x|z) := m_1 - m_0, \qquad \hat{v}_{t,s}(x) := v_{t,s}(x|z) := m_2 - m_0,$$

and, for $(t, s) \in \Delta^2 := \{(t, s) \in [0, 1]^2 : \ t + s \leq 1\}$, define

$$\hat{Q}_{t,s} := \mathcal{N}\Big((1-t-s)m_0 + t\, m_1 + s\, m_2, \Sigma \Big).$$

Let $\mathcal{Q}_{t,s}$ denote the probability measure with density $\hat{Q}_{t,s}$.

Then:

(i) The family $\hat{Q}_{t,s}$ satisfies the two continuity equations

$$\partial_t \hat{Q}_{t,s} + \nabla \cdot (\hat{Q}_{t,s}\, \hat{u}) = 0, \qquad \partial_s \hat{Q}_{t,s} + \nabla \cdot (\hat{Q}_{t,s}\, \hat{v}) = 0,$$

with boundary conditions

$$\hat{Q}_{0,0} = \rho(0), \qquad \hat{Q}_{1,0} = \rho(1), \qquad \hat{Q}_{0,1} = \rho(2).$$

(ii) For every $(t, s) \in \Delta^2$, $\mathcal{Q}_{t,s}$ is the Wasserstein barycenter of $\rho(0), \rho(1), \rho(2)$ with weights $(1-t-s, t, s)$, i.e.,

$$\mathcal{Q}_{t,s} = \operatorname*{arg\,min}_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \Big\{ (1-t-s)\mathcal{W}_2^2(\nu, \rho(0)) + t\, \mathcal{W}_2^2(\nu, \rho(1)) + s\, \mathcal{W}_2^2(\nu, \rho(2)) \Big\}.$$

(iii) The horizontal and vertical slice barycenter identities also hold:

(a) For each fixed $s \in [0,1]$, define

$$A_s := (\Psi_s^0)_{\#}\rho(0), \qquad B_s := (\Psi_s^1)_{\#}\rho(1).$$

Then

$$A_s = \mathcal{N}(m_0 + s(m_2 - m_0), \Sigma), \qquad B_s = \mathcal{N}(m_1 + s(m_2 - m_0), \Sigma),$$

and for every $t \in [0,1]$,

$$\mathcal{Q}_{t,s} = \operatorname*{arg\,min}_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \Big\{ (1-t)\mathcal{W}_2^2(\nu, A_s) + t\,\mathcal{W}_2^2(\nu, B_s) \Big\}.$$

(b) For each fixed $t \in [0,1]$, define

$$C_t := (\Phi_t^0)_{\#}\rho(0), \qquad D_t := (\Phi_t^1)_{\#}\rho(2).$$

Then

$$C_t = \mathcal{N}(m_0 + t(m_1 - m_0), \Sigma), \qquad D_t = \mathcal{N}(m_2 + t(m_1 - m_0), \Sigma),$$

and for every $s \in [0,1]$,

$$\mathcal{Q}_{t,s} = \operatorname*{arg\,min}_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \Big\{ (1-s)\mathcal{W}_2^2(\nu, C_t) + s\,\mathcal{W}_2^2(\nu, D_t) \Big\}.$$

Proof of Theorem A.7. Set

$$\gamma := \mathcal{N}(0, \Sigma), \qquad \tau_m(x) := x + m.$$

Then

$$\rho(i) = (\tau_{m_i})_{\#}\gamma, \qquad i = 0, 1, 2.$$

Step 1: Explicit form of the flow and continuity equations.

Since $\hat{u}$ and $\hat{v}$ are constant vector fields, their flows are translations:

$$\Phi_{t,s}(x) = x + t(m_1 - m_0), \qquad \Psi_{t,s}(x) = x + s(m_2 - m_0).$$

Hence

$$(\Phi_{t,s})_{\#}\rho(0) = \mathcal{N}(m_0 + t(m_1 - m_0), \Sigma),$$

and

$$(\Psi_{t,s})_{\#}\rho(0) = \mathcal{N}(m_0 + s(m_2 - m_0), \Sigma).$$

Hence,

$$\mathcal{Q}_{t,s} = \mathcal{N}\Big(m_0 + t(m_1 - m_0) + s(m_2 - m_0), \Sigma\Big) = \hat{Q}_{t,s}.$$

Since $\hat{Q}_{t,s}$ is a translated Gaussian with constant covariance and its mean evolves according to the constant velocities $\hat{u}, \hat{v}$, it satisfies

$$\partial_t \hat{Q}_{t,s} + \nabla \cdot (\hat{Q}_{t,s}\,\hat{u}) = 0, \qquad \partial_s \hat{Q}_{t,s} + \nabla \cdot (\hat{Q}_{t,s}\,\hat{v}) = 0.$$

The boundary conditions are immediate from the definition.

Step 2: Global barycenter identity.

Let

$$\lambda_0 := 1 - t - s, \qquad \lambda_1 := t, \qquad \lambda_2 := s, \qquad (\lambda_0, \lambda_1, \lambda_2) \in \Delta^2.$$

Set

$$\bar{m}_{t,s} := \lambda_0 m_0 + \lambda_1 m_1 + \lambda_2 m_2.$$

Then

$$\mathcal{Q}_{t,s} = \mathcal{N}(\bar{m}_{t,s}, \Sigma) = (\tau_{\bar{m}_{t,s}})_{\#}\gamma.$$

We claim that $\mathcal{Q}_{t,s}$ minimizes

$$J(\nu, \lambda) := \sum_{i=0}^{2} \lambda_i \mathcal{W}_2^2(\nu, \rho(i)).$$

Take any $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, and let m_ν denote its mean. Write $\tilde{\nu} := (\tau_{-m_\nu})_\# \nu$, so that $\tilde{\nu}$ has mean zero. For each i, since $\rho(i) = (\tau_{m_i})_\# \gamma$, any coupling of ν and $\rho(i)$ may be written in terms of random variables $Y \sim \nu$, $X \sim \gamma$. Expanding the square gives (because the cross term vanishes):

$$\mathbb{E}|Y - (X + m_i)|^2 = \mathbb{E}|(Y - m_\nu) - X|^2 + |m_\nu - m_i|^2,$$

Taking the infimum over couplings yields

$$\mathcal{W}_2^2(\nu, \rho(i)) \geq \mathcal{W}_2^2(\tilde{\nu}, \gamma) + |m_\nu - m_i|^2.$$

Therefore

$$J(\nu, \lambda) \geq \mathcal{W}_2^2(\tilde{\nu}, \gamma) + \sum_{i=0}^{2} \lambda_i |m_\nu - m_i|^2.$$

Since $\sum_i \lambda_i = 1$, the quadratic function

$$m \longmapsto \sum_{i=0}^{2} \lambda_i |m - m_i|^2$$

is uniquely minimized at

$$m = \sum_{i=0}^{2} \lambda_i m_i = \bar{m}_{t,s}.$$

Also,

$$\mathcal{W}_2^2(\tilde{\nu}, \gamma) \geq 0,$$

with equality if and only if $\tilde{\nu} = \gamma$, i.e.

$$\nu = (\tau_{\bar{m}_{t,s}})_\# \gamma = \mathcal{Q}_{t,s}.$$

Thus $\mathcal{Q}_{t,s}$ is the unique minimizer of F.

This step could also be simply demonstrated using Theorem 3.5 as Φ_t^s and Ψ_s^t belong to an admissible family of deformation.

Step 3: Horizontal slice barycenter identity.

Fix $s \in [0, 1]$. Since $\Psi_{t,s}$ is translation by $s(m_2 - m_0)$,

$$A_s = (\Psi_{0,s})_\# \rho(0) = \mathcal{N}(m_0 + s(m_2 - m_0), \Sigma),$$

and

$$B_s = (\Psi_{1,s})_\# \rho(1) = \mathcal{N}(m_1 + s(m_2 - m_0), \Sigma).$$

Their weighted $(1 - t, t)$-barycenter is therefore the Gaussian with the same covariance Σ and mean

$$(1 - t)\big(m_0 + s(m_2 - m_0)\big) + t\big(m_1 + s(m_2 - m_0)\big) = m_0 + t(m_1 - m_0) + s(m_2 - m_0),$$

which is exactly the mean of $\mathcal{Q}_{t,s}$. Hence $\mathcal{Q}_{t,s}$ is the horizontal barycenter.

Step 4: Vertical slice barycenter identity.

Fix $t \in [0, 1]$. Since $\Phi_{t,s}$ is translation by $t(m_1 - m_0)$,

$$C_t = (\Phi_{t,0})_\# \rho(0) = \mathcal{N}(m_0 + t(m_1 - m_0), \Sigma),$$

and

$$D_t = (\Phi_{t,0})_\# \rho(2) = \mathcal{N}(m_2 + t(m_1 - m_0), \Sigma).$$

Their weighted $(1 - s, s)$-barycenter is the Gaussian with covariance Σ and mean

$$(1 - s)\big(m_0 + t(m_1 - m_0)\big) + s\big(m_2 + t(m_1 - m_0)\big) = m_0 + t(m_1 - m_0) + s(m_2 - m_0),$$

again equal to the mean of $\mathcal{Q}_{t,s}$. Thus $\mathcal{Q}_{t,s}$ is also the vertical barycenter. This proves the theorem. □

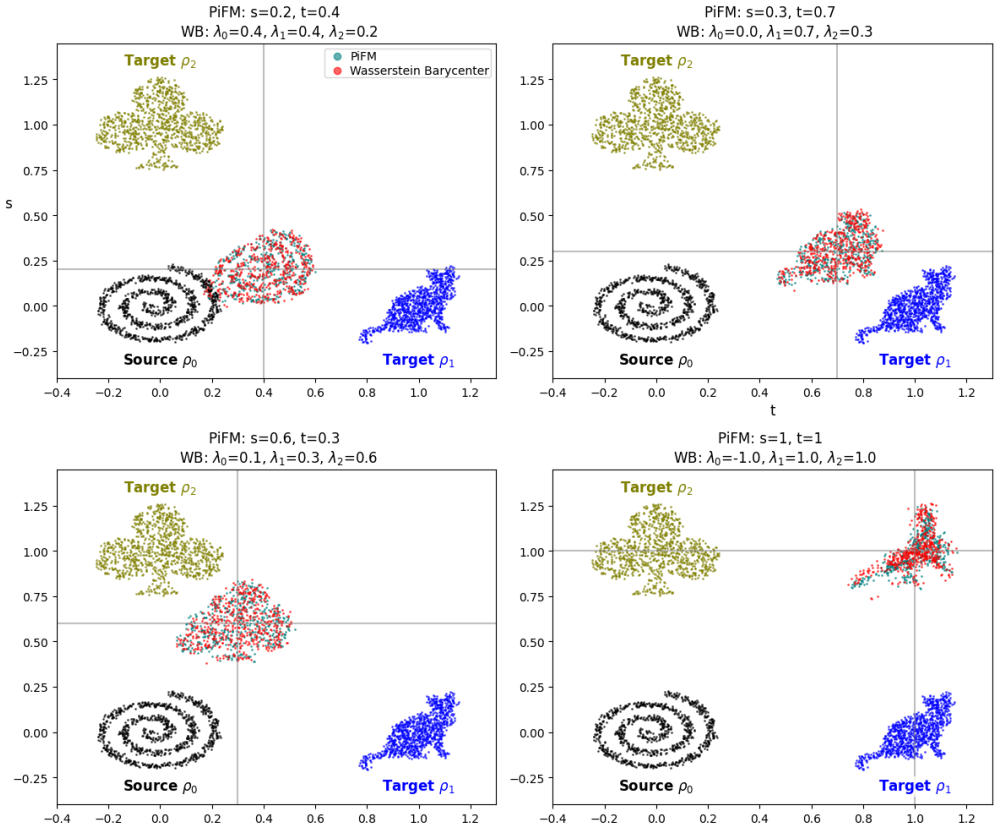


Figure 7: *Comparison between output distributions from PiFM model (ours) and Wasserstein Barycenters for a) within and b) outside the probability simplex for spiral, clover and cat distributions . The predicted distributions from the PiFM model correspond to the Wasserstein Barycenters.*

B Additional Experimental Details and Results

B.1 Comparison of execution times for PiFM and Free Support Algorithm (PyOT)

it is worth noting that computing Wasserstein barycenters is very expensive, even with the fast free-support computation implementation in the POT package. We compare execution time between the former and show that our method considerably reduces the time required to compute the barycenter (Supplementary Figure 8) and state that PiFM can be leveraged for fast inference of Wasserstein barycenters, with most of the computational cost incurred during model training.



Figure 8: Comparison of execution times between free support barycenter (POT package) and PiFM (ours).

B.2 Supplementary Details and Results for Unsupervised Image-to-Image Translation

B.2.1 Dataset, Preprocessing and Model architecture for Image-to-Image Translation

We train and evaluate on the CelebA dataset (Liu et al., 2015). From the official split we extract images and attribute labels, and use two attributes in the joint experiment: `Smiling` (attribute A) and `Black_Hair` (attribute B). Each image is center-cropped to 178, resized to 64×64, converted to a tensor and normalized to $[-1, 1]$ (per-channel mean 0.5, std 0.5). For training we partition the dataset into four subsets by attribute presence: *sources* (neither A nor B), *A-only* (has A, not B), *B-only* (has B, not A), and *both*.

For model architecture, We use a shared UNet backbone with two per-pixel MLP heads implemented as 1×1 convolutions (equivalent to applying a small MLP independently at every spatial location) with $SiLU$ non-linear activation function. Each of these two MLPs outputs the vector field corresponding to the flow or progression along each of the attributes mentioned above. Details of the model architecture used in this experiment are summarized in Table 1 in Appendix B.

Table 1: Summary of the model architecture used in the unsupervised image-to-image translation experiment (Section 4.2).

Subnetwork	Component	# PWE	Input / Output	Details
UNet backbone	Encoder (downpath): [128 → 256 → 256 → 256], 2 Res-Blocks per level	2*	3 / 128	Downsampling encoder with residual blocks per stage
	Bottleneck (mid): 256 channels with attention	2*	128 / 128	Residual blocks + attention at mid-resolution
	Decoder (uppath): [256 → 256 → 128 → 128]	2*	128 / 128	Symmetric upsampling with skip connections
Per-pixel MLP (Head_h)	Conv MLP: 128 → 128 → 3 (Conv1x1 + SiLU + Conv1x1)	2	128 / 3	Predicts per-pixel field $v(t, x)$
Per-pixel MLP (Head_v)	Conv MLP: 128 → 128 → 3 (Conv1x1 + SiLU + Conv1x1)	2	128 / 3	Predicts per-pixel field $v(t, x)$
Prediction semantics	–	–	(B,3,H,W)	Output is per-pixel displacement field $v(t, x)$

* "# PWE" indicates number of residual blocks per level.

B.2.2 Meta Flow Matching (MFM) as a baseline for unsupervised image-to-image translation

The original meta flow matching (MFM) (Atanackovic et al., 2025) work was developed and evaluated on low-dimensional point-cloud tasks (letters, trellis) where each training environment is a compact 2D point distribution and the model architecture is a GNN + flow decoder tailored to that setting. To test whether the MFM method, the alternating meta-style optimization that jointly learns per-environment embeddings and a conditional flow decoder, can scale to high-dimensional image data, we adapted the MFM pipeline to the CelebA (Liu et al., 2015) image dataset. Our goal was to keep the MFM training dynamics intact while replacing dataset- and architecture-specific pieces so that the same meta-learning principle can operate on images and produce composed transformations (e.g., smiling + aging).

Point-cloud environments and images are different in data dimensionality, spatial structure, and inductive biases. GNNs operate on unordered point sets and they are a suitable for letters, but they are not appropriate for dense image signals. To respect image structure and give the flow decoder expressive local features and to have a fair comparison with what we proposed in our PiFM pipeline, we replaced the original encoder/decoder backbone with a the same image backbone we used in our PiFM pipeline (architecture details are presented in Table 1: a U-Net-style feature extractor paired with small per-task MLP heads. This preserves the essential MFM decomposition: an encoder that produces a per-environment representation and a decoder that predicts flow fields conditioned on that representation, while providing the right inductive bias for images. Moreover, similar to MFM, we used two-optimizer alternating training: a flow optimizer that updates the decoder MLP heads to match conditional flow targets and an encoder optimizer that updates the backbone embeddings (the meta parameters). Alternating these updates encourages the encoder to produce embeddings that

generalize across environments while the decoder learns to follow them. All other core MFM choices (the flow-matching objective, sampling of interpolation times, EMA for evaluation snapshots, and the alternating schedule itself) were left intact so that any differences in outcome can be attributed to the change in data domain and architecture rather than to changes in the meta-method. At training, a practical issue when moving to images is how to pair source images (neither attribute) with target images (A-only or B-only) for supervised flow targets. We used nearest-neighbor (OT-style pairing) on flattened image features (or simple L2 in pixel space for diagnostics) to create target displacements. At inference we generate composed A+B outputs by integrating the sum of the two learned vector fields (or by sequential integrations $t \to$ or $s \to$), which mirrors the composition strategy used in our PiFM experiments.

We kept the same evaluation philosophy: visualize sample trajectories (frames along integration paths) and compute distribution distances between predicted distributions and target distributions using the same metrics used for point-cloud experiments. In Figure 9, we provide grids of intermediate frames for the three integration strategies (t-then-s, s-then-t, diagonal/combined) so readers can see whether composition produces plausible joint-attribute images. As shown, MFM fails to modify the source image along either one or both flow axes.

Training and sampling. We train two conditional vector fields simultaneously: (i) $v_h(t, x)$, vector field for attribute A (Smiling), and (ii) $v_v(s, x)$, vector field for attribute B (Black_Hair). On each training step we sample a minibatch of source images $\{x_a\}$ (from *sources*), target-A images $\{x_b\}$ (from *A-only*) and target-B images $\{x_c\}$ (from *B-only*). Each source is paired to a semantically similar target in the corresponding target minibatch using mini-batch discrete optimal transport (EMD) on flattened pixel vectors; denote the aligned targets by x_b^{aligned} and x_c^{aligned}. We optimize the network with Adam (base learning rate of 2×10^{-4}), gradient clipping of 1.0 and a linear warm-up schedule. An exponential moving average (EMA, decay 0.999) of the model parameters is maintained and used for visualization similar to (Tong et al., 2023).

Inference. Let N_{steps} be the integer number of update steps chosen for the inference (we use Euler updates of step size $1/N_{\text{steps}}$). The three flow integration strategies implemented are: (i) *Integrate t then s ($t \to s$)*: First integrate the A-field from $t = 0$ to 1, then integrate the B-field from $s = 0$ to 1. We split the total number of steps into two parts (approximately half for the first phase, half for the second) so the total frame count equals $N_{\text{steps}} + 1$. At each sub-step we update:$x \leftarrow x + \frac{1}{N_{\text{phase}}} v_h(t, x)$ or $x \leftarrow x + \frac{1}{N_{\text{phase}}} v_v(s, x)$, depending on the phase. (ii) *Integrate s then t ($s \to t$)*: Same as above but apply the B-field first and the A-field second. Order-of-application is intentionally swapped to reveal commutativity. (ii) *Integrate diagonally ($t = s$)*: Integrate along the diagonal $t = s = \tau$ from 0 to 1. At each step evaluate both fields and take their sum: $v(\tau, x) = v_h(\tau, x) + v_v(\tau, x), \qquad x \leftarrow x + \frac{1}{N_{\text{steps}}} v(\tau, x)$. This approximates simultaneous application of both vector fields.

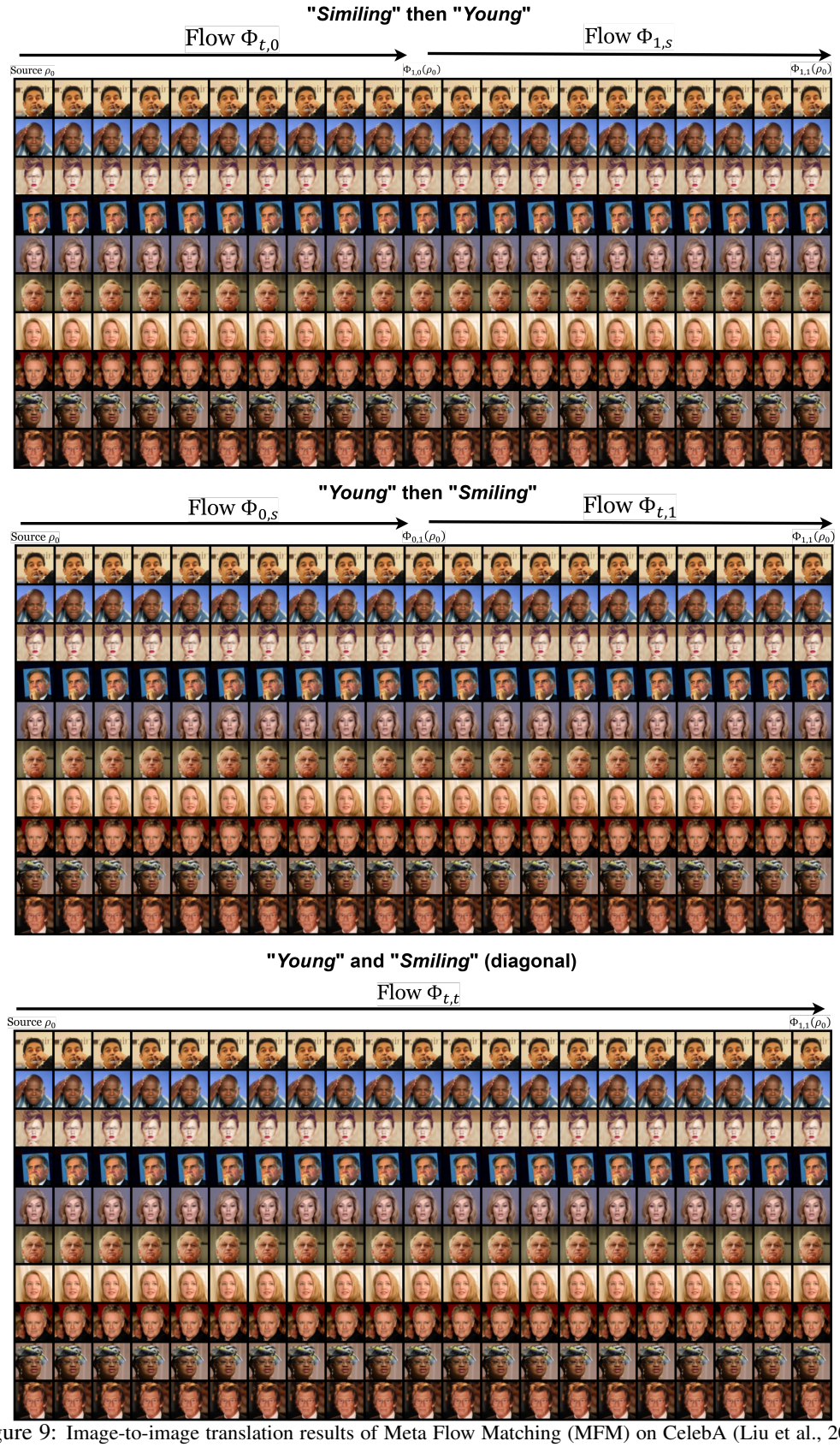


Figure 9: Image-to-image translation results of Meta Flow Matching (MFM) on CelebA (Liu et al., 2015) dataset across three different flow integration strategies. The A-field (`Smiling`) produces localized mouth/cheek edits consistent with smiling, while the B-field (`Young`) darkens and re-textures the hair region. As shown, MFM fails to modify the source image along either one or both flow axes.



Figure 10: More results on the Image-to-image translation task for PiFM on CelebA (Liu et al., 2015) dataset across three different flow integration strategies. The A-field (`Smiling`) produces localized mouth/cheek edits consistent with smiling, while the B-field (`Black Hair`) darkens and re-textures the hair region.

B.2.3 FID Evaluation of Commutativity

Standard image fidelity metrics such as FID are not well suited for evaluating PiFM on the commutativity task (Figure 5), as baseline methods often produce images that are close in pixel space without achieving the desired attribute changes. Figure 11 illustrates this limitation by applying FID to the experiment in Figure 5 in an attempt to quantify commutativity.

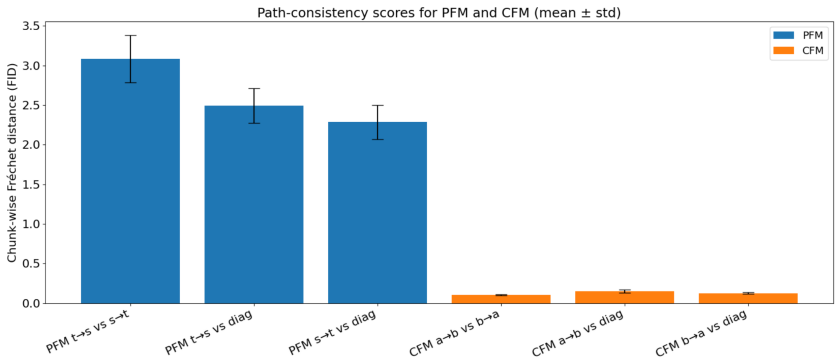


Figure 11: Quantitative experiment for commutativity of PiFM using FID, comparison with CFM. Based on the experiment in Section 4.2.

B.3 Single-Cell Experiment

B.3.1 Data Source and Preprocessing

We use the public MEF-to-iPSC reprogramming single-cell RNA-seq dataset (Schiebinger et al., 2019), using the serum-condition H5AD file from Figshare article 20051492. The raw object is already normalized and log-transformed. We keep cells from the serum condition, downsample to at most 1,500 cells per experimental day, and focus on the early reprogramming window from day 0 to day 6 at half-day resolution. This yields 19,492 cells after filtering.

For the state representation, we select up to 3,000 highly variable genes, scale expression values with maximum value 10, and compute a 50-dimensional PCA embedding. We use this PCA representation as the input space for PiFM; PHATE embeddings are computed only for visualization. The two biological axes are experimental day and the provided `Pluripotency` signature score. Experimental day represents chronological reprogramming, while the pluripotency score measures acquisition of an iPSC-like identity. For balanced grouping and diagnostics, we additionally split the pluripotency score within each day into low, middle, and high tertiles.

B.3.2 Evaluation of Prediction and Commutativity

We evaluate whether the two integrated composition orders produce consistent endpoint distributions, and whether these endpoints occupy the same region as the empirical target distribution. The source distribution is always the day 0, low-pluripotency group, and the comparison is performed over 39 endpoint groups, corresponding to 12 day intervals and 3 pluripotency-bin intervals. To make distances comparable across groups with different source-target scales, we report centroid distance and sliced Wasserstein distance normalized by the corresponding source-to-target distance.

As shown in Table 2, the two composition orders produce endpoint distributions that are close to each other relative to the overall source-target scale. Comparing each endpoint to the empirical target further shows that both compositions land in the same broad target regional, though they are not exactly identical. This imperfect agreement is expected because each endpoint is obtained by composing learned vector fields over relatively long day and pluripotency intervals, so small local integration and approximation errors can accumulate along the trajectory.

Table 2: Target distribution consistency measured by normalized centroid distance and normalized sliced Wasserstein distance. The metrics evaluate whether the two composition orders produce similar endpoint distributions and whether each endpoint lies in the same target region as the empirical distribution.

Comparison	Normalized centroid	Normalized W_2
day then Pluripotency vs true	0.6979 ± 0.1263	0.7917 ± 0.1684
Pluripotency then day vs true	0.5651 ± 0.1255	0.6601 ± 0.1691
day then Pluripotency vs Pluripotency then day	0.3459 ± 0.0572	0.3465 ± 0.0560

B.4 Comparison of distributions obtained from PiFM (ours) and Wasserstein barycenters using Free Support Algorithm (PyOT) for different values t, s

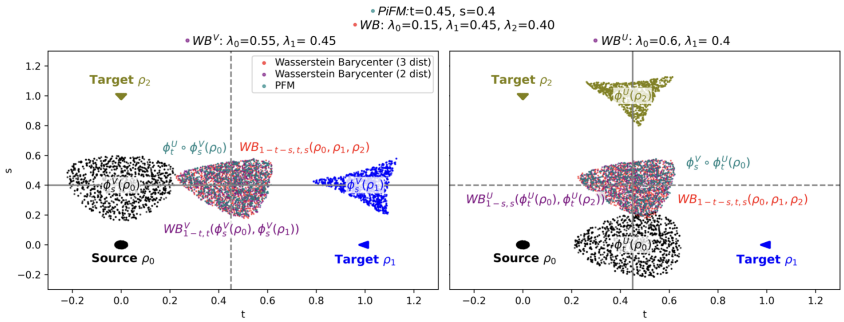
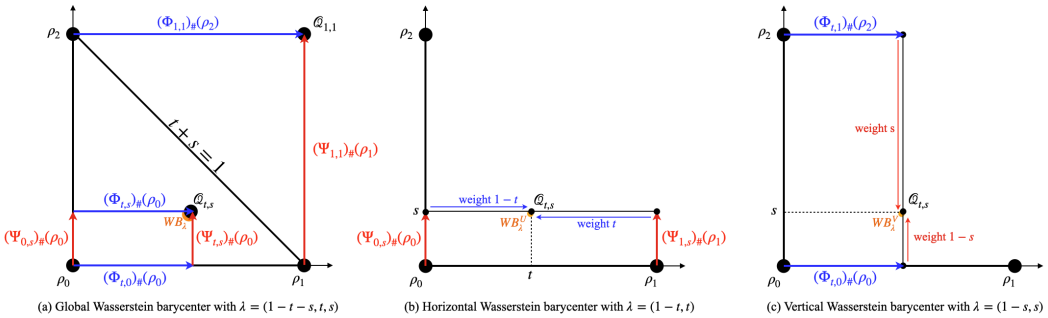


Figure 12: *Comparison between output distributions from PiFM (ours), global Wasserstein barycenter and Wasserstein barycenters between ϕ_s^V (left) and ϕ_t^U (right) projections for t=0.45, s=0.4. Source and target distributions are rescaled for clarity.*

B.5 Comparison of distributions obtained from PiFM (ours), Wasserstein barycenters using Free Support Algorithm (PyOT) between source and targets and vertical and horizontal Wasserstein barycenters

In the following, we define $\phi_{t,s=0}(\rho)$ as the output distribution from the push forward operation $(\phi_t^U)_\#$ of ρ with fixed $s = 0$, i.e the horizontal integration of the vector field for $t' \in [0, t]$ and $s = 0$. Similarly, we define $\phi_{t=1,s}(\rho)$ as the output distribution from the push forward operation $(\phi_s^V)_\#$ of ρ with fixed $t = 1$, i.e the integration of the vector field for $s' \in [0, s]$ and $t = 1$.



(a) Global Wasserstein barycenter with $\lambda = (1-t-s, t, s)$ (b) Horizontal Wasserstein barycenter with $\lambda = (1-t, t)$ (c) Vertical Wasserstein barycenter with $\lambda = (1-s, s)$

Figure 13: a) the global (t, s)-diagram with the simplex $\Delta = \{(t, s) \in [0, 1] : t + s \leq 1\}$ and commuting paths. b) the horizontal slice for fixed s. c) : the vertical slice for fixed t.

Supplementary Figures 14 and 15 provide additional comparisons at different values of (t, s), showing that the PiFM output agrees with the global Wasserstein barycenter and with the corresponding horizontal and vertical slice barycenters.

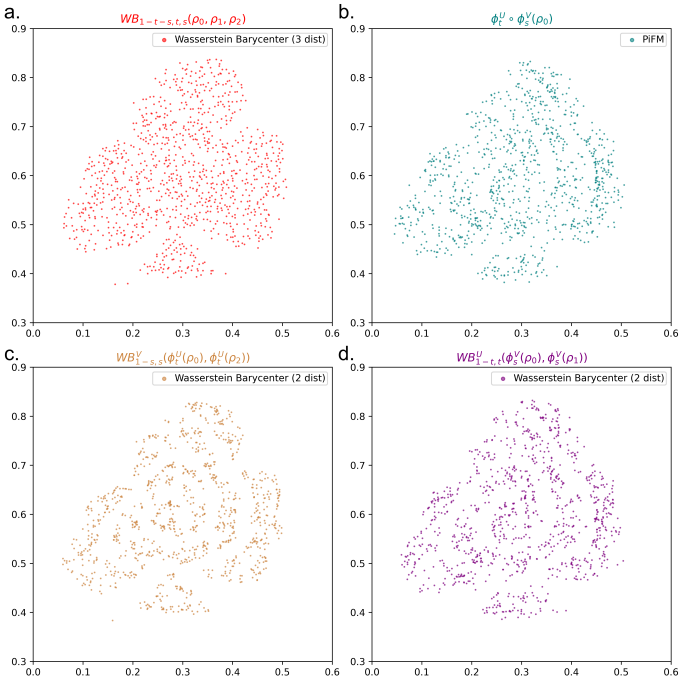


Figure 14: *Comparison between output distributions from a) Wasserstein Barycenters with weights* $(1-t-s, t, s) = (0.1, 0.6, 0.3)$ *(using Free Support Algorithm) b) PiFM model (ours) with* $t = 0.6, s = 0.3$*, c) vertical Wasserstein barycenter with* $\lambda = (1-t, t) = (0.4, 0.6)$ *computed using Free Support Algorithm of* $\phi_{t,s=0}(\rho_0)$ *and* $\phi_{t,s=1}(\rho_2)$ *with* $t = 0.6$*, d) horizontal Wasserstein barycenter* $\lambda = (1-s, s) = (0.7, 0.3)$ *computed using Free Support Algorithm of* $\phi_{t=0,s}(\rho_0)$ *and* $\phi_{t=1,s}(\rho_1)$ *with* $s = 0.3$

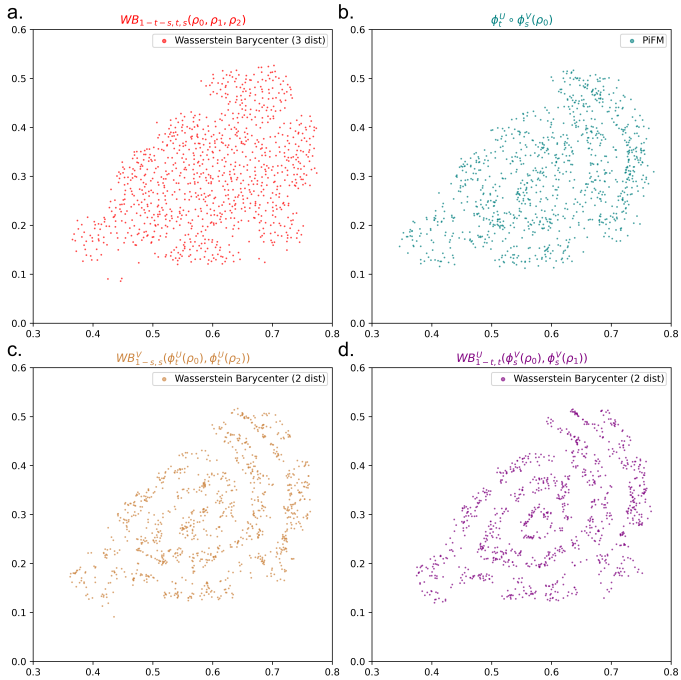


Figure 15: *Comparison between output distributions from a) Wasserstein Barycenters with weights* $(1-t-s, t, s) = (0.1, 0.3, 0.6)$ *(using Free Support Algorithm) b) PiFM model (ours) with* $t = 0.3, s = 0.6$*, c) vertical Wasserstein barycenter with* $\lambda = (1-t, t) = (0.7, 0.3)$ *computed using Free Support Algorithm of* $\phi_{t,s=0}(\rho_0)$ *and* $\phi_{t,s=1}(\rho_2)$ *with* $t = 0.3$*, d) horizontal Wasserstein barycenter* $\lambda = (1-s, s) = (0.4, 0.6)$ *computed using Free Support Algorithm of* $\phi_{t=0,s}(\rho_0)$ *and* $\phi_{t=1,s}(\rho_1)$ *with* $s = 0.6$

B.6 Further Discrepancy Measurements

In the following subsections, we provide additional empirical evidence supporting the results presented in the main manuscript, along with a small ablation study investigating whether the choice of coupling affects distributional path independence.

B.6.1 Figure 1

In the table below we provide a comparison of Wasserstein distance for CFM and PiFM in 3 integration strategies: t then s, s then t and diagonal. This experiment is done across 20 seeds.

Table 3: Comparison of the W_2 distance across different integration strategies.

Model	t **then** s	s **then** t	**Diagonal**
CFM	1.566972 ± 0.132423	0.686938 ± 0.083671	0.954987 ± 0.091452
PiFM	0.027063 ± 0.002321	0.019532 ± 0.001842	0.010326 ± 0.001174

B.6.2 Figure 4

For Figure 4, we run the experiment with varying training batch sizes, obtaining the following results.

Table 4: Comparison of the W_2 distance across different integration strategies.

Model	t **then** s	s **then** t	**Diagonal**
Curly-FM	0.2808 ± 0.0418	1.6132 ± 0.1079	0.7826 ± 0.0662
PiFM (no reg.)	0.2154 ± 0.0536	1.5241 ± 0.1943	0.9853 ± 0.1898
PiFM (with reg.)	0.1715 ± 0.0203	0.5236 ± 0.0468	0.2857 ± 0.0899

By varying the random seed while fixing the batch size at 256, we obtain the following results.

Table 5: Comparison of the W_2 distance across different integration strategies.

Model	t **then** s	s **then** t	**Diagonal**
Curly-FM	0.2297 ± 0.0180	1.5635 ± 0.1151	0.7285 ± 0.0611
PiFM (no reg.)	0.1996 ± 0.0206	0.9204 ± 0.1107	0.8559 ± 0.1446
PiFM (with reg.)	0.1649 ± 0.1272	0.2606 ± 0.0533	0.2791 ± 0.0683

Both experiments are carried out with $n = 10$ runs.

B.6.3 Experiment 4.3

Recovery of held-out day-by-pluripotency populations in the WOT reprogramming experiment. Values show mean $\pm$ standard deviation across held-out experiments. Distances are normalized by the source-to-target distance, with lower values indicating better recovery and one corresponding to no improvement over the source distribution.

Table 6: Comparison of normalized sliced W_2 and MMD^2 across different methods.

Model	**Normalized sliced** W_2	**Normalized MMD**^{2}
Local OT interpolation baseline	0.1020 ± 0.0242	0.0206 ± 0.0096
PiFM: day then pluripotency	0.3370 ± 0.0427	0.2121 ± 0.0672
PiFM: pluripotency then day	0.3108 ± 0.0416	0.1623 ± 0.0512

B.6.4 Ablation for coupling type

Using the Figure 1 toy example, we compare CFM and PiFM under OT and independent couplings over 10 seeds, measuring mean per-sample path discrepancy $\mathbb{E}[x_{t\to s} - x_{s\to t}]$. CFM remains path-dependent, while PiFM maintains low discrepancy for both couplings.

Table 7: Comparison of the W_2 distance for different coupling strategies.

Model	**OT coupling**	**Independent coupling**
CFM	1.886 ± 0.234	1.297 ± 0.520
PiFM	0.023 ± 0.007	0.031 ± 0.068

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claim regarding consistency and path-independent transport of distributions is supported both theoretically in Sections 3.1 and 3.2, and empirically in Sections 4.1 and 4.2. Additional supporting experiments are provided in Appendix B. The claim that PiFM enables practical training via a tractable, simulation-free objective that regresses onto multi-parameter conditional vector fields is addressed in Section 3.3. Finally, the connection between PiFM and the Wasserstein barycenter, as well as its interpretation as a generalization of this concept, is established theoretically in Section 3.4 and supported empirically in the same section, with further experiments presented in Appendix B.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitations we consider most important in the Limitations section. Beyond these, the authors do not identify additional limitations that warrant mention.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best

judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: We present the statements of the theoretical results in Sections 3.1, 3.2, and 3.4 in their two-dimensional form, along with the corresponding assumptions. These results are then extended to the n-dimensional setting in Appendix A, where full proofs for the general case are provided.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: All information required to reproduce the experimental results is summarized in Section 4 and described in further detail in Appendix B. The additional experiments presented in Appendix B likewise include all necessary implementation details for reproducibility.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.

(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: A link to a publicly accessible GitHub repository containing the code is provided in the abstract.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: All experimental details are provided in Section 4. Likewise, for the additional results, all relevant details are included in Appendix B.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

Justification: For experiments involving quantifiable metrics where the inclusion of error bars is appropriate, such as those in Figures 3 and 11, we report error bars to provide a measure of statistical significance. For the bar plot in Figure 4, we do not include error bars, as they are intended to quantify what is illustrated on the left side of the figure for that specific experiment.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We provide as much information on computational resources as we consider necessary. Due to the structure of PiFM training, in the general n-dimensional setting with n distinct parameters or target distributions, the training procedure scales as expected, being n times as computationally complex as standard Flow Matching.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have reviewed the NeurIPS Code of Ethics and, to the best of our knowledge, our work complies with it.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.

- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: As indicated in the paper, we do not identify any specific societal impacts that warrant discussion for this work. Generative modeling, in general, can have both positive and negative societal effects; however, we believe that our contribution does not introduce new capabilities that would require a dedicated assessment of such impacts.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The nature of the work presented in this paper does not have a need for safeguards.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: This work makes use of external software libraries, such as `POT` and `PyTorch`, which are properly cited in the paper.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: This paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: the paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.