Contact Geometry for Generative Models: An Unbalanced Optimal Transport Formulation

Andrea Testa[1] **Søren Hauberg**[2] **Andras Gabor Kupcsik**[1] **Tamim Asfour**[3] **Leonel Rozo**[4]

[1]Bosch Center for Artificial Intelligence (BCAI), [2]Technical University of Denmark (DTU),
[3]Karlsruhe Institute of Technology (KIT), [4]The Italian Institute of Artificial Intelligence (AI4I)

{Andrea.Testa,AndrasGabor.Kupcsik}@de.bosch.com sohau@dtu.dk
asfour@kit.edu leonel.rozo@ai4i.it

Abstract

Real-world generative problems in biology, medical imaging, or robotics, rarely come with perfectly paired source and target distributions. While unbalanced Optimal Transport (uOT) provides a robust framework to bridge unpaired datasets via minimum-length paths in probability space, existing formulations based on the Wasserstein–Fisher–Rao (WFR) geometry introduce a structural bias toward high density regions, often triggering mode collapse. We introduce *Contact unbalanced Optimal Transport (CuOT)*, a novel uOT formulation grounded in *contact geometry* that decouples density transport direction from mass growth to eliminate structural bias. This allows CuOT to adapt sampling steps to the local data structure, naturally reducing step sizes in high-density regions to prevent overshooting and ensuring stable convergence. The result is a stable, scalable path-length minimization solver that consistently outperforms WFR-based methods on biological dynamics reconstruction, unpaired image-to-image translation, and video generation. Project website.

1 Introduction

Generative modeling synthesizes new data by transforming probability distributions, either from noise or more generally between unpaired datasets, resulting in models that capture the temporal evolution of their underlying densities (Yang et al., 2022; Tong et al., 2024b; Neklyudov et al., 2023a; Ding et al., 2022). *The unpaired distributions problem* naturally arises in modeling real-world stochastic dynamics from sparse measurements, like in biology (Yeo et al., 2021; Moon et al., 2019; Zhang et al., 2024), climate science (Rasp and Thuerey, 2021; Verma et al., 2024), and economics (Kazakevičius et al., 2021; Alouadi et al., 2025). It also plays a central role in domain adaptation, tackling distribution shifts in applications such as synthetic-to-real image translation (Peng et al., 2017), medical imaging (Wang et al., 2024a), and systems control (Li et al., 2026a, 2025; Xu et al., 2025). Figure 1-*left* contrasts some leading approaches for connecting probability distributions. Diffusion processes (Ho et al., 2020; Karczewski et al., 2026) require an intermediate Gaussian noise state (Mokrov et al., 2021), whereas

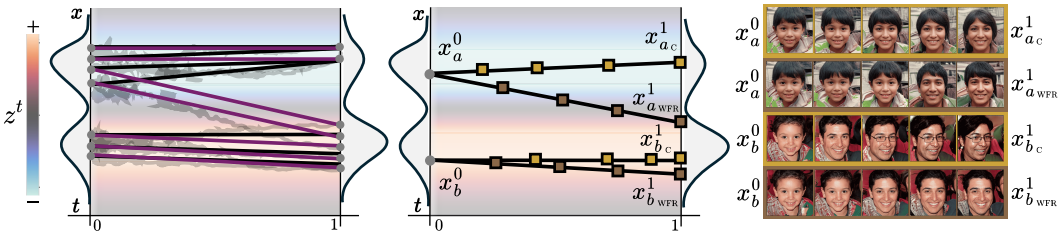

Figure 1: **Left**: Two marginals connected by a diffusion process (noisy gray path), an OT map (—), and a uOT map (—), the latter preventing mode leakage by adapting sample weights according to the growth rate z^t. **Middle-Right**: Comparison between uOT methods, namely WFR (brown squares) and CuOT (gold squares), in a dataset pairing experiment (App. F.3). Our CuOT method avoids bias toward high-growth regions and adapts the step size to the data structure.

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

Optimal Transport (OT) instead enforce direct, minimum-transport connections, yielding minimum-length paths under the Wasserstein metric (Vargas et al., 2021). While often easier to learn and integrate, OT approaches are brittle under class imbalance or outliers (Chizat et al., 2018; Gazdieva et al., 2024). Unbalanced OT (uOT) addresses this by allowing mass variations during transport, so that semantically related but differently sized modes can be matched without cross-mode leakage (Fatras et al., 2021). Yet, current uOT solutions (Sec. 2) based on the Wasserstein–Fisher–Rao (WFR) distance (Liero et al., 2016) introduce a new pathology: they bias the transport dynamics toward regions of maximal mass growth, resulting in soft mode collapse (Chapel et al., 2021; Zhang et al., 2025).

This paper overcomes the foregoing limitations by introducing a novel contact geometric formulation of uOT (Sec. 4), which decouples the direction of sample transport dynamics from its weight growth. The contact geometry (Sec. 3) naturally induces a bias whereby sampling step sizes decrease in regions of increasing weight (i.e., higher data density). Since these regions exhibit richer variability (Kumar et al., 2026), requiring finer resolution to avoid overshooting (Kharitenko et al., 2026; Li et al., 2026b; Weed and Bach, 2019), the WFR geometry falls short. In contrast, our *Contact Unbalanced Optimal Transport* method successfully adapts generative dynamics by producing smaller steps in high-density regions, as shown in Fig. 1-*middle-right*. Furthermore, our contact geometric interpretation of the uOT problem enables a stable and efficient solver via path-length minimization (Sec. 5). We validate our method on reconstructing stochastic dynamics first through an illustrative ablation study and then on biological datasets, unpaired image-to-image translation, and ensemble regression for video generation, consistently outperforming both OT- and WFR-based uOT approaches (Sec. 7).

In summary, **our contributions** are: (1) A contact-geometric formulation of unbalanced optimal transport, where we recast uOT through the lens of the geometry of non-conservative dynamics, resulting in a data-informed regularization that adapts sampling step sizes to the local data density and decouples it from mass growth; (2) A geometry-aware efficient solver, the Contact Unbalanced Optimal Transport (CuOT), which treats generative dynamics as minimum-length curves in probability space. CuOT delivers provably stable computation while maintaining an efficient profile: scaling linearly with dimensionality and achieving near-linear complexity in sample size.

2 Related Works

Table 1 surveys recent approaches for learning OT maps between stochastic distributions. Using OT solvers during training to measure the discrepancy between predicted and reference marginals has become standard in the field. Despite their super-linear complexity in the number of samples N (ranging from $\mathcal{O}(N \log N)$ to $\mathcal{O}(N^2)$), OT solvers achieve high reconstruction accuracy via explicit sample matching, compared to density-estimation methods based on direct parameterization or normalizing flows (Sha et al., 2024; Tong et al., 2024a; Neklyudov et al., 2023b). A key feature of

Method	Unbalanced	Geometry	Networks	Complexity
Schiebinger et al. (2019); Ventre et al. (2023); Lavenant et al. (2024)	✗	Wasserstein	$\varphi_\theta^t, \phi_\theta^t$	$\mathcal{O}(I_{\text{ref}} MDN^2)$
Chizat et al. (2022)	✗	Wasserstein	$\varphi_\theta^t, \phi_\theta^t$	$\mathcal{O}(KMDN^2)$
Pariset et al. (2023)	✗	Wasserstein	$\varphi_\theta^t, \phi_\theta^t$	$\mathcal{O}(I_{\text{IPF}} KDN)$
Tong et al. (2024a,b)	✗	Wasserstein	v_θ^t	$\mathcal{O}(I_{\text{ref}} MDN)$
Sha et al. (2024); Neklyudov et al. (2023a,b)	✓	WFR	$v_\theta^t, z_\theta^t, \rho_\theta^t$	$\mathcal{O}(KDN)$
Peng et al. (2024)	✓	WFR	v_θ^t, z_θ^t	$\mathcal{O}(MDN^2)$
Wang et al. (2025, 2026)	✓	WFR	v_θ^t, z_θ^t	$\mathcal{O}(I_{\text{ref}} MDN^2)$
Zhang et al. (2025)	✓	WFR*	$v_\theta^t, z_\theta^t, s_\theta^t$	$\mathcal{O}(MDN^2)$
Sun et al. (2025)	✓	WFR*	z_θ^t	$\mathcal{O}(MDN^2)$
Testa et al. (2026)	✗	Contact	v_θ^t	$\mathcal{O}(KDN \log N)$
CuOT (ours)	✓	Contact	v_θ^t, z_θ^t	$\mathcal{O}(MDN \log N)$

Table 1: Comparison of OT methods. **Bold** entries are evaluated in Sec. 7. **Unbalanced** indicates whether mass variation is explicitly modeled and learned. **Geometry** denotes the structure of the transport problem. WFR* indicates that the formulation recovers WFR geometry only for a specific choice of the penalty $\psi(z_\theta^t)$, and otherwise lacks a clear geometric interpretation. **Networks** lists the parameterized quantities: $(\varphi_\theta^t, \phi_\theta^t)$ denote the Schrödinger potentials, v_θ^t the drift, z_θ^t the growth rate, ρ_θ^t the density, and s_θ^t the score function. **Complexity** reports the training computational cost: N is the number of samples, D the data dimensionality, K the number of time steps, M the number of intermediate marginals, I_{IPF} the number of IPF iterations, and I_{ref} the number of reference refinement iterations.

the OT's geometric structure is the coupling between the transport drift v^t and the local growth rate z^t. Balanced OT approaches (Tong et al., 2024a) assume zero mass change, or hardcode the growth rate using task-specific knowledge (Chizat et al., 2022; Lavenant et al., 2024; Pariset et al., 2023; Ventre et al., 2023; Schiebinger et al., 2019), which fails to extend to broader cases.

More generally, in unbalanced settings, mass variation is controlled via a penalty $\psi(z^t)$, inducing the constraint $v^t = \nabla \partial_z \psi(z^t)$ (Zhang et al., 2025; Sun et al., 2025; Wang et al., 2025). This constraint is enforced either by parameterizing the growth rate alone, which can yield noisy gradient estimates, or through a regularization term in the loss. The quadratic penalty $\psi(z^t) = \frac{1}{2}\|z^t\|^2$ (Sha et al., 2024; Peng et al., 2024; Wang et al., 2026; Neklyudov et al., 2023a,b) yields a tighter coupling $v^t = \nabla z^t$, which corresponds to a WFR geometry that admits a distance-based regularization that is more stable than Hamilton–Jacobi–Bellman-based (HJB) formulations (Buzun et al., 2026). Instead, contact geometry relaxes the constraint between the direction of v^t and z^t. Building on Testa et al. (2026) (App. D.3), who used a fixed time schedule for z^t to modulate $\|v^t\|$ in balanced OT, we learn z^t as a spatially-varying growth rate field. This enables general uOT while adapting $\|v^t\|$ directly to the data.

3 Geometric Background

We introduce the necessary notation throughout the following sections, and summarize it in App. A.

Dynamics on the probability manifold. Let $\mathcal{P}^+(\mathcal{M})$ be the space of smooth, positive densities $\rho(x) : \mathcal{M} \to \mathbb{R}^+$ with total mass $\int_{\mathcal{M}} \rho(x)\, dx$, where $x \in \mathcal{M}$ denotes the sample space. A time-evolving density is described by a trajectory $\{\rho^t\}_{t\geq 0} \subset \mathcal{P}^+(\mathcal{M})$, whose infinitesimal variation is given by $\partial_t \rho^t(x) \in \mathcal{T}_{\rho^t}\mathcal{P}^+(\mathcal{M})$, with $\mathcal{T}_{\rho^t}\mathcal{P}^+(\mathcal{M})$ being the tangent space. The pair $(\rho^t, \partial_t \rho^t)$ thus defines a trajectory on the tangent bundle $\mathcal{T}\mathcal{P}^+(\mathcal{M})$, the collection of all tangent spaces. The evolution of ρ^t is typically modeled by the Fokker–Planck (FP) equation, combining deterministic density transport with its associated velocity field v^t and diffusion of strength ε,

$$\partial_t \rho^t(x) = -\nabla_x \cdot \big(\rho^t(x)\, v^t(x)\big) + \varepsilon \Delta_x \rho^t(x). \tag{1}$$

The classic OT seeks the trajectory $\rho^t \subset \mathcal{P}^+(\mathcal{M})$ connecting two densities $\rho_a, \rho_b \in \mathcal{P}^+(\mathcal{M})$ that minimizes the kinetic energy $\frac{1}{2}\|v^t(x)\|^2 \rho^t(x)$ (Benamou and Brenier, 2000). The diffusion term introduces an entropy regularization in the transport map, interpretable as a solution of the Schrödinger Bridge (SB) problem (Schrödinger, 1931, 1932), or as a *geodesic* in $\mathcal{P}^+(\mathcal{M})$ under the Wasserstein metric $g^{\mathcal{W}_2}$ (Ambrosio et al., 2005), which induces the corresponding distance $d^{\mathcal{W}_2}$. In particular,

$$g^{\mathcal{W}_2}(\partial_t \rho^t, \partial_t \rho^t) = \int_{\mathcal{M}} \sqrt{\partial_t \rho^t(x)\, (-\Delta_\rho)^\dagger \partial_t \rho^t(x)}\, dx, \tag{2}$$

where $(-\Delta_\rho)^\dagger$ is the inverse of the weighted Laplacian operator $\Delta_\rho = -\nabla_x \cdot (\rho^t \nabla_x)$ (Chow et al., 2020) (see App. B.4 for details). The density dynamics $(\rho^t, \partial_t \rho^t) \subset \mathcal{T}\mathcal{P}^+(\mathcal{M})$ can be equivalently expressed as (ρ^t, S^t) on the cotangent bundle $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$, the dual of $\mathcal{T}\mathcal{P}^+(\mathcal{M})$. Here, the covector $S^t(x) := (-\Delta_\rho)^\dagger \partial_t \rho^t(x)$ is defined to transform covariantly with $\partial_t \rho^t$, ensuring invariance of the Wasserstein metric (2). In the OT setting, S^t is the potential generating the flow via $v^t = \nabla_x S^t(x)$.

Symplectic and contact geometries. On the cotangent bundle $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$, the vector field $(\partial_t \rho^t, \partial_t S^t)$, which is the solution of the OT problem, respects the *symplectic* constraint $\omega\big((\partial_t \rho^t, \partial_t S^t), \cdot\big) = d\mathcal{H}(\partial_t \rho^t, \partial_t S^t)$. That is, the density dynamics evolve along level sets of the *Hamiltonian* function $\mathcal{H}$, corresponding to constant volume sections of $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$ (Chow et al., 2020). This *symplectic form* ω, a geometric structure measuring a generalized notion of area, is preserved along the dynamics in traditional OT settings. uOT relaxes this constraint, which can be geometrically achieved by augmenting the cotangent bundle with an additional scalar variable z^t, governing deviations from Hamiltonian level sets. The resulting dynamics evolve on the extended space $\mathcal{T}^*\mathcal{P}^+(\mathcal{M}) \times \mathbb{R}$ and are described by a *contact structure* satisfying $d\eta\big((\partial_t \rho^t, \partial_t S^t, \partial_t z^t), \cdot\big) - \mathcal{L}_{(\partial_t \rho^t, \partial_t S^t, \partial_t z^t)}\eta(\cdot) = d\mathcal{H}(\partial_t \rho^t, \partial_t S^t, \partial_t z^t)$. Here, η denotes the *contact form*, a generalized signed length quantifying the component of the vector field transversing Hamiltonian flow lines. $\mathcal{L}_{(\partial_t \rho^t, \partial_t S^t, \partial_t z^t)}\eta(\cdot)$ denotes the Lie derivative of the contact form along the vector field. In general, this term is nonzero, implying that the induced dynamics do not preserve the volume element, but instead evolve across sections of $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$ with varying volume (see App. B.1). While the symplectic dynamics correspond to a Wasserstein geodesic, the contact dynamics project to geodesics w.r.t. a dynamically varying metric (see App. B.2). As shown in Sec. 4, this

change in the local volume form across the contact dynamics induces density changes that correspond to mass variations on $\mathcal{P}^+(\mathcal{M})$, thereby moving from symplectic OT to the more general uOT setting.

4 Contact Unbalanced Optimal Transport

Problem definition. We propose an entropy-regularized uOT rooted in contact geometry. By treating the growth rate z^t as an independent dynamical variable, we decouple it from the spatial transport velocity v^t. In contrast, WFR formulations build on symplectic variables (ρ^t, S^t), constraining the growth rate to the action potential via $z^t := S^t$, and thus enforcing coupled dynamics where $v^t = \nabla_x z^t$. Let $\mathcal{M}$ be a domain and let $\{\rho^m\}_{m=0}^M$ be a collection of densities observed at times $\{t_m\}_{m=0}^M$. We seek a continuous trajectory $\{\rho^t\}_{t\in[t_0,t_M]}$, interpolating these marginals, together with a velocity field v^t and a scalar field z^t governing local mass variation. The problem is defined as,

$$\min_{v^t, z^{t_0}} \quad J = \int_{t_0}^{t_M}\int_{\mathcal{M}} \partial_t z^t(x)\, dx\, dt, \tag{3a}$$

$$\text{s.t.} \quad \partial_t z^t(x) = \tfrac{1}{2}\|v^t(x)\|^2\rho^t(x) - \tfrac{1}{2}\|v^{t_0}(x)\|^2\rho^{t_0}(x) + \tfrac{1}{2}w_z\, z^t(x)^2, \tag{3b}$$

$$\partial_t \rho^t(x) + \nabla_x \cdot \big(\rho^t(x)\, v^t(x)\big) = \varepsilon\Delta_x\rho^t(x) + w_z\, z^t(x)\, \rho^t(x), \tag{3c}$$

$$\rho^{t_m} = \rho^m, \quad \forall m \in \{0, \dots, M\}. \tag{3d}$$

The objective functional (3a) minimizes the terminal value of the action z^t, whose evolution is governed by the dynamics (3b). Equation (3c) is a generalized FP equation of the state ρ^t, extending the balanced FP (1) with the term $w_z z^t(x)\rho^t(x)$ to account for mass modulation. In the limit $w_z \to 0$, the problem reduces to minimizing the kinetic energy of the transport map, recovering the balanced OT formulation (details in Sec. 6). However, for $w_z \neq 0$, the dynamics introduce a nonlinear recursive dependence on z^t, such that it is neither a static action nor reducible to a standard state variable. In particular, the system cannot be cast into a symplectic form $\big((\rho^t, z^t), (S^t, \lambda^t)\big)$ with a conjugate variable λ^t acting as dual potential of z^t, and thus, falls outside the scope of classical Hamiltonian formulations on even-dimensional phase spaces. Instead, the system (3) is naturally described within the Herglotz variational framework (Herglotz, 1930), leading to a contact Hamiltonian system on an odd-dimensional contact manifold.

Contact Hamiltonian solution. As demonstrated in App. C.1, the optimal solution of problem (3) follows a contact Hamiltonian dynamics with Hamiltonian $\mathcal{H}$ defined as,

$$\mathcal{H}(\rho^t, S^t, z^t) = \mathcal{K}(\rho^t, S^t) + \varepsilon\mathcal{F}(\rho^t, S^t) - w_z \int_{\mathcal{M}} \tfrac{1}{2} z^t(x)^2 dx + \mathcal{H}_0. \tag{4}$$

At optimality, $v^t = \nabla_x S^t$, so the term $\mathcal{K} = \int_{\mathcal{M}} \frac{1}{2}\|\nabla_x S^t(x)\|^2\, \rho^t(x)\, dx$ represents the kinetic energy. The term $\mathcal{F} = \int_{\mathcal{M}} S^t(x)\, \Delta_x\rho^t(x)\, dx$ introduces an entropic coupling, scaled by the diffusion coefficient ε, which adjusts the transport map to preserve the entropy level imposed by the FP constraint (3c). Finally, the term $\frac{1}{2}z^t(x)^2$ governs energy variation, while the constant $\mathcal{H}_0 = \int_{\mathcal{M}} \frac{1}{2}\|v^{t_0}(x)\|^2\rho^{t_0}(x)dx$ fixes the reference energy level. The contact optimality conditions are,

$$\begin{aligned}
\partial_t\rho^t(x) &= \partial_S\mathcal{H}(\cdot) - \tfrac{1}{2}\rho^t(x)\partial_z\mathcal{H}(\cdot) \\
&= -\nabla_x\cdot\big(\rho^t(x)\nabla_x S^t(x)\big) + \varepsilon\Delta_x\rho^t(x) + w_z\, z^t(x)\rho^t(x), \quad \text{(via the FP eq. (3c))} \qquad (5a)\\
\partial_t S^t(x) &= -\partial_\rho\mathcal{H}(\cdot) - \tfrac{1}{2}S^t(x)\partial_z\mathcal{H}(\cdot) \\
&= -\tfrac{1}{2}\|\nabla_x S^t(x)\|^2 - \varepsilon\Delta_x S^t(x) + w_z\, z^t(x)S^t(x), \qquad (5b)\\
\partial_t z^t(x) &= \tfrac{1}{2}\big(S^t(x)\partial_S\mathcal{H}(\cdot) + \rho^t(x)\partial_\rho\mathcal{H}(\cdot)\big) - \mathcal{H}(\cdot) \\
&= \tfrac{1}{2}\|\nabla_x S^t(x)\|^2\rho^t(x) + \tfrac{1}{2}w_z\, z^t(x)^2 - \tfrac{1}{2}\|v^{t_0}\|^2\rho^t(x), \quad \text{(via the } z^t \text{ dynamics (3b))} \qquad (5c)
\end{aligned}$$

where the partial derivatives of the contact Hamiltonian $\mathcal{H}$ provide the local coordinate form of the contact geometric constraints (see Sec. 3). The explicit dependence on z^t encodes the trajectory traversed so far, distinguishing the contact Hamiltonian framework from the classical Hamiltonian setting. This enables modeling non-conservative, path-dependent forces, allowing the system's energy to evolve over time (Testa et al., 2026). In particular, this system induces a source or sink term in the generalized FP equation (5a), along with a corresponding amplification or damping term in the HJB equation (5b). More generally, the non-conservative dynamics implies that the volume of regions in $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$ explored by (ρ^t, S^t, z^t) is not preserved. As a result, both the transport velocity and the density are

affected: the former adapts to traverse varying-volume regions, while the latter varies according to the local expansion or contraction of the support. This geometric structure is fundamental, as it couples mass variation and transport while keeping them as distinct, independently evolving quantities.

Geodesic interpretation. The optimization (3) can be reformulated on the probability space $\mathcal{P}^+(\mathcal{M})$ by reparametrizing time t with an arc-length variable s, such that $ds = \sqrt{\mathcal{K}}\,dt$. Consequently, the problem reduces to finding a distance-minimizing curve $\rho^s \in \mathcal{P}^+(\mathcal{M})$ w.r.t. a suitable metric (Udrişte, 2000; Abraham and Marsden, 2008; Testa et al., 2025). In the limit $w_z, \varepsilon \to 0$, the cost functional (3a) is equivalent to the Wasserstein distance (see App. B.4), and the problem reduces to a standard Wasserstein geodesic computation. In our formulation, however, two aspects reshape this geometric picture. First, $\mathcal{F}$ introduces an entropy regularization component, so that the resulting metric increases for distributions ρ^s deviating from the reference entropy measure. Sinkhorn-based methods (Feydy et al., 2019) can efficiently incorporate this effect when approximating the distributional distance. Second, the explicit dependence on z^t drives energy variations along the trajectories in $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$, reflecting density volume changes in the underlying space. To account for this, we adjust the Wasserstein metric with a conformal scaling factor $\sqrt{\mathcal{K}}$, yielding the *contact Wasserstein metric* $g^{\mathcal{C}} = \sqrt{\mathcal{K}}\, g^{\mathcal{W}_2}$. Thus, the solution ρ^s is a piecewise entropy-regularized geodesic w.r.t. the dynamically varying metric $g^{\mathcal{C}}$. Computing the solution as a minimum-length path is advantageous, as it promotes optimization stability (see App. E.4), avoiding the need to enforce the HJB equation (5b) in the loss, whose direct minimization is prone to instabilities (Buzun et al., 2026).

5 CuOT: Solving the Contact uOT

Geodesic discretization. Our objective is to compute the geodesic induced by problem (3), i.e., a path $\rho^s \subset \mathcal{P}^+(\mathcal{M})$ consistent with the contact Hamiltonian dynamics and interpolating the observed marginals $\{\rho^m\}_{m=0}^M$. These constraints naturally lead to a discretization of ρ^s as a sequence of nodes $\{\rho^{s_k}\}_{k=0}^K \subset \mathcal{P}^+(\mathcal{M})$, each represented by a weighted particle ensemble $\rho^{s_k} = \{(x_n^{s_k}, w_n^{s_k})\}_{n=0}^N$ approximating the corresponding density. Therefore, each discrete marginal is obtained by integrating: *(1)* a parameterized transport network $v_\theta(x^{s_k}, s_k)$, representing the velocity field; and *(2)* a growth network $z_\theta(x^{s_k}, s_k)$, representing the local mass growth rate. Both networks take as input a batch of particles $\{x_n^{s_k}\}_{n=0}^N \sim \rho^{s_k}$ and the current arc-length parameter s_k. For a given step size τ, this generative process, representing the discrete unbalanced FP equation (3c), is defined as follows,

$$x_n^{s_{k+1}} = x_n^{s_k} + \tau\, v_\theta(x_n^{s_k}, s_k), \tag{6a}$$

$$w_n^{s_{k+1}} = w_n^{s_k} + \tau\, w_z z_\theta(x_n^{s_k}, s_k)\, w_n^{s_k}. \tag{6b}$$

Training procedure. Training these networks requires discretizing the rest of the contact uOT (3a)-(3b)-(3d). The objective minimization (3a)-(3b) is addressed by the *optimality* loss term ℓ_{opt}, which minimizes the total geodesic arc-length and penalizes deviations of the simulated trajectory (6) from the optimality conditions (5). Adherence to the marginal constraints (3d), in terms of total mass and reconstruction accuracy, is then measured by the *feasibility* loss ℓ_{fea}. Thus, our training objective is,

$$\ell = w_{\text{fea}}\,\ell_{\text{fea}} + w_{\text{opt}}\,\ell_{\text{opt}}, \quad \text{where,} \tag{7a}$$

$$\ell_{\text{fea}} = \sum_{m=1}^{M} \left(d^{\mathcal{W}_2}(w^{s_m}, x^{s_m}, w^m, x^m) + w_{\text{mass}}\|W^{s_m} - W^m\|^2\right), \tag{7b}$$

$$\ell_{\text{opt}} = \sum_{k=0}^{K-1}\sum_{n=0}^{N-1} \left((\nu_n^{s_k})^2 w_n^{s_k}\, \|x_n^{s_{k+1}} - x_n^{s_k}\|^2 + \|z_n^{s_{k+1}} - z_\theta(x_n^{s_k}, s_{k+1})\|^2\right). \tag{7c}$$

Minimizing the geodesic arc-length requires computing the contact Wasserstein metric $g^{\mathcal{C}} = \sqrt{\mathcal{K}}\, g^{\mathcal{W}_2}$, obtained by scaling the Wasserstein metric by the kinetic energy (Sec. 4). To this end, we integrate the optimality conditions (5) forward, discretized along the estimated path (see App. E for details),

$$\nu_n^{s_{k+1}} = \nu_n^{s_k} + \frac{1}{1+\nu_n^{s_k}}\left(\frac{\tau}{2} w_z z_\theta(x_n^{s_k}, s_k)\,\nu_n^{s_k} + w_z\left((z_\theta(x_n^{s_{k+1}}, s_k) - z_\theta(x_n^{s_k}, s_k))\, S_n^{s_k}\right)\right., \tag{8a}$$

$$S_n^{s_{k+1}} = S_n^{s_k} - \frac{\tau}{2}(\nu_n^{s_k})^2 + \frac{\tau}{2} w_z\, z_\theta(x_n^{s_k}, s_k)\, S_n^{s_k}, \tag{8b}$$

$$z_n^{s_{k+1}} = z_n^{s_k} + \frac{\tau}{2}(\nu_n^{s_k})^2 w_n^{s_k} + \frac{\tau}{2} w_z\, z_\theta(x_n^{s_k}, s_k)^2 - \frac{\tau}{2}(\nu_n^{s_0})^2 w_n^{s_0}. \tag{8c}$$

Here, $\nu_n^{s_k}$ denotes the scalar speed of the particle $x_n^{s_k}$, evolving jointly with the dual potential $S_n^{s_k}$ under the learned growth rate z_θ. The per-sample kinetic energy $(\nu_n^{s_k})^2 w_n^{s_k}$ acts as a conformal

scaling factor for particle displacements, weighting the distance $\|x_n^{s_{k+1}} - x_n^{s_k}\|^2$ in the optimality loss (7c), to provide a discrete realization of the contact metric $g^{\mathcal{C}}$. High metric values, occurring in regions of positive growth z_θ, penalize large displacements, forcing v_θ to predict smaller steps. As empirically shown in Sec. 7, this mechanism effectively refines the sampling resolution in high-density regions (see App. E.6). In the optimality loss (7c), we also regularize the inter-marginal evolution of the learned growth rate z_θ via the optimal action dynamics (8c). Reconstruction accuracy is enforced by requiring that the learned weighted ensemble $\rho^{s_k} = \{(x_n^{s_k}, w_n^{s_k})\}_{n=1}^N$ remains close to the reference marginal $\rho^m = \{(x_n^m, w_n^m)\}_{n=1}^{N_m}$ at each observation step $s_k = m$. We achieve this by defining the feasibility loss (7b) as a function of the total mass $W^{s_k} = \int_{\mathcal{M}} \rho^{s_k}(x)dx \approx \sum_{n=0}^N w_n^{s_k}$, and the distributional accuracy quantified by the weighted Wasserstein distance $d^{\mathcal{W}_2}$. The weighted distance is necessary as the number N_m of available samples from the marginal ρ^m may differ from the batch size N used for the sample discretization of the geodesic. The training phase is summarized in Alg. 1, while implementation details and a computational complexity analysis are given in App. E.

6 Contact Geometry Generalizes Optimal Transport Formulations

Symplectic geometry of OT. Since symplectic dynamics is a special case of contact dynamics in which the action z^t does not influence the evolution of (ρ^t, S^t), the entropy-regularized OT problem is recovered as a limiting case of the contact formulation (3) when $w_z \to 0$, as,

$$\min_{v^t} \quad J(v^t, \rho^t) = \int_{t_0}^{t_M} \int_{\mathcal{M}} \frac{1}{2} \|v^t(x)\|^2 \, \rho^t(x) \, dx \, dt, \tag{9a}$$

$$\text{s.t.} \quad \partial_t \rho^t(x) = -\nabla_x \cdot \big(\rho^t(x) \, v^t(x)\big) + \varepsilon \Delta_x \rho^t(x), \tag{9b}$$

$$\rho^{t_m} = \rho^m, \quad \forall m \in \{0, \dots, M\}. \tag{9c}$$

The Hamiltonian of the optimization problem (9) is $\mathcal{H}(\rho^t, S^t) = \mathcal{K}(\rho^t, S^t) + \varepsilon \mathcal{F}(S^t, \rho^t)$, and consists solely of the kinetic energy and an entropy regularization term (Chen et al., 2014) (see App. D.1). The optimality conditions define a Hamiltonian system evolving on the symplectic manifold $(\mathcal{T}^*\mathcal{P}^+(\mathcal{M}), \omega)$ (Cui et al., 2024; Chow et al., 2020; Neklyudov et al., 2023b). As a consequence, the flow preserves volume within $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$, and the density $\rho^t \in \mathcal{P}^+(\mathcal{M})$ conserves its total mass $\int_{\mathcal{M}} \rho^t(x) \, dx$ throughout the evolution, reflecting the balanced nature of classical OT.

Wasserstein-Fisher-Rao geometry of uOT. To connect distributions with different mass, several related works (Sec. 2) propose WFR-based extensions of the OT problem (9), such as,

$$\min_{v^t, z^t} \quad J(v^t, z^t, \rho^t) = \int_{t_0}^{t_M} \int_{\mathcal{M}} \frac{1}{2} \big(\|v^t(x)\|^2 + w_z \psi(z^t(x))\big) \, \rho^t(x) \, dx \, dt, \tag{10a}$$

$$\text{s.t.} \quad \partial_t \rho^t(x) = -\nabla_x \cdot \big(\rho^t(x) \, v^t(x)\big) + \varepsilon \Delta_x \rho^t(x) + z^t(x) \rho^t(x), \tag{10b}$$

$$\rho^{t_m} = \rho^m, \quad \forall m \in \{0, \dots, M\}. \tag{10c}$$

The dynamics constraint (10b) extends the FP equation (9b) by introducing a local mass growth rate z^t (Lu et al., 2019), yielding a formulation equivalent to (3c). The growth rate z^t enters the cost functional (10a) through a penalty term $\psi(z^t)$, but unlike the contact Hamiltonian case, it does not correspond to the minimized action. The optimization problem (10) therefore retains the standard variational form and still yields a classical Hamiltonian structure.

When setting $\psi(z^t) = z^t(x)^2$, the Hamiltonian takes form $\mathcal{H}(\rho^t, S^t) = \mathcal{K}(\rho^t, S^t) + \varepsilon \mathcal{F}(S^t, \rho^t)$, where the kinetic energy is given by $\mathcal{K}(\rho^t, S^t) = \frac{1}{2} \int_{\mathcal{M}} \Big(\|\nabla_x S^t(x)\|^2 + \frac{1}{w_z} S^t(x)^2 \Big) \rho^t(x) dx$, and defines the WFR metric, which reinterprets the symplectic dynamics on $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$ as a tractable geodesic $\rho^s \subset \mathcal{P}^+(\mathcal{M})$ (Kondratyev et al., 2016; Neklyudov et al., 2023b). This yields a conic space that converges to the Wasserstein manifold as $w_z \to \infty$, where mass variation is suppressed and the WFR-based uOT reduces to the classic OT (9). The optimality condition $v^t(x) = \nabla_x z^t(x)$ (see App. D.2 for details) biases the transport dynamics toward regions of maximal mass growth. This induces a structural bias that can lead to mode collapse by discouraging the reconstruction of low-mass target modes (Chapel et al., 2021; Zhang et al., 2025). The same mechanism also introduces instability, as the transport velocity becomes sensitive to gradients of the growth field, which are often noisy in real datasets. This limitation is inherent to the volume-preserving symplectic dynamics that can only accommodate mass variation only through a rigid coupling between z^t and v^t. Notably, CuOT (3) overcomes these structural constraints by leveraging contact geometry to decouple transport from mass growth, providing a principled framework for more accurate and controlled adaptation to mass variations.

7 Results

We benchmark the proposed CuOT framework against established OT and uOT baselines summarized in Table 1. Additionally, to isolate the individual sources of performance gains, we conduct an ablation study comparing our full model against variants that systematically omit key components.

Ablation study We first evaluate CuOT on a synthetic 2D distribution-bridging task with known mode correspondences and mass dynamics. The source distribution comprises three Gaussian modes (400 samples each), and the target distribution features three distinct shapes: a Swiss roll, concentric rings, and a pinwheel (600, 600, and 200 samples, respectively). Matching the ground truth requires mass expansion ($z_\theta > 0$) on the first two modes and decay on the third, posing a challenge for OT frameworks. Furthermore, the deliberate misalignment between the transport direction and the growth gradient on the third mode highlights key failure modes in WFR. Table 2 summarizes ablation results and App. F.1 details all variants. The CuOT variant w/o Adaptive Step differs from the full model only in Equation (7c), where the contact metric $g^{\mathcal{C}}$ is replaced by the standard Wasserstein metric $g^{\mathcal{W}_2}$. Although this produces a virtually identical growth field (Figure 2a), removing the adaptive step reduces overall transport accuracy ($\mathcal{W}_2$: $0.10 \to 0.13$; $\mathcal{W}_1$: $0.08 \to 0.10$). The variant with growth only dependent on time $z_\theta(s_k)$ performs the worst because it cannot capture spatially varying dynamics (Figure 2d). WFR-based variants suffer from mode leakage as forward transport directions are incompatible with negative growth rates (Figure 2b). Vanishing growth regularization ($w_z \to 0$) eliminates leakage but results in a high Chamfer distance (> 0.3) and mode collapse, as the WFR geometry enforces positive growth across all modes. Finally, OT-based frameworks with constant sample weights cannot prevent misrouting, even when learning to scale global mass. This yields the highest mode leakage (17.7%) by assigning positive growth to samples in the low-mass cluster (Figure 2c).

Table 2: Quantitative ablation study on the 2D multi-mode synthetic transport problem. Mode coverage is evaluated using the Chamfer distance (CD) between the target and generated distributions.

Method	$\mathcal{W}_2$	$\mathcal{W}_1$	Mass	Leakage (%)	CD
CuOT (Complete)	$\mathbf{0.10}_{\pm 0.01}$	$\mathbf{0.08}_{\pm 0.01}$	$0.0006_{\pm 0.0001}$	$\mathbf{0.2}\%$	$0.09_{\pm 0.02}$
CuOT w/ Time-Only Growth	$0.27_{\pm 0.02}$	$0.19_{\pm 0.02}$	$0.1484_{\pm 0.0104}$	14.1%	$0.31_{\pm 0.03}$
CuOT w/ WFR Geometry	$0.16_{\pm 0.02}$	$0.13_{\pm 0.01}$	$0.0009_{\pm 0.0001}$	10.4%	$0.11_{\pm 0.02}$
CuOT w/o Adaptive Step	$0.13_{\pm 0.01}$	$0.10_{\pm 0.01}$	$0.0009_{\pm 0.0001}$	$\mathbf{0.2}\%$	$0.10_{\pm 0.01}$
VarRUOT	$0.16_{\pm 0.01}$	$0.13_{\pm 0.01}$	$0.0006_{\pm 0.0001}$	6.5%	$0.13_{\pm 0.02}$
OTCFM	$0.20_{\pm 0.01}$	$0.13_{\pm 0.01}$	$\mathbf{0.0001}_{\pm \mathbf{0.0001}}$	17.7%	$\mathbf{0.07}_{\pm \mathbf{0.01}}$

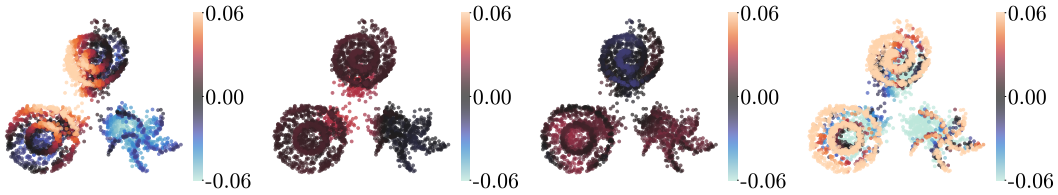

Figure 2: Learned growth field $z_\theta(x^{s_k}, s_k)$ across variants in the ablation study. a CuOT (Complete), b WFR Geometry, c Standard OT, and d w/ Time-Only Growth.

Single-cell reconstruction. We next benchmark the proposed CuOT framework on the trajectory inference from temporal snapshots of single-cell data, considering 5 well-established benchmarks (see Table 3). Analysis on these datasets reveals a fundamental mismatch: the Wasserstein distance $d^{\mathcal{W}_2}$ between successive marginals $\{\rho^m\}_{m=0}^M$ is not proportional to the growth rate gradient $\nabla_x z^t(x)$. This means that real-world biological processes often violate the assumption underlying WFR-based methods. Even in the simulated three-gene dataset (Zhang et al., 2025), the distance between successive distributions decreases monotonically, while $\nabla_x z^t(x)$ remains nearly constant (see Table 12 in App. F.2), exposing a structural decoupling that WFR cannot accommodate. Figure 3 shows the 2D projection of the gene dynamics, alongside the predicted growth rate. The ground-truth dynamics consist of two distinct modes: a stationary primary mode and a dynamically evolving secondary mode. Balanced OT methods fail to model this dynamics, suffering from spurious cross-mode leakage. While uOT approaches successfully isolate the evolving mode, WFR-based

Table 3: Summary of single-cell reconstruction experiments (details in App. F.2).

Dataset	Dim.	Biological Process	Details	Results
Three-gene simulation (Zhang et al., 2025)	2D	Simulated gene expression dynamics	Table 12	Fig. 3, Table 4
Mouse Hematopoiesis (Weinreb et al., 2020)	2D	Blood stem cell specialization	Table 13	Fig. 8, Table 14
EMT (Cook and Vanderhyden, 2020)	10D	Cancer cell transition	Table 15	Fig. 9, Table 16
Pancreatic β-cell (Veres et al., 2019)	30D	Stem cell maturation	Table 17	Fig. 10, Tables 18, 19
Zebrafish Embryogenesis (Farrell et al., 2018)	50D	Cell specialization dynamics	Table 20	Fig. 4, Tables 5, 21

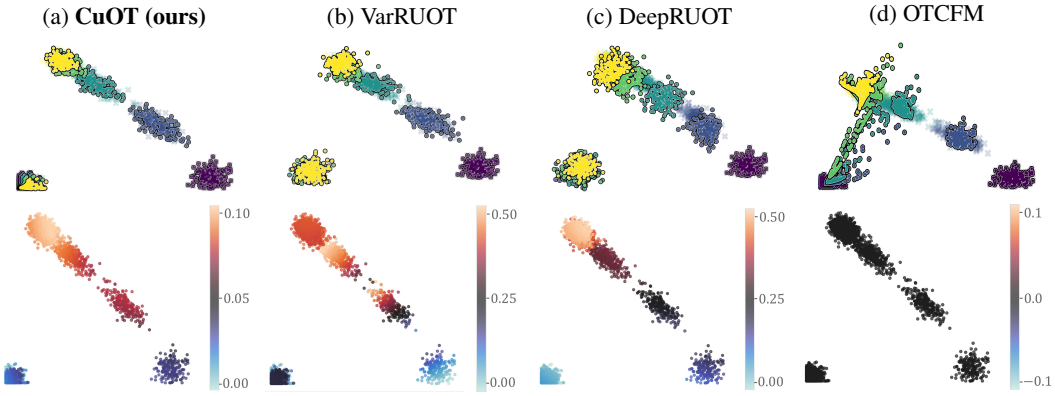

Figure 3: **Top:** Dynamics $(t_0, t_1, t_2, t_3, t_4)$. **Bottom:** Growth rate (bottom) for Zhang et al. (2025).

Table 4: Quantitative analysis on the three-gene simulated dataset (Zhang et al., 2025).

Model	Metric	$t=1$	$t=2$	$t=3$	$t=4$
OTCFM (Tong et al., 2024a)	$\mathcal{W}_1$	$0.0889_{\pm 0.0019}$	$0.1149_{\pm 0.0018}$	$0.2086_{\pm 0.0019}$	$0.4054_{\pm 0.0019}$
	$\mathcal{W}_2$	$0.1430_{\pm 0.0028}$	$0.1257_{\pm 0.0021}$	$0.2793_{\pm 0.0027}$	$0.5505_{\pm 0.0022}$
	Mass	$0.0110_{\pm 0.0000}$	$0.1056_{\pm 0.0000}$	$0.5256_{\pm 0.0000}$	$2.0235_{\pm 0.0000}$
TIGON (Sha et al., 2024)	$\mathcal{W}_1$	$0.0636_{\pm 0.0012}$	$0.0998_{\pm 0.0011}$	$0.1601_{\pm 0.0022}$	$0.1885_{\pm 0.0023}$
	$\mathcal{W}_2$	$0.1040_{\pm 0.0008}$	$0.1925_{\pm 0.0025}$	$0.2534_{\pm 0.0030}$	$0.2871_{\pm 0.0027}$
	Mass	$0.0263_{\pm 0.0009}$	$0.0490_{\pm 0.0009}$	$0.0527_{\pm 0.0009}$	$0.0531_{\pm 0.0010}$
DeepRUOT (Zhang et al., 2025)	$\mathcal{W}_1$	$0.0448_{\pm 0.0063}$	$0.0832_{\pm 0.0074}$	$0.1229_{\pm 0.0074}$	$0.1450_{\pm 0.0074}$
	$\mathcal{W}_2$	$0.0525_{\pm 0.0061}$	$0.1391_{\pm 0.0096}$	$0.2325_{\pm 0.0103}$	$0.2429_{\pm 0.0105}$
	Mass	$0.0005_{\pm 0.0001}$	$0.0012_{\pm 0.0001}$	$0.0004_{\pm 0.0001}$	$0.0002_{\pm 0.0001}$
VarRUOT (Sun et al., 2025)	$\mathcal{W}_1$	$0.0476_{\pm 0.0039}$	$0.0518_{\pm 0.0049}$	$0.0524_{\pm 0.0055}$	$0.0569_{\pm 0.0053}$
	$\mathcal{W}_2$	$0.0431_{\pm 0.0101}$	$0.0865_{\pm 0.0117}$	$0.0884_{\pm 0.0129}$	$0.1009_{\pm 0.0104}$
	Mass	$0.0002_{\pm 0.0001}$	$0.0005_{\pm 0.0001}$	$0.0065_{\pm 0.0006}$	$0.0032_{\pm 0.0008}$
CuOT (Ours)	$\mathcal{W}_1$	$\mathbf{0.0337_{\pm 0.0018}}$	$\mathbf{0.0466_{\pm 0.0064}}$	$\mathbf{0.0381_{\pm 0.0044}}$	$\mathbf{0.0225_{\pm 0.0008}}$
	$\mathcal{W}_2$	$\mathbf{0.0296_{\pm 0.0020}}$	$\mathbf{0.0589_{\pm 0.0115}}$	$\mathbf{0.0624_{\pm 0.0092}}$	$\mathbf{0.0355_{\pm 0.0054}}$
	Mass	$\mathbf{9 \times 10^{-8}_{\pm 6\times 10^{-8}}}$	$\mathbf{2 \times 10^{-6}_{\pm 2\times 10^{-7}}}$	$\mathbf{1 \times 10^{-5}_{\pm 3\times 10^{-6}}}$	$\mathbf{5 \times 10^{-5}_{\pm 1\times 10^{-5}}}$

Table 5: Quantitative analysis on the dataset Farrell et al. (2018). Results averaged over $t = 1, \ldots, 6$.

Method	$\mathcal{W}_1$	$\mathcal{W}_2$	Mass
OTCFM (Tong et al., 2024a)	$6.8356_{\pm 2.2546}$	$5.0909_{\pm 1.2883}$	$0.8359_{\pm 0.4004}$
TIGON (Sha et al., 2024)	$8.3305_{\pm 1.5746}$	$7.2321_{\pm 2.0083}$	$0.5274_{\pm 0.1145}$
DeepRUOT (Zhang et al., 2025)	$4.2386_{\pm 1.2133}$	$4.9448_{\pm 1.0268}$	$0.1823_{\pm 0.1875}$
VarRUOT (Sun et al., 2025)	$5.2476_{\pm 0.9197}$	$4.8008_{\pm 0.7942}$	$0.4192_{\pm 0.2929}$
CuOT (Ours)	$\mathbf{3.2689_{\pm 0.4946}}$	$\mathbf{3.0357_{\pm 0.4140}}$	$\mathbf{0.0243_{\pm 0.0205}}$

solutions are hindered by the structural biases that degrade their reconstruction accuracy (see Table 4). In contrast, our CuOT approach achieves superior performance across all metrics. On real-world datasets, the gradient $\nabla_x z^t$ is highly variable and noisy, rendering the WFR-based estimation of the transport term v^t unstable. This is evident in the zebrafish embryogenesis dynamics (Farrell et al., 2018), where a single progenitor population branches into multiple lineage modes. As shown in Fig. 4, balanced OT methods achieve broad modality coverage, but fail to adapt sample weights, hindering precise matching. Conversely, although most uOT methods achieve higher scores in Table 5 through weight adjustments, their coverage of the different modalities remains unsatisfactory. By avoiding these structural biases, our CuOT approach consistently outperforms the state-of-the-art baselines, successfully capturing both the diversity and the precision of the biological dynamics.

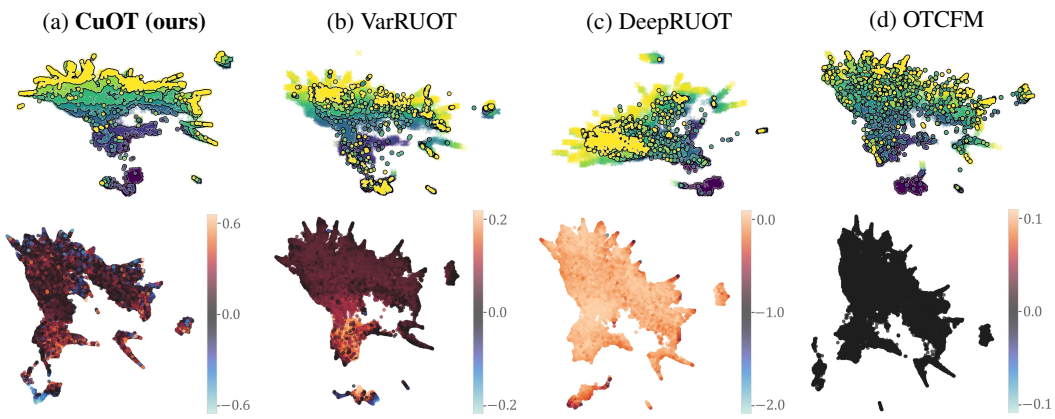

Figure 4: Qualitative analysis on the zebrafish embryogenesis dataset (Farrell et al., 2018). **Top:** Predicted dynamics $(t_0, t_1, t_2, t_3, t_4, t_5, t_6)$. **Bottom:** Growth rate.

Unpaired dataset matching. We then test CuOT on unpaired image-matching tasks using the Flickr-Faces-HQ (FFHQ) human face dataset (Karras et al., 2019). We employ pre-trained encoders to compute the stochastic dynamics in two distinct latent representations: a 512-dimensional space provided by ALAE (Pidhorskyi et al., 2020) and a 9216-dimensional space provided by e4e (Richardson et al., 2021). We partition the dataset into two age-based groups: *adults* ($45K$ samples) and *children* ($15K$ samples). The Adult $\rightarrow$ Child task requires accurate early-stage dynamics in a highly-dense source to ensure correct transport to the target mode. Conversely, the Child $\rightarrow$ Adult task demands high-fidelity reconstructions and a correct coverage of the target distribution support. Within the ALAE latent space (Table 6), balanced OT approaches yield larger distances from the target distribution than WFR-based uOT methods; however, they achieve the second-best FID scores on decoded images, which better measures mode coverage in the reconstruction. As illustrated in Figure 5, although higher latent distances correlate with lower visual fidelity, balanced OT methods provide broader distribution coverage, producing images that faithfully reflect the target age group. In contrast, WFR-based uOT baselines frequently fail to complete the transport trajectory effectively and suffer from severe scalability bottlenecks in the higher-dimensional e4e latent space (Table 7). By leveraging contact geometry over the WFR structure, CuOT avoids mode collapse, achieving state-of-the-art performance across all metrics and latent representations with superior sample quality and target coverage. Additional details are provided in App. F.3, where Figs. 12–15 show how CuOT allocates more sampling steps in high-density regions to improve accuracy.

Table 6: Adult↔Child image generation on the FFHQ dataset (Karras et al., 2019), ALAE encoder.

Method	FID	$\mathcal{W}_1$	$\mathcal{W}_2$	Mass
OTCFM (Tong et al., 2024a)	$5.5307_{\pm 0.0816}$	$14.4850_{\pm 0.1087}$	$10.3316_{\pm 0.1097}$	$6.5052_{\pm 0.0000}$
TIGON (Sha et al., 2024)	$21.8858_{\pm 0.4533}$	$16.4599_{\pm 0.1795}$	$11.5193_{\pm 0.1489}$	$0.7945_{\pm 0.0696}$
DeepRUOT (Zhang et al., 2025)	$6.5654_{\pm 0.1801}$	$13.7665_{\pm 0.1090}$	$9.8020_{\pm 0.1058}$	$0.0102_{\pm 0.0082}$
VarRUOT (Sun et al., 2025)	$27.2228_{\pm 0.2637}$	$14.7688_{\pm 0.1265}$	$10.5592_{\pm 0.1171}$	$0.8849_{\pm 0.3058}$
CuOT (Ours)	$\mathbf{5.1764_{\pm 0.0761}}$	$\mathbf{12.4042_{\pm 0.1080}}$	$\mathbf{8.8574_{\pm 0.1005}}$	$\mathbf{0.0043_{\pm 0.0018}}$

Table 7: Adult↔Child image generation on the FFHQ dataset (Karras et al., 2019), e2e encoder.

Method	FID	$\mathcal{W}_1$	$\mathcal{W}_2$	Mass
OTCFM (Tong et al., 2024a)	$9.3859_{\pm 0.1646}$	$19.3396_{\pm 0.3013}$	$14.8317_{\pm 0.1881}$	$6.5049_{\pm 0.0000}$
TIGON (Sha et al., 2024)	$45.2745_{\pm 0.3889}$	$28.8340_{\pm 0.3516}$	$20.5483_{\pm 0.3577}$	$1.5906_{\pm 0.1553}$
DeepRUOT (Zhang et al., 2025)	$39.7057_{\pm 0.1952}$	$27.7649_{\pm 0.2185}$	$20.7265_{\pm 0.2127}$	$0.0735_{\pm 0.0103}$
VarRUOT (Sun et al., 2025)	$54.6319_{\pm 0.0800}$	$31.3152_{\pm 0.2591}$	$24.5797_{\pm 0.3155}$	$1.1373_{\pm 0.4025}$
CuOT (Ours)	$\mathbf{8.3972_{\pm 0.0972}}$	$\mathbf{16.0868_{\pm 0.1131}}$	$\mathbf{12.1086_{\pm 0.1408}}$	$\mathbf{0.0153_{\pm 0.0144}}$

Image-to-Video generation. We further evaluate our method by performing video trajectory reconstruction on the UCF101 dataset (Soomro et al., 2012). Here, the learned transport map reconstructs temporal frame sequences across $M = 60$ intermediate marginal distributions, with each marginal corresponding to a discrete time step along the video dynamics. Because the frame count per video remains constant, mass variation is negligible. Therefore, this benchmark primarily evaluates scalability



Figure 5: Child → Adult generation on the FFHQ dataset (Karras et al., 2019) with ALAE encoder.

and stability over high-dimensional, long-transport horizons. We operate in pre-trained latent spaces to focus on semantic video dynamics over pixel redundancies. Specifically, we ablate performance across two representations: (i) the Deep Compression Autoencoder (Chen et al., 2025), which maps video frames into a $32 \times 8 \times 8 = 2048$-dimensional space, and (ii) the TiTok encoder (Yu et al., 2024), which compresses frames into a 128×32 sequence representation. We consider three action classes with approximately 100 videos each: *BreastStroke*, *BlowingCandles*, and *CuttingInKitchen*. As summarized in Table 8, WFR-based uOT methods fail to scale to this regime, often producing collapsed trajectories that do not evolve beyond the initial frame, or exhibiting poor distributional alignment that leads to high noise and distortion. In contrast, CuOT successfully recovers the underlying video dynamics. Fig. 6 shows a qualitative example for the *BreastStroke* class. In both latent representations, CuOT maintains high-fidelity reconstructions across the full temporal horizon. Detailed breakdowns across individual classes and latent representations are provided in App. F.4.

Table 8: Image-to-Video generation on the UCF101 dataset (Soomro et al., 2012).

Method	FID	$\mathcal{W}_1$	$\mathcal{W}_2$
OTCFM	$208.5315_{\pm 19.7064}$	$25.6506_{\pm 3.5847}$	$17.3783_{\pm 2.3124}$
TIGON	$336.5115_{\pm 17.0147}$	$31.0243_{\pm 4.9427}$	$20.1477_{\pm 2.8925}$
DeepRUOT	$\underline{138.2666}_{\pm 4.4798}$	$\underline{20.6668}_{\pm 2.8854}$	$\underline{14.3858}_{\pm 2.1248}$
VarRUOT	$325.7000_{\pm 13.2486}$	$28.1000_{\pm 4.7682}$	$18.3665_{\pm 3.3568}$
CuOT (Ours)	$\mathbf{18.3783}_{\pm \mathbf{1.7926}}$	$\mathbf{11.2006}_{\pm \mathbf{1.6584}}$	$\mathbf{8.1669}_{\pm \mathbf{1.0897}}$

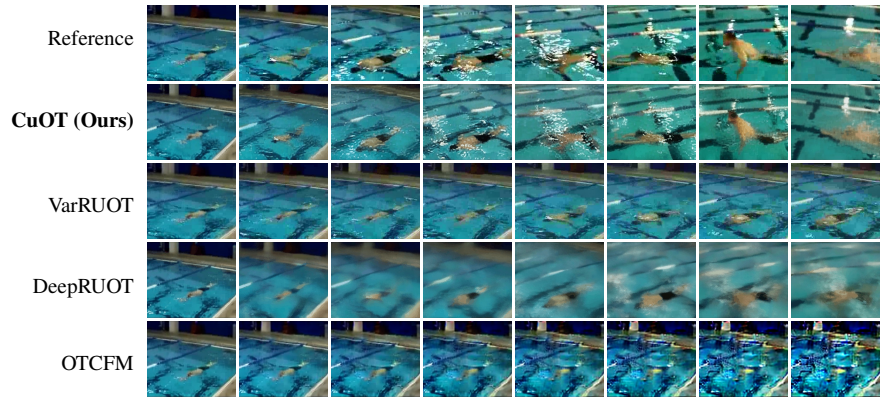

Figure 6: Image-to-Video generation on the UCF101 dataset (Soomro et al., 2012), for DC encoder.

8 Conclusions and Limitations

We introduced Contact unbalanced Optimal Transport (CuOT), a stable and scalable generative model for unpaired dataset matching grounded in contact geometry. CuOT outperforms traditional uOT approaches by eliminating structural biases toward high-density regions and adapting the sampling step size to the data geometry. As detailed in App. H, while CuOT is architecture-agnostic, validation on structured data with spatial or sequential relationships, where Transformers are the established backbone, remains open. Furthermore, establishing formal connections with diffusion methods that match data via Fisher–Rao geometry in noise space remains an open direction for future work.

References

Abraham, R. and Marsden, J. E. (2008). *Foundations of mechanics*. Number 364. American Mathematical Soc.

Alouadi, A., Barreau, B., Carlier, L., and Pham, H. (2025). Robust time series generation via Schrödinger bridge: a comprehensive evaluation. In *ACM International Conference on AI in Finance*, pages 906–914.

Ambrosio, L., Gigli, N., and Savaré, G. (2005). *Gradient flows: in metric spaces and in the space of probability measures*. Springer Science & Business Media.

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

Betts, J. T. (1998). Survey of numerical methods for trajectory optimization. *Journal of guidance, control, and dynamics*, 21(2):193–207.

Blohmann, C. (2024). Lagrangian field theory. Lecture Notes.

Bortoli, V. D., Thornton, J., Heng, J., and Doucet, A. (2021). Diffusion Schrödinger bridge with applications to score-based generative modeling. In Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems (NeurIPS)*.

Bravetti, A., Cruz, H., and Tapias, D. (2017). Contact Hamiltonian mechanics. *Annals of Physics*, 376:17–39.

Bravetti, A., Daza-Torres, M. L., Flores-Arguedas, H., and Betancourt, M. (2023). Bregman dynamics, contact transformations and convex optimization. *Information Geometry*, 6(1):355–377.

Buzun, N., Shlenskii, D., Bobrin, M., and Dylov, D. V. (2026). HOTA: Hamiltonian framework for optimal transport advection. In *International Conference on Learning Representations (ICLR)*.

Chapel, L., Flamary, R., Wu, H., Févotte, C., and Gasso, G. (2021). Unbalanced optimal transport through non-negative penalized linear regression. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Chen, J., Cai, H., Chen, J., Xie, E., Yang, S., Tang, H., Li, M., and Han, S. (2025). Deep compression autoencoder for efficient high-resolution diffusion models. In *The Thirteenth International Conference on Learning Representations*.

Chen, Y., Georgiou, T., and Pavon, M. (2014). On the relation between optimal transport and Schrödinger bridges: A stochastic control viewpoint. *arXiv 1412.4430*.

Chizat, L., Peyré, G., Schmitzer, B., and Vialard, F.-X. (2018). Unbalanced optimal transport: Dynamic and kantorovich formulations. *Journal of Functional Analysis*, 274(11):3090–3123.

Chizat, L., Zhang, S., Heitz, M., and Schiebinger, G. (2022). Trajectory inference via mean-field Langevin in path space. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 16731–16742.

Chow, S.-N., Li, W., and Zhou, H. (2020). Wasserstein Hamiltonian flows. *Journal of Differential Equations*, 268(3):1205–1219.

Cook, D. P. and Vanderhyden, B. C. (2020). Context specificity of the emt transcriptional response. *Nature communications*, 11(1):2142.

Cui, J., Liu, S., and Zhou, H. (2024). Stochastic Wasserstein Hamiltonian flows. *Journal of Dynamics and Differential Equations*, 36(4):3885–3921.

Ding, J., Sharon, N., and Bar-Joseph, Z. (2022). Temporal modelling using single-cell transcriptomics. *Nature Reviews Genetics*, 23(6):355–368.

Farrell, J. A., Wang, Y., Riesenfeld, S. J., Shekhar, K., Regev, A., and Schier, A. F. (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science*, 360(6392):eaar3131.

Fatras, K., Séjourné, T., Flamary, R., and Courty, N. (2021). Unbalanced minibatch optimal transport; applications to domain adaptation. In *International Conference on Machine Learning (ICML)*, pages 3186–3197. PMLR.

Feydy, J., Séjourné, T., Vialard, F.-X., Amari, S.-i., Trouve, A., and Peyré, G. (2019). Interpolating between optimal transport and mmd using Sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 2681–2690.

Gazdieva, M., Asadulaev, A., Burnaev, E., and Korotin, A. (2024). Light unbalanced optimal transport. In *Conference on Neural Information Processing Systems (NeurIPS)*.

Geiges, H. (2001). A brief history of contact geometry and topology. *Expositiones Mathematicae*, 19(1):25–53.

Geiges, H. (2008). An introduction to contact topology. *Cambridge Studies in Advanced Mathematics*, 109.

Gushchin, N., Kholkin, S., Burnaev, E., and Korotin, A. (2024). Light and optimal Schrödinger bridge matching. In *International Conference on Machine Learning (ICML)*, pages 17100–17122. PMLR.

Han, J., Jentzen, A., and E, W. (2018). Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510.

Herglotz, G. (1930). Berührungstransformationen. Lectures at the University of Göttingen.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:6840–6851.

Karczewski, R., Heinonen, M., Pouplin, A., Hauberg, S., and Garg, V. K. (2026). The spacetime of diffusion models: An information geometry perspective. In *International Conference on Learning Representations (ICLR)*.

Karras, T., Laine, S., and Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pages 4401–4410.

Kazakevičius, R., Kononovicius, A., Kaulakys, B., and Gontis, V. (2021). Understanding the nature of the long-range memory phenomenon in socioeconomic systems. *Entropy*, 23(9):1125.

Kharitenko, A., Shen, Z., Santi, R. D., He, N., and Dorfler, F. (2026). Landing with the score: Riemannian optimization through denoising. In *International Conference on Learning Representations (ICLR)*.

Kondratyev, S., Monsaingeon, L., and Vorotnikov, D. (2016). A new optimal transport distance on the space of finite radon measures. *Advances in Differential Equations*, 21(11-12):1117–1164.

Kumar, S., Wang, Y., and Lin, L. (2026). Flow matching is adaptive to manifold structures. *arXiv preprint arXiv:2602.22486*.

Lavenant, H., Zhang, S., Kim, Y.-H., Schiebinger, G., et al. (2024). Toward a mathematical theory of trajectory inference. *The Annals of Applied Probability*, 34(1A):428–500.

Li, W., Sun, R., Mai, H., and Zhang, T. (2025). Towards unsupervised domain bridging via image degradation in semantic segmentation. In *Conference on Neural Information Processing Systems (NeurIPS)*.

Li, Y., Xiong, K., Guo, X., Li, F., Yan, S., Xu, G., Zhou, L., Chen, L., Sun, H., WANG, B., Ma, K., Chen, G., Ye, H., Liu, W., and Wang, X. (2026a). ReCogDrive: A reinforced cognitive framework for end-to-end autonomous driving. In *International Conference on Learning Representations (ICLR)*.

Li, Z., Hu, H., Lim, S. H., Li, X., Gao, F., Diao, E., Ding, Z., Vazirgiannis, M., and Bostrom, H. (2026b). A kinetic-energy perspective of flow matching. *arXiv preprint arXiv:2602.07928*.

Liero, M., Mielke, A., and Savaré, G. (2016). Optimal transport in competition with reaction: The hellinger–kantorovich distance and geodesic curves. *SIAM Journal on Mathematical Analysis*, 48(4):2869–2911.

Lu, Y., Lu, J., and Nolen, J. (2019). Accelerating langevin sampling with birth-death. *arXiv preprint arXiv:1905.09863*.

Mokrov, P., Korotin, A., Li, L., Genevay, A., Solomon, J., and Burnaev, E. (2021). Large-scale Wasserstein gradient flows. In *Advances in Neural Information Processing Systems (NeurIPS)*.

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. (2019). Visualizing structure and transitions in high-dimensional biological data. *Nature biotechnology*, 37(12):1482–1492.

Neklyudov, K., Brekelmans, R., Severo, D., and Makhzani, A. (2023a). Action matching: Learning stochastic dynamics from samples. In *International Conference on Machine Learning (ICML)*, pages 25858–25889. PMLR.

Neklyudov, K., Brekelmans, R., Tong, A., Atanackovic, L., Liu, Q., and Makhzani, A. (2023b). A computational framework for solving Wasserstein Lagrangian flows. *arXiv preprint arXiv:2310.10649*.

Ni, H., Shi, C., Li, K., Huang, S. X., and Min, M. R. (2023). Conditional image-to-video generation with latent flow diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 18444–18455.

Pariset, M., Hsieh, Y.-P., Bunne, C., Krause, A., and De Bortoli, V. (2023). Unbalanced diffusion Schrödinger bridge. *arXiv preprint arXiv:2306.09099*.

Peng, Q., Zhou, P., and Li, T. (2024). stvcr: Reconstructing spatio-temporal dynamics of cell development using optimal transport. *bioRxiv*, pages 2024–06.

Peng, X., Usman, B., Kaushik, N., Hoffman, J., Wang, D., and Saenko, K. (2017). Visda: The visual domain adaptation challenge. *arXiv preprint arXiv:1710.06924*.

Pidhorskyi, S., Adjeroh, D. A., and Doretto, G. (2020). Adversarial latent autoencoders. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14104–14113.

Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., Skinner, D., Ramadhan, A., and Edelman, A. (2020). Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*.

Rasp, S. and Thuerey, N. (2021). Data-driven medium-range weather prediction with a resnet pretrained on climate simulations: A new model for weatherbench. *Journal of Advances in Modeling Earth Systems*, 13(2):e2020MS002405.

Richardson, E., Alaluf, Y., Patashnik, O., Nitzan, Y., Azar, Y., Shapiro, S., and Cohen-Or, D. (2021). Encoding in style: a stylegan encoder for image-to-image translation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2287–2296. IEEE.

Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., et al. (2019). Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943.

Schrödinger, E. (1931). Über die umkehrung der naturgesetze. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-Mathematische Klasse*, pages 144–153.

Schrödinger, E. (1932). Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. In *Annales de l'institut Henri Poincaré*, volume 2, pages 269–310.

Sha, Y., Qiu, Y., Zhou, P., and Nie, Q. (2024). Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. *Nature Machine Intelligence*, 6(1):25–39.

Silva, A. C. d. (2001). Lectures on symplectic geometry. *Lecture Notes in Mathematics*, 1764.

Soomro, K., Zamir, A. R., and Shah, M. (2012). Ucf101: A dataset of 101 human actions classes from videos in the wild. *arXiv preprint arXiv:1212.0402*.

Sun, Y., Zhang, Z., Wang, Z., Li, T., and Zhou, P. (2025). Variational regularized unbalanced optimal transport: Single network, least action. In *Conference on Neural Information Processing Systems (NeurIPS)*.

Testa, A., Hauberg, S., Asfour, T., and Rozo, L. (2025). Geometric contact flows: Contactomorphisms for dynamics and control. In *International Conference on Machine Learning (ICML)*.

Testa, A., Hauberg, S., Asfour, T., and Rozo, L. (2026). Contact Wasserstein geodesics for non-conservative Schrödinger bridges. In *International Conference on Learning Representations (ICLR)*.

Tokasi, S. and Pickl, P. (2022). Symplectic reduction of classical mechanics on shape space. *Foundations of Physics*, 52(5):106.

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

Tong, A., Malkin, N., Fatras, K., Atanackovic, L., Zhang, Y., Huguet, G., Wolf, G., and Bengio, Y. (2024b). Simulation-free Schrödinger bridges via score and flow matching. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*.

Udrişte, C. (2000). Geometric dynamics. *Southeast Asian Bulletin of Mathematics*, 24(2):313–322.

Vargas, F., Thodoroff, P., Lamacraft, A., and Lawrence, N. (2021). Solving Schrödinger bridges via maximum likelihood. *Entropy*, 23(9):1134.

Ventre, E., Forrow, A., Gadhiwala, N., Chakraborty, P., Angel, O., and Schiebinger, G. (2023). Trajectory inference for a branching sde model of cell differentiation. *arXiv preprint arXiv:2307.07687*.

Veres, A., Faust, A. L., Bushnell, H. L., Engquist, E. N., Kenty, J. H.-R., Harb, G., Poh, Y.-C., Sintov, E., Gürtler, M., Pagliuca, F. W., et al. (2019). Charting cellular identity during human in vitro β-cell differentiation. *Nature*, 569(7756):368–373.

Verma, Y., Heinonen, M., and Garg, V. (2024). ClimODE: Climate and weather forecasting with physics-informed neural ODEs. In *International Conference on Learning Representations (ICLR)*.

Wang, C., Li, S., He, D., and Wang, L. (2022). Is l^2 physics informed loss always suitable for training physics informed neural network? In Oh, A. H., Agarwal, A., Belgrave, D., and Cho, K., editors, *Advances in Neural Information Processing Systems (NeurIPS)*.

Wang, D., Jiang, Y., Zhang, Z., Gu, X., Zhou, P., and Sun, J. (2025). Joint velocity-growth flow matching for single-cell dynamics modeling. In *Conference on Neural Information Processing Systems (NeurIPS)*.

Wang, H., Wang, X., and Cui, Z. (2024a). Structure-preserving diffusion model for unpaired medical image translation. In *International Workshop on Machine Learning in Medical Imaging*, pages 218–227. Springer.

Wang, X., Wang, R., Peng, Q., Zhou, P., and Li, T. (2026). WFR-MFM: One-step inference for dynamic unbalanced optimal transport. *arXiv preprint arXiv:2601.20606*.

Wang, Y., Chen, Z., Chen, X., Wei, Y., Zhu, J., and Chen, J. (2024b). Framebridge: Improving image-to-video generation with bridge models. *arXiv preprint arXiv:2410.15371*.

Weed, J. and Bach, F. (2019). Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. *Bernoulli*, 25(4A):2620–2648.

Weinreb, C., Rodriguez-Fraticelli, A., Camargo, F. D., and Klein, A. M. (2020). Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science*, 367(6479):eaaw3381.

Xu, W., Fu, H., Dong, H., Zhou, Z., and Chen, C. (2025). DEAL: Diffusion evolution adversarial learning for sim-to-real transfer. In *Conference on Neural Information Processing Systems (NeurIPS)*.

Yang, L., Daskalakis, C., and Karniadakis, G. E. (2022). Generative ensemble regression: Learning particle dynamics from observations of ensembles with physics-informed deep generative models. *SIAM Journal on Scientific Computing*, 44(1):B80–B99.

Yeo, G. H. T., Saksena, S. D., and Gifford, D. K. (2021). Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions. *Nature communications*, 12(1):3222.

Yu, Q., Weber, M., Deng, X., Shen, X., Cremers, D., and Chen, L.-C. (2024). An image is worth 32 tokens for reconstruction and generation. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Zadra, F. (2023). *Topics in contact Hamiltonian systems: analytical and numerical perspectives*. PhD thesis, University of Groningen.

Zhang, J., Larschan, E., Bigness, J., and Singh, R. (2024). scNODE: generative model for temporal single cell transcriptomic data prediction. *Bioinformatics*, 40(Supplement_2):ii146–ii154.

Zhang, Z., Li, T., and Zhou, P. (2025). Learning stochastic dynamics from snapshots through regularized unbalanced optimal transport. In *International Conference on Learning Representations (ICLR)*.

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: Abstract and Section 1.

   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: Section 7, Appendix H.

   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: Section 4, 6, Appendix B.4, C.1, D.1, D.2.

   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: Section 5, Appendix E, F.

   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: [No]

   Justification: We will release code for full reproducibility upon publication of the paper.

   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: Appendix E, F.

   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: Section F, Appendix F.

   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: Appendix F.

   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: N/A

   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: Appendix G.

    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: N/A

    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: Appendix F.

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: [No]

Justification: We will release code along with examples and documentation for full reproducibility upon publication of the paper.

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: N/A

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: N/A

    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: N/A

    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.

A Notation

Table 9: Summary of mathematical notation.

Symbol	Description	Reference
Probability Space		
$\mathcal{M}$	Data domain (sample space).	Sec. 3
$\mathcal{P}^+(\mathcal{M})$	Space of smooth, positive probability densities supported on $\mathcal{M}$ (Wasserstein manifold). The associated tangent and cotangent bundles are $\mathcal{T}\mathcal{P}^+(\mathcal{M})$ and $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$, respectively.	Sec. 3
$\rho^t(x)$	Time-dependent probability density; $\rho^t \in \mathcal{P}^+(\mathcal{M})$.	Sec. 3
$\{\rho^m\}_{m=0}^M$	Observed marginal densities at times $\{t_m\}_{m=0}^M$.	Eq. (3)
Transport Dynamics		
$v^t(x)$	Velocity field governing sample transport; $v^t = \nabla_x S^t$.	Eq. (1)
$z^t(x)$	Local mass growth rate field.	Eq. (3)
$S^t(x)$	Dual potential (canonical momentum); generates the velocity via $\nabla_x S^t(x) = v^t(x)$.	Sec. 3
ε	Diffusion coefficient for entropy regularization.	Eq. (1)
w_z	Weighting factor for the influence of mass variation.	Eq. (3)
J	Objective functional minimized.	Eq. (3)
Hamiltonian Functions		
$\mathcal{H}(\rho^t, S^t)$	Symplectic Hamiltonian; $\mathcal{H} = \mathcal{K} + \varepsilon\mathcal{F}$.	Sec. 6
$\mathcal{H}(\rho^t, S^t, z^t)$	Contact Hamiltonian; $\mathcal{H} = \mathcal{K} + \varepsilon\mathcal{F} + \frac{w_z}{2}\int (z^t)^2 dx + \mathcal{H}_0$.	Eq. (4)
$\mathcal{K}(\rho^t, S^t)$	Kinetic energy; $\mathcal{K} = \int_{\mathcal{M}} \frac{1}{2}\|\nabla_x S^t\|^2 \rho^t \, dx$.	Sec. 4
$\mathcal{F}(\rho^t, S^t)$	Entropic coupling term (scaled by ε).	Sec. 4
$\mathcal{H}_0$	Reference energy constant; $\mathcal{H}_0 = \int_{\mathcal{M}} \frac{1}{2}\|v^{t_0}\|^2 \rho^{t_0} \, dx$.	Eq. (4)
Geometric Structure		
$g^{\mathcal{W}_2}$	Wasserstein metric tensor on $\mathcal{P}^+(\mathcal{M})$, defining distance $d^{\mathcal{W}_2}$.	Eq. (2)
g^C	Contact Wasserstein metric, $g^C = \sqrt{\mathcal{K}}\, g^{\mathcal{W}_2}$, defining distance $d^{\mathcal{C}}$..	Sec. 4
Δ_{ρ^t}	Weighted Laplacian, $\Delta_{\rho^t} = -\nabla_x \cdot (\rho^t \nabla_x)$.	Eq. (2)
ω	Symplectic form on $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$; preserved under OT.	Sec. 3
η	Contact form on $\mathcal{T}^*\mathcal{P}^+(\mathcal{M}) \times \mathbb{R}$.	Sec. 3
Discretization and Networks		
$\{x_n^{s_k}, w_n^{s_k}\}_{n=1}^N$	Weighted particle ensemble approximating ρ^{s_k}.	Eq. (6)
$v_\theta(x^{s_k}, s_k)$	Neural network parameterizing the velocity field.	Sec. 5
$z_\theta(x^{s_k}, s_k)$	Neural network parameterizing the mass growth rate.	Sec. 5
$\nu_n^{s_k}$	Scalar speed of particle $x_n^{s_k}$; conformal scaling factor for displacement.	Eq. (8a)
$S_n^{s_k}$	Per-particle dual potential at arc-length step s_k.	Eq. (8b)
τ	Geodesic discretization step size.	Eq. (8)
K, M, N, D	Number of geodesic steps, intermediate marginals, samples, and data dimensions.	Sec. 5

B Geometric details

B.1 Symplectic and Contact Geometries

Hamiltonian and contact Hamiltonian dynamics are governed by specific energy constraints that can be analyzed via differential geometry as flows on specialized manifolds. Hamiltonian dynamics is energy-conserving and evolves on a *symplectic manifold*. Contact Hamiltonian dynamics is more general, allowing for variable energy levels, and takes place on a *contact manifold*. Their formal definitions and key differences are discussed next.

Symplectic and contact structures. Let $\mathcal{N}$ be a smooth compact manifold, and let $\mathcal{T}_x\mathcal{N}$ denote the tangent space at $x \in \mathcal{N}$. The collection of all the tangent spaces identifies the tangent bundle $\mathcal{T}\mathcal{N} = \cup_{x\in\mathcal{N}}\mathcal{T}_x\mathcal{N}$. A vector field $X : \mathcal{N} \to \mathcal{T}\mathcal{N}$ assigns a tangent vector v to each point $x \in \mathcal{N}$. The set of all the vector fields over $\mathcal{T}\mathcal{N}$ is denoted as $\Gamma(\mathcal{T}\mathcal{N})$. A differential 1-form $\alpha : \mathcal{T}\mathcal{N} \to \mathbb{R}$ is a smooth map field acting on vectors of the tangent bundle. For a smooth function $f : \mathcal{N} \to \mathbb{R}$, the 1-form $\alpha = df$ generalizes the gradient from Euclidean spaces. Specifically, df measures the variation of f under an infinitesimal displacement on $\mathcal{N}$. This displacement is locally described by a starting point x and a direction v, such that $(x, v) \in \mathcal{T}\mathcal{N}$. Alternatively, it can be globally expressed by a vector field X. The variation of f along the vector field X is given by $df(X)$. This variation is independent of the choice of reference frame. To preserve this invariance, df must transform covariantly with X. Consequently, the 1-form $\alpha = df$ resides in the cotangent bundle $\mathcal{T}^*\mathcal{N}$, the dual space to $\mathcal{T}\mathcal{N}$. The symplectic and contact structures provide two distinct mechanisms for associating a 1-form to a vector field, thereby establishing connections between the tangent and cotangent bundles. By considering the dynamics governed by the vector field and the scalar function defining the 1-form, a relationship between these elements emerges, as illustrated in Fig. 7.

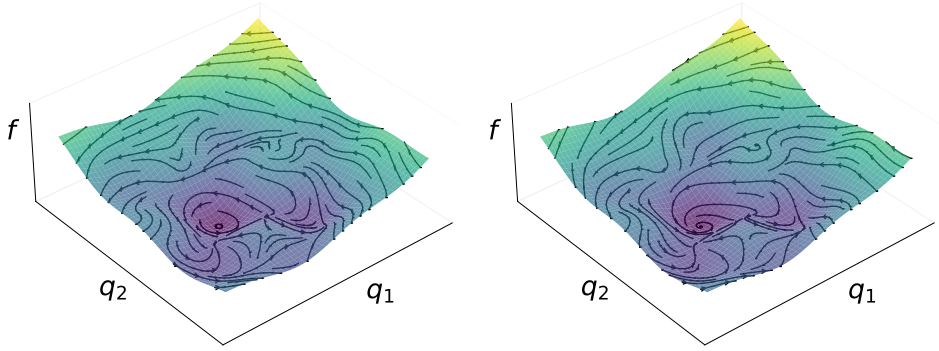

Figure 7: The same scalar function f, associated with the 1-form $\alpha = df$, gives rise to two distinct vector fields under the symplectic (left) and contact (right) geometric structures. The streamlines of these vector fields are illustrated on a representation of the state manifold. In symplectic geometry, the streamlines are tangent to the level curves of f, representing isoenergetic trajectories where f remains constant, thus describing the dynamics of conservative systems. In contrast, in contact geometry, a single flow line can traverse different energy levels.

Symplectic geometry. A differential 2-form $\omega : \mathcal{T}\mathcal{N} \times \mathcal{T}\mathcal{N} \to \mathbb{R}$ is a skew-symmetric, bilinear, and smooth field of maps acting on pairs of tangent vectors. A 2-form is called *symplectic* if it is both closed ($d\omega = 0$) and non-degenerate. The symplectic form lacks the properties required to define an inner product. However, it still establishes a fundamental relation between differential 1-forms and vector fields: Given a 1-form df, the symplectic form ω uniquely determines a vector field X_f that is tangent to the level sets of f. This relation is defined by,

$$df(X) = \omega(X_f, X), \quad \forall X \in \Gamma(\mathcal{T}\mathcal{N}). \tag{11}$$

By definition, f remains constant along the flow of X_f, which in turn preserves the symplectic form ω, i.e., $\mathcal{L}_{X_f}\omega = 0$ where $\mathcal{L}_{X_f}$ denotes the Lie derivative (Silva, 2001). Since the symplectic form captures a generalized notion of signed area on pairs of tangent vectors, its invariance under the flow implies the preservation of volume. In this framework, the function f is interpreted as a conserved energy, or equivalently, as a Hamiltonian $\mathcal{H}$. The symplectic structure thereby endows $\mathcal{N}$ with a natural geometric framework for formulating Hamiltonian dynamics (Tokasi and Pickl, 2022). The pair $(\mathcal{N}, \omega)$ is referred to as a symplectic manifold. Notably, the non-degeneracy of ω implies that $\mathcal{N}$ must be even-dimensional.

Contact geometry. While symplectic manifolds provide a geometric framework for modeling the dynamics of volume-preserving, conservative systems in classical mechanics, a more general approach is required to describe non-conservative systems. This is addressed by contact manifolds, the odd-dimensional counterparts of symplectic manifolds (Geiges, 2001; Bravetti et al., 2017). A *contact manifold* is defined as $(\mathcal{N}, \eta)$, where $\mathcal{N}$ is an odd-dimensional smooth manifold, and η is a non-degenerate 1-form known as the contact form (Geiges, 2008). The contact form satisfies the *maximal non-integrability* condition, meaning that the top-degree differential form $\eta \wedge (d\eta)^d \neq 0$ is nowhere vanishing on $\mathcal{N}$. This form is constructed by taking the exterior product of η with the d-fold wedge product of its exterior derivative $d\eta$, i.e.,

$$(d\eta)^d = \underbrace{d\eta \wedge \cdots \wedge d\eta}_{d \text{ times}}. \tag{12}$$

The $(2d+1)$-form defines a volume form on $\mathcal{N}$, ensuring that the hyperplanes $\ker(\eta) \subset \mathcal{TN}$, constraining the dynamics on the contact manifold, do not form a foliation, i.e., they do not partition the manifold into lower-dimensional submanifolds (Geiges, 2001, 2008). Geometrically, this means that the contact distribution imposes *non-holonomic constraints*: it restricts the admissible directions of motion at each point without confining the dynamics to a fixed submanifold or energy level. This property is crucial for modeling systems where energy can change over time, enabling constraints on energy behavior without enforcing conservation.

Like symplectic geometry, contact geometry connects scalar functions to vector fields, enabling the description of dynamical systems (Zadra, 2023). Given an energy function $\mathcal{H} : \mathcal{N} \to \mathbb{R}$, the dynamics on a contact manifold are defined by a contact Hamiltonian vector field $X_{\mathcal{H}}$, as follows,

$$d\mathcal{H}(X) = d\eta(X_{\mathcal{H}}, X) - \mathcal{L}_{X_{\mathcal{H}}}\eta(X), \ \forall X \in \Gamma(\mathcal{TN}). \tag{13}$$

Unlike symplectic geometry, where dynamics are confined to energy-preserving flows along the level sets of the Hamiltonian, contact geometry allows for an additional component of motion. Specifically, the dynamics on a contact manifold are not restricted to the term $d\eta(X_{\mathcal{H}}, X)$, which lies tangent to the level sets of $\mathcal{H}$, but also include a transverse component $\mathcal{L}_{X_{\mathcal{H}}}\eta(X)$, arising from the non-degeneracy of the contact form. Consequently, while in symplectic geometry the symplectic form ω is strictly preserved, contact geometry allows the contact form η to be preserved only up to a scaling factor $a \in \mathbb{R}$ (Bravetti et al., 2017). Consequently, the volume form $\eta \wedge (d\eta)^d$ need not be conserved along the flow, permitting the dynamics to unfold on regions of the contact manifold $\mathcal{N}$ of varying measure.

B.2 Riemannian Geometry

The Riemannian metric. Let $\mathcal{M}$ be a smooth compact manifold. A Riemannian metric $g : \mathcal{TM} \times \mathcal{TM} \to \mathbb{R}$ is a smooth, symmetric, and positive-definite bilinear field of maps defined on pairs of vectors in the tangent bundle. This enables the introduction of an inner product on the tangent spaces of the manifold, allowing us to measure distances and curve lengths. For a smooth curve $x(t) : [t_0, t_1] \to \mathcal{M}$, the length l w.r.t. the metric g is $l = \int_{t_0}^{t_1} \sqrt{g(\dot{x}(t), \dot{x}(t))}dt$, where $\dot{x}(t) \in \mathcal{T}_{x(t)}\mathcal{M}$ is the vector tangent to the curve at $x(t)$. The curve minimizing this length between two points $x(t_0)$ and $x(t_1)$ on $\mathcal{M}$ is called a *geodesic*. Geodesics generalize straight lines in Euclidean space to curved spaces, representing the shortest paths in the geometry induced by g.

Hamiltonian geodesics. The geodesic flow $x(t)$ on a Riemannian manifold $(\mathcal{M}, g)$ lifts to the joint evolution of coordinates $(x(t), \alpha(x(t), \dot{x}(t)))$ on the cotangent bundle $\mathcal{T}^*\mathcal{M}$ (Abraham and Marsden, 2008). This extended dynamics is governed by an energy function $\mathcal{H}(x, \alpha) : \mathcal{T}^*\mathcal{M} \to \mathbb{R} = g^{-1}(\alpha, \alpha)$, which remains constant along the flow. A reparameterization $ds = \sqrt{\mathcal{H}}dt$ links the trajectory of the integrated dynamical system at time t on $\mathcal{T}^*\mathcal{M}$ with the length of the corresponding geodesic on $\mathcal{M}$. This framework reveals a fundamental connection between geodesic flows and Hamiltonian dynamics in the special case where the Hamiltonian consists solely of a kinetic energy term $\mathcal{H} = \mathcal{K}$. The cotangent bundle $\mathcal{T}^*\mathcal{M}$ is naturally equipped with a symplectic structure, making it a symplectic manifold $(\mathcal{T}^*\mathcal{R}, \omega)$.

Contact Hamiltonian geodesics. This formulation admits a further generalization within the framework of contact Hamiltonian systems. In this setting, the contact Hamiltonian $\mathcal{H}(x, \alpha, z)$ depends explicitly on the action-like variable $z \in \mathbb{R}$, which evolves along the flow on the augmented space $(\mathcal{T}^*\mathcal{R} \times \mathbb{R}, \eta)$. Since the reparameterization $ds = \sqrt{\mathcal{K}}dt$ is no longer constant, the Hamiltonian

does not define a unique metric, rather inducing a family of Riemannian metrics $\{\sqrt{\mathcal{K}}g\}$ scaled by a z-dependent conformal factor. Consequently, the resulting dynamics can be interpreted as a piecewise geodesic flow on a manifold where the metric evolves according to the value of z.

B.3 Geometry of the Probability Space

Symplectic and Contact Structures on the Wasserstein Manifold. Denoting the Wasserstein manifold by $\mathcal{P}^+(\mathcal{M})$, its cotangent bundle $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$ is an even-dimensional space whose state is described by the local coordinate chart (ρ^t, S^t). The balanced OT problem induces a volume-preserving Hamiltonian dynamics on $(\mathcal{T}^*\mathcal{P}^+(\mathcal{M}), \omega)$, with symplectic form given by

$$\omega\big((\partial_t\rho_1^t, \partial_t S_1^t),\, (\partial_t\rho_2^t, \partial_t S_2^t)\big) := \int_{\mathcal{M}} \partial_t S_1^t(x)\, \partial_t\rho_2^t(x)\, \mathrm{d}x - \int_{\mathcal{M}} \partial_t S_2^t(x)\, \partial_t\rho_1^t(x)\, \mathrm{d}x, \tag{14}$$

where $(\partial_t\rho_1^t, \partial_t S_1^t),\, (\partial_t\rho_2^t, \partial_t S_2^t) \in \mathcal{T}_{(\rho^t, S^t)}\big(\mathcal{T}^*\mathcal{P}^+(\mathcal{M})\big)$ are admissible tangent vectors. The conservation properties of this dynamics are evident from the associated evolution equations (Appendix D.1), which preserve both the kinetic energy, as the velocity field remains constant along trajectories, and the total probability mass $\int_{\mathcal{M}} \rho^t(x)\, dx$. These conservation properties can be relaxed by augmenting the state space with an additional scalar variable $z^t \in \mathbb{R}$, yielding a *contact dynamical system* with extended state $(\rho^t, S^t, z^t) \in \mathcal{T}^*\mathcal{P}^+(\mathcal{M}) \times \mathbb{R}$. The additional degree of freedom z^t tracks the accumulated action of the dynamics and defines a direction transverse to $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$, foliating the contact manifold into a family of symplectic hyperplanes parametrized by the value of z^t. The contact manifold $(\mathcal{T}^*\mathcal{P}^+(\mathcal{M}) \times \mathbb{R}, \eta)$ is endowed with a form η, representing a signed length, specifically the projection of the velocity $(\partial_t\rho^t, \partial_t S^t, \partial_t z^t) \in \mathcal{T}_{(\rho^t, S^t, z^t)}\big(\mathcal{T}^*\mathcal{P}^+(\mathcal{M}) \times \mathbb{R}\big)$ onto the transverse direction, i.e., the part of the dynamics that escapes the symplectic hyperplane. The expression of η in local coordinates depends on the choice of coordinate system. The two most common options are the canonical coordinates, leading to η_{std_1}, and the Legendre-variant coordinates, leading to η_{std_2} (Bravetti et al., 2023),

$$\eta_{\mathrm{std}_1} = \partial_t\tilde{z}^t - \int_{\mathcal{M}} \tilde{S}^t(x)\, \partial_t\tilde{\rho}^t(x)\, \mathrm{d}x, \tag{15a}$$

$$\eta_{\mathrm{std}_2} = \partial_t z^t - \frac{1}{2}\int_{\mathcal{M}} S^t(x)\, \partial_t\rho^t(x)\, \mathrm{d}x + \frac{1}{2}\int_{\mathcal{M}} \rho^t(x)\, \partial_t S^t(x)\, \mathrm{d}x. \tag{15b}$$

As the contact form governs the relationship between the scalar contact Hamiltonian $\mathcal{H}$ and the associated vector field, each of the two forms gives rise to a distinct set of evolution equations,

$$\begin{aligned} \partial_t\tilde{\rho}^t &= \partial_{\tilde{S}}\, \mathcal{H}(\cdot), \\ \partial_t\tilde{S}^t &= -\partial_{\tilde{\rho}}\, \mathcal{H}(\cdot) - \tilde{S}^t\, \partial_{\tilde{z}}\, \mathcal{H}(\cdot), \\ \partial_t\tilde{z}^t &= \tilde{S}^t\, \partial_{\tilde{S}}\, \mathcal{H}(\cdot) - \mathcal{H}(\cdot), \end{aligned} \tag{16a}$$

$$\begin{aligned} \partial_t\rho^t &= \partial_S\, \mathcal{H}(\cdot) - \tfrac{1}{2}\, \rho^t\, \partial_z\, \mathcal{H}(\cdot), \\ \partial_t S^t &= -\partial_\rho\, \mathcal{H}(\cdot) - \tfrac{1}{2}\, S^t\, \partial_z\, \mathcal{H}(\cdot), \\ \partial_t z^t &= \tfrac{1}{2}\big(S^t + \rho^t\, \partial_\rho\, \mathcal{H}(\cdot)\big) - \mathcal{H}(\cdot). \end{aligned} \tag{16b}$$

A contact dynamics is characterized by η being rescaled by a constant factor along its flow, a condition strictly weaker than that of a symplectic dynamics: the symplectic form ω on each hyperplane need not be preserved, allowing the contact flow to traverse hyperplanes of varying volume. As shown in Appendix C.1, this volume variation on the augmented cotangent bundle induces mass variation on the probability space, providing a geometrically principled framework for the uOT without complicating the metric structure of the problem.

Geodesics on the Wasserstein Manifold. Analogously to the framework detailed in Appendix B.2, the symplectic dynamics $(\rho^t, S^t) \subset (\mathcal{T}^*\mathcal{P}^+(\mathcal{M}), \omega)$ project to a geodesic on the Wasserstein manifold $\mathcal{P}^+(\mathcal{M})$. Since the Hamiltonian function $\mathcal{H} = \int_{\mathcal{M}} \frac{1}{2}\|\nabla_x S^t(x)\|^2 \rho^t(x) dx$ represents the minimal constant kinetic energy $\mathcal{K}$ associated with the stochastic dynamics, the projected trajectory minimizes the Wasserstein distance, equivalent to the action functional (see Appendix B.4), thereby constituting a geodesic. In the contact Hamiltonian setting, the dynamics $(\rho^t, S^t, z^t) \subset (\mathcal{T}^*\mathcal{P}^+(\mathcal{M}) \times \mathbb{R}, \eta)$ are characterized by a contact Hamiltonian $\mathcal{H}(\rho^t, S^t, z^t)$ whose value evolves over time. This implies that the Wasserstein metric is no longer scaled uniformly. Instead, the geometry is governed by a dynamic scaling factor, resulting in a conformal transformation of the metric expressed as $\sqrt{\mathcal{K}}g^{\mathcal{W}_2}$.

B.4 Equivalence of Wasserstein Metric Formulations

Proposition 1. *Consider the dynamic formulation of OT (Benamou and Brenier, 2000),*

$$J = \int_{t_0}^{t_M} \int_{\mathcal{M}} \|v^t(x)\|^2 \rho^t(x)\, dx\, dt. \tag{17}$$

At optimality, this functional coincides with the Wasserstein distance between the endpoint densities.

Proof. To interpret this functional as a distance, consider a curve ρ^t connecting two probability densities. At optimality, the density evolution satisfies the continuity equation $\partial_t \rho^t + \nabla_x \cdot (\rho^t \nabla_x S^t(x))$ for some potential S^t, where $\nabla_x S^t(x) = v^t(x)$ represents a velocity field. Substituting this representation into (17) yields a kinetic energy of the form,

$$\int_{t_0}^{t_M} \int_{\mathcal{M}} \|\nabla_x S^t(x)\|^2 \rho^t(x)\, dx\, dt. \tag{18}$$

By focusing on the temporal integrand and leveraging the scalar nature of ρ^t, we can write,

$$\int_{\mathcal{M}} \|\nabla_x S^t(x)\|^2 \rho^t(x)\, dx = \int_{\mathcal{M}} \nabla_x S^t(x) \cdot \left(\rho^t(x) \nabla_x S^t(x)\right) dx. \tag{19}$$

Applying integration by parts (assuming suitable boundary conditions),

$$\int_{\mathcal{M}} S^t(x) \left(-\nabla_x \cdot (\rho^t(x) \nabla_x S^t(x))\right) dx = \int_{\mathcal{M}} S^t(x)\, \Delta_\rho\, S^t(x)\, dx. \tag{20}$$

The continuity equation connects the density variation $\partial_t \rho^t$ and the potential S^t by the weighted Laplacian operator, $\Delta_\rho = -\nabla_x \cdot (\rho^t \nabla_x)$ (Chow et al., 2020), as

$$\partial_t \rho^t(x) = \Delta_\rho S^t(x), \quad \text{equivalently} \quad S^t(x) = \Delta_\rho^\dagger \partial_t \rho^t(x). \tag{21}$$

Using (21) in , this becomes

$$g^{\mathcal{W}_2} = \int_{\mathcal{M}} S^t(x)\, \partial_t \rho^t(x)\, dx. \tag{22}$$

Substituting $S^t(x) = \Delta_\rho^\dagger \partial_t \rho^t(x)$ in (20), we obtain

$$g^{\mathcal{W}_2^2}(\partial_t \rho^t, \partial_t \rho^t) = \int_{\mathcal{M}} \partial_t \rho^t(x)\, (-\Delta_\rho)^\dagger \partial_t \rho^t(x)\, dx, \tag{23}$$

which matches the standard formulation of the Wasserstein metric (2). Therefore, the kinetic energy functional (17) can be equivalently expressed as

$$J = \int_{t_0}^{t_M} g^{\mathcal{W}_2^2}(\partial_t \rho^t, \partial_t \rho^t)\, dt, \tag{24}$$

which corresponds to the energy of a curve in $\mathcal{P}^+(\mathcal{M})$ endowed with the Wasserstein metric. Minimizing J thus amounts to computing a geodesic. This interpretation can be made explicit by introducing an arc-length parametrization. Defining $ds = \sqrt{\|v^t(x)\|^2 \rho^t(x)} dt = \sqrt{g^{\mathcal{W}_2}(\partial_t \rho^t, \partial_t \rho^t)}\, dt$, the functional reduces to $J = \int_{s_0}^{s_M} ds$, showing that the minimization of kinetic energy is equivalent to minimizing the length of the curve. □

C Contact Unbalanced Optimal Transport

C.1 Contact Hamiltonian Structure of Unbalanced Optimal Transport

This part presents the derivation of the Hamiltonian structure of the Contact uOT problem (3) (Section 4), obtained via the method of Lagrange multipliers. To transform the constrained optimization

problem (3) into an unconstrained one, we incorporate the generalized Fokker–Planck equation, scaled by the Lagrange multiplier S^t, into the original running cost, yielding the augmented Lagrangian,

$$J(v^t, \rho^t, S^t, z^t) = \int_{t_0}^{t_M} \int_{\mathcal{M}} \mathcal{L}(v^t, \rho^t, S^t, z^t)\, dx\, dt; \tag{25}$$

$$\mathcal{L}(v^t, \rho^t, S^t, z^t) = \tfrac{1}{2}\|v^t(x)\|^2 + \tfrac{1}{2} w_z z^t(x)^2 + S^t(x) \underbrace{\Big(\partial_t \rho^t(x) + \nabla_x \cdot \big(\rho^t(x)\, v^t(x)\big)\Big)}_{\text{Fokker–Planck equation}}, \tag{26}$$

dependent on the optimization variable v^t and the contact state (ρ^t, S^t, z^t). The problem (3) falls outside the scope of classical variational calculus due to the explicit dependence of the cost functional J on the evolving action z^t. Instead, it is naturally framed within non-conservative variational principles, where J evolves dynamically with the system. This setting is naturally addressed by the Herglotz variational principle, which extends the Euler–Lagrange equations to systems with dissipative effects. Applying this framework to derive the optimality conditions leads to a contact Hamiltonian dynamical system expressed in canonical coordinates (Bravetti et al., 2023). For our purposes, however, it is particularly convenient to employ an equivalent non-canonical representation (Appendix B.1), as it simplifies the structure of the Hamiltonian. To carry out the Hamiltonian derivation using Herglotz-type Euler–Lagrange equations, it is therefore necessary to transition from the non-canonical state (ρ^t, S^t, z^t) to the canonical one $(\tilde{\rho}^t, \tilde{S}^t, \tilde{z}^t)$, via the substitution introduced in (Bravetti et al., 2023),

$$\rho^t(x) = \tfrac{1}{2}\tilde{\rho}^t(x) - \tilde{S}^t(x), \tag{27a}$$

$$S^t(x) = \tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x), \tag{27b}$$

$$z^t(x) = \tilde{z}^t(x) + \tfrac{1}{8}\tilde{\rho}^t(x)^2 - \tfrac{1}{2}\tilde{S}^t(x)^2. \tag{27c}$$

Accordingly, the Lagrangian (26) transforms as follow,

$$\begin{aligned}\mathcal{L} =& \tfrac{1}{2}\|v^t\|^2 \left(\tfrac{1}{2}\tilde{\rho}^t(x) - \tilde{S}^t(x)\right) + \tfrac{1}{2} w_z \left(\tilde{z}^t(x) + \tfrac{1}{8}\tilde{\rho}^t(x)^2 - \tfrac{1}{2}\tilde{S}^t(x)^2\right)^2 + \left(\tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x)\right) \cdot \\ & \left(\tfrac{1}{2}\partial_t \tilde{\rho}^t(x) - \partial_t \tilde{S}^t(x) + \nabla_x \cdot \left(\left(\tfrac{1}{2}\tilde{\rho}^t(x) - \tilde{S}^t(x)\right) \cdot v^t\right) - \varepsilon \Delta_x \left(\tfrac{1}{2}\tilde{\rho}^t(x) - \tilde{S}^t(x)\right)\right)\end{aligned} \tag{28}$$

The optimality conditions obtained from the variations of $\mathcal{L}$, namely, the Herglotz-type Euler–Lagrange equations, for a generic field argument $\psi^t(x)$ in this case take the form,

$$d_\psi \mathcal{L} = \partial_\psi \mathcal{L} - \partial_t(\partial_{\partial_t \psi}\mathcal{L}) - \nabla_x \cdot (\partial_{\nabla_x \psi}\mathcal{L}) + \Delta_x(\partial_{\Delta_x}\mathcal{L}) - \partial_z \mathcal{L}\, \partial_{\partial_t \psi}\mathcal{L} = 0. \tag{29}$$

Applying the Herglotz-type Euler–Lagrange equations (29) to the Lagrangian $\mathcal{L}$ defined in equation (28), with respect to the fields v^t, $\tilde{\rho}^t$, and $\tilde{S}^t$, yields the following optimality conditions,

$$\begin{aligned} d_v \mathcal{L} &= v \left(\tfrac{1}{2}\tilde{\rho}^t(x) - \tilde{S}^t(x)\right) - \nabla_x \left(\tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x)\right) \left(\tfrac{1}{2}\tilde{\rho}^t(x) - \tilde{S}^t(x)\right) = 0 \\ &\implies v^t = \nabla_x \left(\tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x)\right), \end{aligned} \tag{30a}$$

$$\begin{aligned} d_{\tilde{\rho}} \mathcal{L} =& \tfrac{1}{4}\|v\|^2 + w_z \left(\tilde{z}^t(x) + \tfrac{1}{8}\tilde{\rho}^t(x)^2 - \tfrac{1}{2}\tilde{S}^t(x)^2\right) \tfrac{1}{4}\tilde{\rho}^t(x) \\ &+ \tfrac{1}{2}\left(\tfrac{1}{2}\partial_t \tilde{\rho}^t(x) - \partial_t \tilde{S}^t(x) + \nabla_x \cdot \left(\left(\tfrac{1}{2}\tilde{\rho}^t(x) - \tilde{S}^t(x)\right) \cdot v^t\right) - \varepsilon \left(\tfrac{1}{2}\Delta_x \tilde{\rho}^t(x) - \Delta_x \tilde{S}^t(x)\right)\right) \\ &- \tfrac{1}{2}\nabla_x(\tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x))v^t - \tfrac{1}{2}\partial_t \tilde{S}^t(x) - \tfrac{1}{4}\partial_t \tilde{\rho}^t(x) - \tfrac{1}{2}\varepsilon \Delta_x \tilde{S}^t(x) - \tfrac{1}{4}\varepsilon \Delta_x \tilde{\rho}^t(x) \\ &- \tfrac{1}{2} w_z (\tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x)) \left(\tilde{z}^t(x) + \tfrac{1}{8}\tilde{\rho}^t(x)^2 - \tfrac{1}{2}\tilde{S}^t(x)^2\right) = 0, \end{aligned} \tag{30b}$$

$$\begin{aligned} \partial_{\tilde{S}} \mathcal{L} =& -\tfrac{1}{2}\|v\|^2 - w_z \left(\tilde{z}^t(x) + \tfrac{1}{8}\tilde{\rho}^t(x)^2 - \tfrac{1}{2}\tilde{S}^t(x)^2\right) \tilde{S}^t(x) \tilde{S}^t(x) \\ &+ \left(\tfrac{1}{2}\partial_t \tilde{\rho}^t(x) - \partial_t \tilde{S}^t(x) + \nabla_x \cdot \left(\left(\tfrac{1}{2}\tilde{\rho}^t(x) - \tilde{S}^t(x)\right) \cdot v^t\right) - \varepsilon \left(\tfrac{1}{2}\Delta_x \tilde{\rho}^t(x) - \Delta_x \tilde{S}^t(x)\right)\right) \\ &+ \nabla_x(\tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x))v^t + \tfrac{1}{2}\partial_t \tilde{\rho}^t(x) + \partial_t \tilde{S}^t(x) + \varepsilon \Delta_x \tilde{S}^t(x) + \tfrac{1}{2}\varepsilon \Delta_x \tilde{\rho}^t(x) \\ &+ w_z (\tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x)) \left(\tilde{z}^t(x) + \tfrac{1}{8}\tilde{\rho}^t(x)^2 - \tfrac{1}{2}\tilde{S}^t(x)^2\right) = 0 \end{aligned} \tag{30c}$$

Rearranging the expression and performing the necessary algebraic simplifications yields,

$$\partial_t \tilde{S}^t(x) = \tfrac{1}{2}\nabla_x \cdot \left(\left(\tfrac{1}{2}\tilde{\rho}^t(x) - \tilde{S}^t(x)\right)\nabla_x\left(\tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x)\right)\right) - \tfrac{1}{4}\left\|\nabla_x\left(\tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x)\right)\right\|^2 - \tfrac{1}{2}\varepsilon\Delta_x\tilde{\rho}^t(x) - \tfrac{1}{2}w_z\left(\tilde{z}^t(x) + \tfrac{1}{8}\tilde{\rho}^t(x)^2 - \tfrac{1}{2}\tilde{S}^t(x)^2\right)\tilde{S}^t(x). \tag{31a}$$

$$\partial_t \tilde{\rho}^t(x) = -\nabla_x \cdot \left(\left(\tfrac{1}{2}\tilde{\rho}^t(x) - \tilde{S}^t(x)\right)\nabla_x\left(\tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x)\right)\right) - \tfrac{1}{2}\left\|\nabla_x\left(\tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x)\right)\right\|^2 - 2\varepsilon\Delta_x\tilde{S}^t(x) + \tfrac{1}{2}w_z\left(\tilde{z}^t(x) + \tfrac{1}{8}\tilde{\rho}^t(x)^2 - \tfrac{1}{2}\tilde{S}^t(x)^2\right)\tilde{\rho}^t(x), \tag{31b}$$

Taking the sum and difference of equations (31b) and (31a), we obtain,

$$\begin{aligned}\tfrac{1}{2}\partial_t\tilde{\rho}^t(x) - \partial_t\tilde{S}^t(x) &= -\nabla_x \cdot \left(\left(\tfrac{1}{2}\tilde{\rho}^t(x) - \tilde{S}^t(x)\right)\nabla_x\left(\tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x)\right)\right) \\ &\quad + \varepsilon\Delta_x\left(\tfrac{1}{2}\tilde{\rho}^t(x) - \tilde{S}^t(x)\right) \\ &\quad + \tfrac{1}{2}w_z\left(\tilde{z}^t(x) + \tfrac{1}{8}\tilde{\rho}^t(x)^2 - \tfrac{1}{2}\tilde{S}^t(x)^2\right)\left(\tfrac{1}{2}\tilde{\rho}^t(x) - \tilde{S}^t(x)\right), \qquad (32a)\\ \partial_t\tilde{S}^t(x) + \tfrac{1}{2}\partial_t\tilde{\rho}^t(x) &= -\tfrac{1}{2}\left\|\nabla_x\left(\tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x)\right)\right\|^2 - \varepsilon\Delta_x\left(\tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x)\right) \\ &\quad + \tfrac{1}{2}w_z\left(\tilde{z}^t(x) + \tfrac{1}{8}\tilde{\rho}^t(x)^2 - \tfrac{1}{2}\tilde{S}^t(x)^2\right)\left(\tilde{S}^t(x) + \tfrac{1}{2}\tilde{\rho}^t(x)\right). \qquad (32b)\end{aligned}$$

At this stage, we transform equations (32) using the inverse of the canonical coordinate transformation (27), namely,

$$\tilde{\rho}^t(x) = \rho^t(x) + S^t(x), \tag{33a}$$
$$\tilde{S}^t(x) = \tfrac{1}{2}(S^t(x) - \rho^t(x)), \tag{33b}$$
$$\tilde{z}^t(x) = z^t(x) - \tfrac{1}{2}\rho^t(x)S^t(x), \tag{33c}$$

and recover the dynamical equation for the action z^t (3b). This yields the optimality conditions in equation (5), expressed in non-canonical coordinates,

$$\begin{aligned}\partial_t\rho^t(x) &= \partial_S\mathcal{H}(\cdot) - \frac{1}{2}\rho^t(x)\partial_z\mathcal{H}(\cdot) \\ &= -\nabla_x\cdot\left(\rho^t(x)\nabla_x S^t(x)\right) + \varepsilon\Delta_x\rho^t(x) + w_z\, z^t(x)\rho^t(x), \\ \partial_t S^t(x) &= -\partial_\rho\mathcal{H}(\cdot) - \frac{1}{2}S^t(x)\partial_z\mathcal{H}(\cdot) \\ &= -\tfrac{1}{2}\|\nabla_x S^t(x)\|^2 - \varepsilon\Delta_x S^t(x) + w_z\, z^t(x)S^t(x), \\ \partial_t z^t(x) &= \frac{1}{2}\left(S^t(x)\partial_S\mathcal{H}(\cdot) + \rho^t(x)\partial_\rho\mathcal{H}(\cdot)\right) - \mathcal{H}(\cdot) \\ &= \tfrac{1}{2}\|\nabla_x S^t(x)\|^2\rho^t(x) + \tfrac{1}{2}w_z\, z^t(x)^2 - \tfrac{1}{2}\|v_0\|^2\rho^t(x),\end{aligned}$$

This defines a contact Hamiltonian system associated with the following contact Hamiltonian function,

$$\begin{aligned}\mathcal{H}(\rho^t, S^t, z^t) = &\int_{\mathcal{M}} \tfrac{1}{2}\|\nabla_x S^t(x)\|^2\rho^t(x)\,dx + \varepsilon\int_{\mathcal{M}} S^t(x)\,\Delta_x\rho^t(x)\,dx \\ &- w_z\int_{\mathcal{M}}\tfrac{1}{2}z^t(x)^2\,dx - \int_{\mathcal{M}}\tfrac{1}{2}\|\nabla_x S^t(x)\|^0\rho^0(x)\,dx.\end{aligned} \tag{35}$$

Here, the first term represents the kinetic energy of the flow, while the second term introduces an entropic coupling scaled by the diffusion coefficient ε. The latter modifies the drift of the transport map so as to maintain a level of entropy consistent with the reference process in the Fokker–Planck equation (3c). The third term constitutes the damping factor, responsible for drawing or increasing energy. As direct consequences, this have the introduction of a damping or accelerating term in the Hamilton-Jacobi-Bellman equation (5b), and a source or sink term in the Fokker-Planck equation (5a). The last term represents a constant fixing the reference energy level.

D Geometric Analysis of State-of-the-Art Methods

D.1 Hamiltonian Structure of the Optimal Transport

This part presents the derivation of the Hamiltonian structure of the entropy-regularized OT problem (9) (Section 3), obtained via the method of Lagrange multipliers (Testa et al., 2026). To transform the constrained optimization problem into an unconstrained one, we incorporate the Fokker–Planck equation, scaled by the Lagrange multiplier S^t, into the original running cost $\mathcal{L}$ (i.e., the Lagrangian), as follows,

$$J(v^t, \rho^t, S^t) = \int_{t_0}^{t_M} \mathcal{L}(v^t, \rho^t, S^t)\, dt; \tag{36}$$

$$\mathcal{L}(v^t, \rho^t, S^t) = \int_{\mathcal{M}} \frac{1}{2}\|v^t(x)\|^2 \rho^t(x)\, dx + \int_{\mathcal{M}} S^t(x) \underbrace{\left(\partial_t \rho^t(x) + \nabla_x \cdot (\rho^t(x) v^t(x)) - \varepsilon \Delta \rho^t(x)\right)}_{\text{Fokker-Planck equation}} dx. \tag{37}$$

The optimality conditions resulting from the extremization of the cost functional J in equation (36) follow from the Euler–Lagrange equations, generalized to the setting of classical field theory (Blohmann, 2024). In this framework, the arguments of the Lagrangian $\mathcal{L}$, in equation (37), are viewed as smooth fields defined over space and time. By setting to zero the variations of $\mathcal{L}$ with respect to these fields, we obtain the stationarity conditions for J. For a generic field $\psi^t(x)$, the corresponding Euler–Lagrange equation takes the form,

$$d_\psi \mathcal{L} = \partial_\psi \mathcal{L} - \partial_t(\partial_{\partial_t \psi}\mathcal{L}) - \nabla_x \cdot (\partial_{\nabla_x \psi}\mathcal{L}) + \Delta_x(\partial_{\Delta_x}\mathcal{L}) = 0. \tag{38}$$

Applying equation (38) to equation (37) for the fields v^t, ρ^t, and S^t, we obtain the following system of optimality conditions,

$$d_v \mathcal{L} = v^t(x)\rho^t(x) - \rho^t(x)\nabla_x S^t(x) = 0 \implies v^t(x) = \nabla_x S^t(x), \tag{39a}$$

$$d_\rho \mathcal{L} = \frac{1}{2}\|v^t(x)\|^2 - \partial_t S^t(x) - \nabla_x S^t(x) \cdot v^t(x) - \varepsilon \Delta_x S^t(x) = 0, \tag{39b}$$

$$d_S \mathcal{L} = \partial_t \rho^t(x) + \nabla_x \cdot (\rho^t(x) v(x)) - \varepsilon \Delta_x \rho^t(x) = 0. \tag{39c}$$

Substituting the expression for the optimal velocity from equation (39a) into equations (39b) and (39c), we obtain the following Hamiltonian system,

$$\partial_t \rho^t(x) = \partial_S \mathcal{H}(\cdot) = -\nabla_x \cdot (\rho^t(x)\nabla_x S^t(x)) + \varepsilon \Delta \rho^t(x), \tag{40a}$$

$$\partial_t S^t(x) = -\partial_\rho \mathcal{H}(\cdot) = -\frac{1}{2}\|\nabla_x S^t(x)\|^2 - \varepsilon \Delta_x S^t(x), \tag{40b}$$

with the corresponding Hamiltonian function,

$$\mathcal{H}(\rho^t, S^t) = \frac{1}{2}\int_{\mathcal{M}} \|\nabla_x S^t(x)\|^2 \rho^t(x)\, dx + \varepsilon \int_{\mathcal{M}} S^t(x) \Delta_x \rho^t(x)\, dx. \tag{41}$$

Here, the first term represents the kinetic energy of the flow, while the second term introduces an entropic coupling scaled by the diffusion coefficient ε. The latter modifies the drift of the transport map so as to maintain a level of entropy consistent with the Brownian motion used as the reference process. In the limiting case $\varepsilon \to 0$, the entropic correction vanishes and one recovers a classical OT plan, wherein the minimization of the Hamiltonian kinetic energy on $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$ is equivalent to the computation of a geodesic on the Wasserstein manifold $\mathcal{P}^+(\mathcal{M})$, with the kinetic energy coinciding with the Wasserstein metric $g^{\mathcal{W}_2}$ defined in equation (2) (Appendix B.4). In the general case $\varepsilon > 0$, the Hamiltonian governs the minimization of an entropy-regularized transport cost, whose solution is equivalent to that obtained via the Sinkhorn algorithm (Feydy et al., 2019).

D.2 Hamiltonian Structure of the Fisher-Rao Unbalanced Optimal Transport

This part presents the derivation of the Hamiltonian structure of the Fisher–Rao uOT problem (10) (Section 3), obtained via the method of Lagrange multipliers. To transform the constrained optimization problem (10) into an unconstrained one, we incorporate the generalized Fokker–Planck

equation, scaled by the Lagrange multiplier S^t, into the original running cost, yielding the augmented Lagrangian,

$$J(v^t, z^t, \rho^t, S^t) = \int_{t_0}^{t_M} \mathcal{L}(v^t, z^t, \rho^t, S^t)\, dt; \tag{42}$$

$$\begin{aligned}\mathcal{L}(v^t, z^t, \rho^t, S^t) = &\int_{\mathcal{M}} \frac{1}{2}\Big(\|v^t(x)\|^2 + w_z\, z^t(x)^2\Big)\rho^t(x)\, dx \\ &+ \int_{\mathcal{M}} S^t(x) \underbrace{\Big(\partial_t \rho^t(x) + \nabla_x \cdot \big(\rho^t(x)\, v^t(x)\big) - \varepsilon \Delta_x \rho^t(x) - z^t(x)\rho^t(x)\Big)}_{\text{Generalized Fokker–Planck equation}} dx.\end{aligned} \tag{43}$$

Applying the Euler–Lagrange equations (43) to $\mathcal{L}$ with respect to the fields v^t, z^t, ρ^t, and S^t, we obtain the following optimality conditions,

$$d_v\mathcal{L} = v^t(x)\,\rho^t(x) - \rho^t(x)\,\nabla_x S^t(x) = 0 \quad \Longrightarrow \quad v^t(x) = \nabla_x S^t(x), \tag{44a}$$

$$d_g\mathcal{L} = w_z\, z^t(x)\,\rho^t(x) - S^t(x)\,\rho^t(x) = 0 \quad \Longrightarrow \quad z^t(x) = \frac{S^t(x)}{w_z}, \tag{44b}$$

$$d_\rho\mathcal{L} = \frac{1}{2}\|v^t(x)\|^2 + \frac{w_z}{2} z^t(x)^2 - \partial_t S^t(x) - \nabla_x S^t(x)\cdot v^t(x) - \varepsilon\Delta_x S^t(x) - z^t(x) S^t(x) = 0, \tag{44c}$$

$$d_S\mathcal{L} = \partial_t \rho^t(x) + \nabla_x \cdot \big(\rho^t(x)\, v^t(x)\big) - \varepsilon\Delta_x\rho^t(x) - z^t(x)\rho^t(x) = 0. \tag{44d}$$

Substituting the optimal velocity (44a) and growth rate (44b) into equations (44c) and (44d), we obtain the Hamiltonian system,

$$\partial_t \rho^t(x) = \quad \partial_S \mathcal{H}(\cdot) = -\nabla_x \cdot \big(\rho^t(x)\,\nabla_x S^t(x)\big) + \varepsilon\Delta_x\rho^t(x) + \frac{S^t(x)}{w_z}\rho^t(x), \tag{45a}$$

$$\partial_t S^t(x) = -\partial_\rho \mathcal{H}(\cdot) = -\frac{1}{2}\|\nabla_x S^t(x)\|^2 - \frac{1}{2w_z} S^t(x)^2 - \varepsilon\Delta_x S^t(x), \tag{45b}$$

governed by the Hamiltonian function,

$$\mathcal{H}(\rho^t, S^t) = \frac{1}{2}\int_{\mathcal{M}} \left(\|\nabla_x S^t(x)\|^2 + \frac{1}{w_z} S^t(x)^2\right)\rho^t(x)\, dx + \varepsilon \int_{\mathcal{M}} S^t(x)\,\Delta_x\rho^t(x)\, dx. \tag{46}$$

Compared to the OT Hamiltonian (41), the Fisher–Rao unbalanced formulation introduces an explicit dependence on S^t in the kinetic energy. This dependence arises from the deformation of the underlying conic space induced by the growth rate S^t. Despite this structural complication, the quadratic Fisher–Rao penalty, $\psi(z^t(x)) = \big(z^t(x)\big)^2$, remains the most tractable choice within the class of ψ-divergences. Indeed, it is the only choice that preserves a polynomial structure of the Hamiltonian. By contrast, a generic convex ψ-divergence induces a transcendental dependence on S^t, thereby significantly increasing the analytical complexity of the Hamiltonian and hindering the explicit computation of geodesics (Kondratyev et al., 2016). In this setting, distance-based regularization cannot be recovered, requiring Hamilton-Jacobi-Bellman regularization (Buzun et al., 2026), whose PDE-based formulation lacks comparable optimization stability (see Appendix E.4).

The dynamical equations (45) describe a symplectic flow on $\mathcal{T}^*\mathcal{P}^+(\mathcal{M})$, corresponding to a geodesic with respect to the WFR metric. Indeed, as in problems (3) and (9), $\mathcal{K}$ defines a proper geometric metric that is linear in ρ^t, allowing the WFR geodesic length of $\rho^s \subset \mathcal{P}^+(\mathcal{M})$ to be computed as an expectation over samples $x_n^{s_k}{}_{n=0}^{N}$ (Neklyudov et al., 2023b). In the limit $w_z \to \infty$, mass variation is suppressed and this formulation recovers the classical OT (9), constituting a principled extension thereof. However, the optimality conditions $v^t(x) = \nabla_x S^t(x)$ and $z^t(x) = S^t(x)$ drives the transport dynamics to regions of maximal mass growth. This structural bias leads to mode collapse by discouraging the reconstruction of small target modes (Chapel et al., 2021; Zhang et al., 2025), where sample weights must naturally decrease. This bias is also unstable as it regulates the transport velocity w.r.t. the gradient of the weight growth, often noisy in real datasets. This limitation is baked into the symplectic manifold $(\mathcal{T}^*\mathcal{P}^+(\mathcal{M}), \omega)$, which enforcing volume-preserving dynamics, can only accommodate mass variation by rigidly coupling z^t and v^t within a fixed bound. Notably, CuOT overcomes the symplectic structural limitations by exploiting contact geometry to decouple transport from mass growth, offering a principled framework for more accurate, controlled adaptation to mass changes.

D.3 Contact Hamiltonian Structure of the Non-Conservative Optimal Transport

Our formulation is inspired by Testa et al. (2026), who reformulated the conservative Hamiltonian OT problem (9) via a contact framework to let energy vary during transport, yielding more flexible maps with non-constant velocity v^t that better capture real-world processes involving energy dissipation and external interactions,

$$\min_{v^t} J(v^t, \rho^t) = \int_0^1 \partial_t z^t \, dt; \tag{47a}$$

$$\text{s.t. } \partial_t z^t = \int_{\mathcal{M}} \frac{1}{2} \|v^t(x)\|^2 \rho^t(x) \, dx - \gamma z^t; \tag{47b}$$

$$\partial_t \rho^t(x) + \nabla_x \cdot \big(\rho^t(x)\, v^t(x)\big) = \varepsilon \Delta_x \rho^t(x); \tag{47c}$$

$$\rho^{t_m} = \rho_m, \, \forall m \in \{0, \dots, M\}. \tag{47d}$$

Here, z^t represents a scalar Lagrangian action and $\gamma \in \mathbb{R}$ is a constant damping factor. Our problem (3) extends their framework in two fundamental ways by introducing (i) a spatially varying action field through the quadratic term $\frac{1}{2} z^t(x)^2$ in equation (5c), instead of the scalar linear action z^t in (47b), and (ii) a corresponding source/sink term $w_z \rho^t(x) z^t(x)$ in equation (5a), which is missing in (47c). These differences reflect a different interpretation of the contact action. In Testa et al. (2026), z^t is a time-dependent, non-learnable scalar obtained by integrating the dynamics generated by the learned velocity field. Through the relation (47b), it acts as a global damping or amplification variable that modulates the energy of the dynamics according to a prescribed coefficient γ, thereby allowing non-conservative dynamics with a predefined energy trend. In contrast, we reinterpret $z^t(x)$ as a local mass growth-rate field learned from the data, allowing the energy of the transport map to adapt to the geometry of the space. To this end, we introduce a quadratic term $\frac{1}{2} z^t(x)^2$ to obtain a source/sink term $w_z \rho^t(x) z^t(x)$ in equation (5a) via the contact Hamiltonian equations, ensuring a theoretically consistent evolution of the density. Consequently, while their approach is restricted to balanced OT problems, our method extends to unbalanced OT and adapts the velocity field v^t to the local data density,refining the integration step size in high-density regions rather than relying on an a priori schedule for v^t based on a prescribed damping coefficient γ.

Moreover, the formal derivation of problem (3) diverges from that of Testa et al. (2026) by employing a different basis on the contact manifold (the Legendre-variant coordinates). A discussion of this alternative representation, alongside the formal mathematical proofs for our derivation, is provided in Appendices B.3 and C.1.

E Implementation Details

E.1 Architecture

The transport and growth networks v_θ and z_θ share a common backbone: a time-conditioned multilayer perceptron (MLP) with a sinusoidal Fourier feature encoding of the input. Concretely, given a particle position $x_n^{s_k} \in \mathbb{R}^d$ and a scalar variable $s_k \in [0, 1]$, the concatenated input $[x_n^{s_k}, s_k] \in \mathbb{R}^{d+1}$ is lifted to a higher-dimensional representation via a multi-scale Fourier encoding,

$$\begin{aligned} \phi([x_n^{s_k}, s_k]) = \big[&[x_n^{s_k}, \, s_k, \, \sin(2^0 [x_n^{s_k}, s_k]), \, \cos(2^0 [x_n^{s_k}, s_k]), \, \dots, \\ &\sin(2^{F-1} [x_n^{s_k}, s_k]), \, \cos(2^{F-1} [x_n^{s_k}, s_k])\big], \end{aligned} \tag{48}$$

where F denotes the number of frequency bands. The encoded input $\phi([x_n^{s_k}, s_k])$ is then processed by a shared trunk consisting of L hidden layers of width H, with `SiLU` activations and `LayerNorm` normalization applied after each hidden layer to prevent mode collapse in the learned vector fields. The trunk output is then passed to two separate linear heads: a velocity head producing $v_\theta(x_n^{s_k}, s_k) \in \mathbb{R}^d$, and a growth head producing $z_\theta(x_n^{s_k}, s_k) \in \mathbb{R}$. The full architecture is summarized as

$$h = \text{Trunk}\big(\phi([x_n^{s_k}, s_k])\big), \qquad v_\theta(x_n^{s_k}, s_k) = W_v h, \qquad z_\theta(x_n^{s_k}, s_k) = W_g h, \tag{49}$$

where $W_v \in \mathbb{R}^{d \times H}$ and $W_g \in \mathbb{R}^{1 \times H}$ are the learned head weights. The number of trunk layers L, hidden width H, and number of frequency bands F are treated as hyperparameters and tuned per dataset (see Table 11).

Algorithm 1 Training procedure for the CuOT method.

Input: Marginal samples $\{x_n^m\}_{n=1}^{N_m} \sim \rho^m$ for $m = 0, \dots, M$; integration step τ.
Output: Trained transport and growth networks v_θ, z_θ.

```
Initialize network parameters θ for v_θ and z_θ.
for each training epoch do
    for each mini-batch do
        Sample a batch {x_n^0}_{n=1}^N ~ ρ^0 from the initial marginal.
        Propagate the discrete geodesic {(x_n^{s_k}, w_n^{s_k})}_{k=0}^K forward via (6).
        Compute the optimality variables {(ν_n^{s_k}, z_n^{s_k})}_{k=0}^K from the current z_θ via (8).
        Sample reference batches {x_n^m}_{n=1}^{N_m} ~ ρ^m for m = 1, ..., M, and evaluate the loss (7a).
        Update θ via gradient descent on (7a).
    end for
end for
```

E.2 Discrete optimal equations

Given an estimated discrete geodesic $\{(x_n^{s_k}, w_n^{s_k})\}_{k=0}^K$, the contact optimality conditions (5) can be deterministically discretized by evaluating all fields at the current sample positions. Since all distributional distances in our solver are computed with entropy-regularized Sinkhorn divergences, the entropy contribution associated with the diffusion term is already incorporated by the distance evaluation. Consequently, in the discretized dynamics we consider the limit $\varepsilon \to 0$. Moreover, the solver integrates the dynamics along the particle trajectories, so each transition $k \to k+1$ advances the trajectory in time and updates the particle position. Therefore, all derivatives appearing in equations (5) are evaluated along the same discrete path using forward finite differences, $Df = (f_n^{s_{k+1}} - f_n^{s_k})/\tau$, where D denotes the corresponding derivative along the trajectory (time derivative for scalar fields and directional derivative for spatial quantities). Under this trajectory-based discretization, equations (5b) and (5c) directly yield equations (8b) and (8c), respectively. Equation (8a) is then obtained by taking the spatial derivative of equation (5b) along the particle trajectory and using,

$$\frac{1}{2}\nabla_x \|\nabla_x S^t(x)\|^2 = \frac{1}{2}\nabla_x \|v^t(x)\|^2 = v^t(x)\nabla_x v^t(x) \approx \nu_n^{s_k} \frac{\nu_n^{s_{k+1}} - \nu_n^{s_k}}{\tau}. \tag{50}$$

Crucially, the resulting discrete equations (8) depend exclusively on the growth network z_θ: given the current estimate of z_θ, they can be integrated forward along the geodesic to obtain the contact Wasserstein metric $g^{\mathcal{C}} = \sqrt{\mathcal{K}}\, g^{\mathcal{W}_2}$. This metric then governs the correction of the geodesic estimate via the transport network v_θ, through the arc-length minimization objective (7c). Consequently, z_θ and v_θ are trained in a mutually consistent manner: z_θ determines the geometry against which v_θ is optimized, while v_θ updates the particle trajectories along which z_θ is subsequently re-evaluated, tightening the contact Hamiltonian constraints at each training step. The parameter w_z must be negative to ensure that $z_n^{s_k}$ decays toward zero over the optimization interval $[t_0, t_M]$, thereby regularizing the learned initial condition $z_\theta(x^{s_0}, s_0)$ and reducing its magnitude irrespective of the sign of the growth rate. This mechanism governs the temporal evolution of $z_\theta(x^{s_k}, s_k)$ without imposing spatial constraints. Consequently, since the growth rate is only relevant where the distribution has support, the optimizer can initialize z_θ so that, under the induced dynamics, it attains suitable values when the distribution reaches those regions. The discrete geodesic $\{(x_n^{s_k}, w_n^{s_k})\}_{k=0}^K$ is computed via explicit Euler integration of the learned vector fields over K uniform steps of step size $\tau = (s_1 - s_0)/K$. The full particle trajectory $\{x_n^{s_k}\}_{k=0}^K$ and weight trajectory $\{w_n^{s_k}\}_{k=0}^K$ are stored at every integration step and used to evaluate the loss (7a). At each training iteration, the optimality variables $\{(\nu_n^{s_k}, z_n^{s_k})\}_{k=0}^K$ are computed from the current growth network z_θ by forward integration of the discretized contact Hamiltonian dynamics (8). The variables are initialized as $\nu_n^{s_0} = 1$ and $S_n^{s_0} = 0$ for all particles.

E.3 Training

Optimization. The training procedure is summarized in Algorithm 1. All models are trained using the Adam optimizer with weight decay 10^{-4} and automatic mixed precision enabled for memory efficiency on GPU. Gradient norms are clipped to a maximum of 1.0 at each update step to stabilize training. At each iteration, a batch of $B \times N_b$ particles is pre-sampled from the full dataset at the start of each epoch and partitioned into N_b mini-batches of size B, each processed as an independent

gradient step. All particles are initialized with unit mass $w_n^{s_0} = 1$, and the relative total mass at each observed marginal ρ^m is computed from the empirical cell counts, providing the reference mass targets for the feasibility loss.

Loss function. The total training objective combines an optimality term and a feasibility term (7a), where w_{opt} and w_{fea} are scalar weights balancing the two objectives (see Table 11). The optimality term enforces minimization of the contact Wasserstein arc-length. It is computed as the sum of a geodesic energy term and a mass regularity term as shown in equation (7c),

$$\ell_{\text{opt}} = \sum_{k=0}^{K-1} \sum_{n=0}^{N-1} \left((\nu_n^k)^2 w_n^{s_k} \, \|x_n^{s_{k+1}} - x_n^{s_k}\|^2 + \|z_n^{s_{k+1}} - z_\theta(x_n^{s_k}, s_{k+1})\|^2 \right).$$

The geodesic energy $(\nu_n^k)^2 w_n^{s_k} \, \|x_n^{s_{k+1}} - x_n^{s_k}\|^2$ computed as $(\overline{\nu}_n^{s_k})^2 w_n^{s_k} \, (\overline{d}_n^{s_k})^2$, with $\overline{\nu}_n^{s_k}$ and $\overline{d}_n^{s_k}$ denoting the column-normalized scalar speeds and squared particle displacements at step k, respectively. The column-wise normalization

$$\bar{u}_k = u_k \cdot \frac{K}{\sum_{k'} u_{k'} + \varepsilon} \tag{51}$$

is applied to both $\nu_n^{s_k}$ and $d_n^{s_k}$ before multiplication, ensuring that the energy is scale-invariant across integration steps and robust to vanishing gradients. The product $(\overline{\nu}_n^{s_k})^2 w_n^{s_k} \, (\overline{d}_n^{s_k})^2$ implements the conformal scaling of the displacement by the contact Wasserstein metric $g^{\mathcal{C}} = \sqrt{\mathcal{K}} \, g^{\mathcal{W}_2}$: steps with higher kinetic energy (large $\nu_n^{s_k}$) are penalized more heavily, aligning the learned velocity magnitude with the optimality condition for the given growth rate z_θ.

The feasibility term enforces consistency of the predicted trajectory with the observed marginals $\{\rho^m\}_{m=0}^M$ at the designated integration steps $\{s_{k_m}\}$. The loss (7b) comprises two components,

$$\ell_{\text{fea}} = \sum_{m=1}^{M} \left(d^{\mathcal{W}_2}(w^{s_m}, x^{s_m}, w^m, x^m) + w_{\text{mass}} \|W^{s_m} - W^m\|^2 \right).$$

where the weighted Wasserstein distance $d^{\mathcal{W}_2}$ is practically computed as the Sinkhorn approximation between the predicted weighted ensemble $\{(x_n^{s_{k_m}}, \tilde{w}_n^{s_{k_m}})\}$ and the reference marginal ρ^m, computed via the `geomloss` library (Feydy et al., 2019). Here $\tilde{w}_n^{s_{k_m}} = w_n^{s_{k_m}} / \sum_n w_n^{s_{k_m}}$ denotes the normalized predicted weights used as a discrete probability measure. The mass term $\|W^{s_m} - W^m\|^2$ penalizes the deviation of the predicted total mass W^{s_m} from the reference mass W^m. At the integration steps corresponding to observed marginals, the particle positions are detached from the computational graph, so that the geodesic is constrained to interpolate the observed marginals as fixed waypoints without back-propagating through them.

This loss formulation is consistent with our uOT baselines (Zhang et al., 2025; Sun et al., 2025), which also combine forward simulation with feasibility and optimality objectives. The key difference lies in the enforcement of the optimality conditions. Existing methods regularize the learned dynamics by minimizing Monte Carlo estimates of the Hamilton-Jacobi-Bellman residual, introducing additional objectives whose optimization may become unstable (Buzun et al., 2026). In contrast, our formulation discretizes and simulates the optimality system itself. Since the learned variables have a direct contact-geometric interpretation, they can be evolved jointly with the transport dynamics and their corresponding action can be evaluated directly along the trajectory. As a result, the loss retains the variational interpretation of the continuous Contact-uOT problem while avoiding residual minimization. The empirical consequences of this design choice are discussed in the following stability analysis.

E.4 Stability Considerations

The CuOT solver discretizes the geodesic flow $\{(x_n^{s_k}, w_n^{s_k})\}_{k=0}^K$ on the Wasserstein manifold $\mathcal{P}^+(\mathcal{M})$ and can be interpreted as a shooting direct method (Betts, 1998) for the optimal control problem of finding parameters θ^* minimizing the loss $\ell = w_{\text{opt}} \, \ell_{\text{opt}} + w_{\text{fea}} \, \ell_{\text{fea}}$ (7a). We show that each gradient update does not destabilize training and converges toward a local minimum.

Proposition 2. *The gradient $\nabla_\theta \ell_{\text{fea}}$ is Lipschitz-continuous in θ.*

Proof. The feasibility loss ℓ_{fea} comprises two terms,

(1) Wasserstein term. The weighted Wasserstein distance $d^{\mathcal{W}_2}$ is computed via the Sinkhorn approximation, whose solution corresponds to the optimal transport plan $\pi_{il}^m(\theta)$ between the predicted ensemble $\{(x_n^{s_{km}}, w_n^{s_{km}})\}$ and the reference marginal $\{(x_n^m, w_n^m)\}$. Since $\pi_{il}^m(\theta)$ solves a linear program with feasibility constraint $\sum_{il} \pi_{il}^m = 1$, and its dependence on θ enters through the predicted positions and weights, which are Lipschitz-continuous functions of θ by construction, the transport plan satisfies,

$$\|\pi_{il}^m(\theta') - \pi_{il}^m(\theta)\| \leq L_\pi \|\theta' - \theta\|, \tag{52}$$

for a scalar constant $L_\pi > 0$. Consequently, $\nabla_\theta \pi_{il}^m(\theta)$ is Lipschitz-continuous.

(2) Mass term. The total mass $W^{s_m} = \sum_n w_n^{s_m}$ depends nonlinearly on θ through the growth network z_θ, so the map $\theta \mapsto W^{s_m}$ is in general non-convex. However, the squared mass deviation $\|W^{s_m} - W^m\|^2$ is locally convex in a neighborhood of any local minimizer θ^*, since the Hessian $\nabla_\theta^2 \|W^{s_m} - W^m\|^2 \succeq 0$ in that neighborhood. Therefore $\nabla_\theta \|W^{s_m} - W^m\|^2$ is locally Lipschitz-continuous.

Combining (1) and (2), $\nabla_\theta \ell_{\text{fea}}$ is locally Lipschitz-continuous in θ. □

Proposition 3. *The gradient $\nabla_\theta \ell_{\text{opt}}$ is locally Lipschitz-continuous in θ under mild smoothness assumptions on the learned fields.*

Proof. We show that $\nabla_\theta \ell_{\text{opt}}$ is locally Lipschitz continuous by establishing this property for the dynamical variables $\nu_n^{s_k}$, $S_n^{s_k}$, and $z_n^{s_k}$ via induction on k.

The initial conditions $\nu_n^{s_0}$, $S_n^{s_0}$, and $z_n^{s_0}$ are fixed and independent of θ. Consequently, their gradients with respect to θ are identically zero and trivially locally Lipschitz. The growth network $z_\theta(x_n^{s_k}, s_k)$ employs smooth activations, therefore, both z_θ and $\nabla_\theta z_\theta$ are locally Lipschitz. The update equations (8) are rational functions of z_θ and the step-k variables. Since rational operations preserve local Lipschitz continuity, the gradients of the updated variables $(\nu_n^{s_1}, S_n^{s_1}, z_n^{s_1})$ are also locally Lipschitz. Via induction, the gradients of $(\nu_n^{s_k}, S_n^{s_k}, z_n^{s_k})$ are locally Lipschitz for all k.

The loss ℓ_{opt} is a finite sum of products involving these dynamical variables and the squared displacements $\|x_n^{s_{k+1}} - x_n^{s_k}\|^2$. The particle positions depend on θ through the velocity field v_θ, which is locally Lipschitz. Since finite sums and products of locally Lipschitz functions remain locally Lipschitz, it follows that $\nabla_\theta \ell_{\text{opt}}$ is locally Lipschitz continuous. □

The parameter update rule

$$\theta_{(j+1)} = \theta(j) - \alpha \nabla_\theta \ell(\theta_{(j)}), \tag{53}$$

with step size $\alpha > 0$, is a gradient descent step on the total loss ℓ. Since $\nabla_\theta \ell = w_{\text{opt}} \nabla_\theta \ell_{\text{opt}} + w_{\text{fea}} \nabla_\theta \ell_{\text{fea}}$ is locally Lipschitz-continuous by Propositions 1 and 2, and the loss is bounded below by zero, the gradient descent iterates $\{\theta_{(j)}\}_{j \geq 0}$ satisfy the sufficient decrease condition

$$\ell(\theta_{(j+1)}) \leq \ell(\theta_{(j)}) - \frac{\alpha}{2} \left\|\nabla_\theta \ell(\theta_{(j)})\right\|^2, \tag{54}$$

for sufficiently small α. Each update therefore strictly decreases the loss unless $\nabla_\theta \ell = 0$, guaranteeing convergence to a stationary point θ^*. Geometrically, θ^* corresponds to a local coordinate chart of the tangent bundle $\mathcal{TP}^+(\mathcal{M})$ that is consistent with the observed marginal constraints $\{\rho^m\}_{m=0}^M$.

Stability of shooting versus HJB residual minimization. An alternative approach to solving the optimal control problem (3) is to directly parameterize the dual potential (value function) $S_\theta(x_n^{s_k}, s_k)$ with a neural network and minimize the residual of the associated Hamilton-Jacobi-Bellman (HJB) equation (5b), as explored in Neklyudov et al. (2023a,b); Sun et al. (2025); Buzun et al. (2026). This strategy requires to learning the optimality conditions, rather than enforcing them directly, with important implications for both stability and numerical tractability.

In our formulation, the discretized optimality conditions (8) are enforced explicitly along predicted trajectories, reducing the problem to evaluating scalar quantities along sample paths $(x_n^{s_{km}}, w_n^{s_{km}})$ that are already available. As a result, training only requires backpropagation through $\nabla_\theta z_\theta$, leading to a comparatively well-conditioned optimization procedure. In contrast, minimizing the HJB residual necessitates differentiating through mixed second-order terms $\nabla_{x\theta}^2 S_\theta$, which are notoriously ill-conditioned and often induce vanishing or exploding gradients (Han et al., 2018; Rackauckas et al., 2020). Moreover, the HJB residual admits a constant degenerate solution $S_\theta(x_n^{s_k}, s_k) = C$ that trivially minimize the loss while encoding no transport information, thereby risking collapse

of the optimization (Wang et al., 2022). In this setting, methods that derive the drift solely from $\nabla_x S_\theta$ tend to produce noisy approximations (Neklyudov et al., 2023b; Sun et al., 2025; Buzun et al., 2026). When no additional optimal transport solver is used to anchor the learned dynamics to the prescribed marginals, the problem further inherits the limitations of collocation-based schemes, including the curse of dimensionality and the need for extensive domain sampling (Neklyudov et al., 2023b). These issues are exacerbated by the intrinsic regularity properties of optimal transport value functions, which frequently develop non-differentiable shocks. Accurately capturing such viscosity solutions is challenging for standard neural network parameterizations and typically requires substantial regularization (Han et al., 2018; Rackauckas et al., 2020).

Taken together, these considerations highlight a key advantage of the proposed shooting formulation: by enforcing optimality conditions directly along trajectories, it avoids both the degeneracies and the numerical instabilities inherent to HJB residual minimization.

E.5 Computational Complexity and Practical Considerations

At each iteration of the geodesic optimization procedure (Alg. 1), we sample N points $x_n^{s_0} \in \mathcal{M}$ of dimension D from the initial marginal ρ^0. We then compute the discrete geodesic $\{(x_n^{s_k}, w_n^{s_k})\}$ using equations (6), along with the optimality variables $\{(\nu_n^{s_k}, S_n^{s_k}, z_n^{s_k})\}$ via equations (8). Each integration step requires 7 evaluations of the neural networks v_θ and z_θ, implemented as MLPs with L layers and hidden width W. For K integration steps, the computational cost of this forward pass scales as $\mathcal{O}(7\,N\,D\,K\,L\,W)$. The loss function ℓ in equation (7a) requires M evaluations of the Wasserstein distance between sample batches. We employ the `geomloss` library (Feydy et al., 2019), which uses the Sinkhorn algorithm with time complexity $\mathcal{O}(I_{\text{sh}} M D N \log N)$, where I_{sh} is the number of Sinkhorn iterations until convergence (Feydy et al., 2019). Combining both contributions, and noting that in practice $N \gg D, K, L, W$, the overall time complexity becomes,

$$\mathcal{O}(7\,N\,D\,K\,L\,W) + \mathcal{O}(I_{\text{sh}} D N \log N) \approx \mathcal{O}(I_{\text{sh}} D N \log N). \tag{55}$$

Overall, the method exhibits favorable scaling properties: the complexity grows linearly with the data dimension D and nearly linearly with the batch size N. Additionally, it depends only weakly on the number of marginals and avoids costly outer iterative procedures required by alternative approaches.

E.6 Geodesic Parameterization: Finer discretization in high-variability regions

The result of integration equation (6), $(x_n^{s_{k_m}}, w_n^{s_{k_m}})$, discretizes the geodesic ρ^s on $\mathcal{P}^+(\mathcal{M})$ into a piecewise-linear path segmented at nodes $\{\rho_\theta^{s_k}\}_{k=0}^K$, thereby defining a specific sampling of points along the geodesic. By default, minimizing the curve energy induced by the Wasserstein metric $g^{\mathcal{W}_2}$,

$$d^2_{g^{\mathcal{W}_2}}(\rho^0, \rho^M) = \sum_{k=1}^{K} d^2_{\mathcal{W}_2}(\rho_\theta^{s_k}, \rho_\theta^{s_{k-1}}), \tag{56}$$

corresponds to a uniform arc-length discretization s, where each segment has approximately the same length, i.e., $d^2_{\mathcal{W}_2}(\rho_\theta^{s_k}, \rho_\theta^{s_{k-1}}) \approx C$. The relationship between the arc-length discretization s and the time discretization t, which describes the progression along the geodesic with respect to the time variable, is given by $ds = \sqrt{\mathcal{K}}\,dt$, where $\mathcal{K}$ denotes the kinetic energy of the system whose dynamics is interpreted as a geodesic. In the energy-conserving case, $\mathcal{K}$ remains constant, and the time and arc-length discretizations differ only by a constant rescaling of the geodesic length. Nevertheless, the discretization nodes $\{\rho_\theta^{s_k}\}_{k=0}^K$ remain unchanged, uniformly distributed in space and in time. In contrast, when modeling an energy-varying system whose associated trajectory projects onto a geodesic, the curve energy induced by the contact metric $g^{\mathcal{C}}$ takes the form,

$$d^2_{g^{\mathcal{C}}}(\rho^0, \rho^M) = \sum_{k=1}^{K} \mathcal{K}^{s_k}\, d^2_{\mathcal{W}_2}(\rho_\theta^{s_k}, \rho_\theta^{s_{k-1}}). \tag{57}$$

This optimization enforces an approximately uniform geodesic parameterization under $g^{\mathcal{C}}$, resulting in a non-uniform node spacing Δs on the Wasserstein manifold, as $\Delta s = d^2_{\mathcal{W}_2}(\rho_\theta^{s_k}, \rho_\theta^{s_{k-1}}) \approx \frac{C}{\mathcal{K}^{s_k}}$. As follows from equations (5), the scaling term $\mathcal{K}$ increases with the growth rate z^t. Consequently, regions with higher density concentration induce a larger effective metric. This, in turn, leads to a geodesic discretization that allocates more nodes in these areas, effectively assigning finer resolution to regions of higher variability and complexity.

Table 10: Model size and training time across methods and experimental settings. Memory is reported in MB and time in minutes.

Dataset / Autoencoder	OTCFM	TIGON	DeepRUOT	VarRUOT	**CuOT (Ours)**
Model Size (million of parameters)					
Simulation (Zhang et al., 2025)	0.2	0.2	0.2	0.5	0.5
Mouse (Weinreb et al., 2020)	0.2	0.2	0.2	0.5	0.8
EMT (Cook and Vanderhyden, 2020)	0.2	0.2	0.2	0.5	0.2
Pancreatic (Veres et al., 2019)	0.5	0.5	0.5	0.8	1
Zebrafish (Farrell et al., 2018)	1	1	1	0.8	1
FFHQ (Karras et al., 2019)					
ALAE encoder (Pidhorskyi et al., 2020)	12	1	3	1	3
e4e encoder (Richardson et al., 2021)	18	3	14	5	15
UCF101 (Soomro et al., 2012)					
DC encoder (Chen et al., 2025)	18	3	14	5	18
TiTok encoder (Yu et al., 2024)	18	3	14	5	18
Training Time (minutes)					
Simulation (Zhang et al., 2025)	10	300	5	15	3
Mouse (Weinreb et al., 2020)	10	300	5	15	3
EMT (Cook and Vanderhyden, 2020)	10	300	5	30	10
Pancreatic (Veres et al., 2019)	15	300	25	75	15
Zebrafish (Farrell et al., 2018)	20	300	30	80	15
FFHQ (Karras et al., 2019)					
ALAE encoder (Pidhorskyi et al., 2020)	20	400	60	115	25
e4e encoder (Richardson et al., 2021)	60	300	120	300	300
UCF101 (Soomro et al., 2012)					
DC encoder (Chen et al., 2025)	60	400	60	240	200
TiTok encoder (Yu et al., 2024)	90	300	120	300	300

Table 11: Parameter configurations for each experimental setting. K: number of integration steps; L: number of hidden layers in the shared trunk network used by v_θ and z_θ; F: number of sinusoidal frequencies; H: hidden feature dimension; B: batch size; N_b: number of batches per iteration; N_{iter}: number of training iterations; η: learning rate; ε: entropy level in Sinkhorn algorithm. $w_{(\cdot)}$: weighting coefficients. Additional details on these parameters are provided in Appendix E.

Dataset / Autoencoder	Dim.	K	L	F	H	B	N_b	N_{iter}	η	ε	w_z	w_m	w_g	w_{mass}
Simulation (Zhang et al., 2025)	2	20	3	3	512	512	1	2500	1e-5	0.005	-1	10	0.01	10
Mouse (Weinreb et al., 2020)	2	20	4	3	512	512	1	1500	1e-5	0.005	-1	10	0.01	1
EMT (Cook and Vanderhyden, 2020)	10	21	3	3	256	512	4	2500	1e-5	0.005	-1	10	0.3	1
Pancreatic (Veres et al., 2019)	30	21	4	4	512	1000	12	500	1e-5	0.005	-1	1	0.3	100
Zebrafish (Farrell et al., 2018)	50	20	4	3	512	1000	7	500	5e-4	0.005	-1	1	0.1	1
FFHQ (Karras et al., 2019)														
ALAE encoder (Pidhorskyi et al., 2020)	512	20	4	3	512	1000	5	10000	1e-5	0.005	-1	1	0.1	5
e4e encoder (Richardson et al., 2021)	9216	20	3	3	2048	1000	5	10000	3e-5	0.005	-1	1	0.01	500
UCF101 (Soomro et al., 2012)														
DC encoder (Chen et al., 2025)	2048	180	3	5	2048	101	1	15000	3e-5	0.005	-1	1	0.01	500
TiTok encoder (Yu et al., 2024)	d	180	3	5	2048	101	1	15000	3e-5	0.005	-1	1	0.01	500

F Extended Results

Details on the experimental settings adopted in each experiments are provided in Table 11. The synthetic ablation study (Appendix F.1) and the single-cell experiments (Appendix F.2) are conducted on a local machine equipped with an NVIDIA GeForce RTX 5090 GPU (32 GB memory, CUDA 13.0, driver 580.126.09). Methods are evaluated over 20 runs with different random seeds. The dataset-pairing (Appendix F.3) and image-to-video generation experiments (Appendix F.4) are run on a shared LSF-managed GPU cluster using the `batch_a100` queue. Each run is allocated a single GPU in exclusive mode, together with 1 CPU core, 32 GB of system memory, and CUDA 12.6 support. Both Child $\to$ Adult and Adult $\to$ Child transports are evaluated over 10 independent retrainings each, while for image-to-video generation each method is retrained 10 times per action class.

F.1 Ablation Studies and Analysis of Mode Collapse

To isolate the individual sources of performance gains in the proposed CuOT framework, evaluate the impact of adaptive step-size selection, and quantitatively assess resilience to soft mode collapse, we perform a detailed empirical study on a synthetic dataset with known mode correspondences and mass dynamics.

The source distribution consists of three 2D Gaussian modes (variance $\sigma = 0.15$) centered at $[0.0, 1.0]$, $[-0.8, 0.5]$, and $[0.8, 0.5]$, each with 400 samples (total $N = 1200$). The target distribution comprises three distinct shapes centered at $[0.0, 1.4]$ (Swiss roll, $N = 600$), $[-1.2, 0.7]$ (concentric rings, $N = 600$), and $[1.2, 0.7]$ (pinwheel, $N = 200$). Under correct mode-to-mode matching, the first two source modes undergo mass expansion (requiring positive growth rates, $z^t > 0$), whereas the third mode undergoes mass decay (requiring a negative growth rate, $z^t < 0$). This setup intentionally misaligns the transport direction with the growth gradient for the third mode to directly target the failure modes of standard WFR-based formulations.

Performance is evaluated across global transport, mass conservation, and mode-level fidelity using, 2-Wasserstein ($\mathcal{W}_2$) and 1-Wasserstein ($\mathcal{W}_1$) distances between generated and ground-truth target distributions; absolute error in total reconstructed target mass; the fraction of source samples transported to incorrect target modes (Leakage); the chamfer distance from ground-truth target samples to generated samples (CD); and individual W_2 metrics for each of the three target modes $[M_1, M_2, M_3]$ to quantify minority mode recovery. The results of the comparison are resumed in Table 2.

We compare the complete CuOT model against five baselines and structural variants:

1. CuOT (Complete): The proposed full contact geometry formulation with spatial-temporal growth dynamics $z_\theta(x^{s_k}, s_k)$ and adaptive step-size allocation.
2. CuOT w/ Time-Only Growth: Replaces $z_\theta(x^{s_k}, s_k)$ with a spatially uniform, time-dependent growth rate parameter $z_\theta(s_k)$.
3. CuOT w/ WFR Geometry: Replaces the Herglotz variational uOT problem (3) with the WFR uOT (10) and applies the geodesic-based discretization described in Section 5. The discrete FP equation (6) and the feasibility loss term (7b) remain unchanged, but the discrete optimality conditions (8) now discretize the WFR conditions (45), and the optimality loss optimizes the WFR geodesic objective coupled with growth regularization:
$$\ell_{\text{opt}} = \frac{1}{2}(v_n^{s_k})^2 w_n^{s_k} d^{\mathcal{W}_2} + \|z_n^{s_k} - z_\theta(x_n^{s_k}, s_k)\|^2.$$
4. CuOT w/o Adaptive Step: Retains the growth regularization term from the contact formulation but uses standard 2-Wasserstein geodesic minimization for trajectory parameterization:
$$\ell_{\text{opt}} = (d^{\mathcal{W}_2})^2 + \|z_n^{s_k} - z_\theta(x_n^{s_k}, s_k)\|^2.$$
This isolates the contact-metric step adaptation but preserves mass variation dynamics.
5. VarRUOT: Standard variational WFR baseline.
6. OTCFM: OT-based baseline assisted by an independent global mass network.

Comparing CuOT (Complete) with CuOT w/o Adaptive Step directly isolates the impact of dynamic temporal resolution allocation. Although both variants achieve near-identical mass reconstruction errors (0.0006 vs. 0.0009) and mode leakage (0.2%), removing the adaptive step degrades overall transport accuracy ($\mathcal{W}_2$: $0.10 \to 0.13$; $\mathcal{W}_1$: $0.08 \to 0.10$). Inspecting per-mode target errors reveals that these performance gains are concentrated in regions undergoing mass expansion: CuOT (Complete) yields per-mode $\mathcal{W}_2$ errors of $[0.09, 0.10, 0.13]$, compared to $[0.13, 0.14, 0.13]$ for the non-adaptive variant. This difference arises because the non-adaptive variant distributes discrete trajectory points uniformly due to the standard 2-Wasserstein geodesic objective. In contrast, the contact metric dynamically reduces step sizes in regions with positive local mass growth ($z_\theta > 0$), allocating higher temporal resolution to high-density target regions (Modes 1 and 2). Shrinking regions (Mode 3) retain standard resolution, maintaining consistent accuracy ($\mathcal{W}_2 = 0.13$).

WFR-based methods suffer from mode leakage, improperly assigning samples to distant target modes when transport directions conflict with local growth dynamics. This behavior is demonstrated in Table 2, where VarRUOT and CuOT w/ WFR Geometry exhibit 6.5% and 10.4% mode leakage,

respectively. Examining per-mode $\mathcal{W}_2$ errors for VarRUOT ($[0.15, 0.16, 0.20]$) reveals progressive accuracy degradation on the third mode, which requires negative growth, because samples are instead transported toward other directions. As growth regularization vanishes, leakage decreases to 0, but WFR geometry induces positive growth for all modes and and high CD (> 0.3), causing samples to cover only partial modes and leading to mode collapse. The time-only growth variant cannot capture spatially varying dynamics, yielding the worst performance, as the target requires simultaneous growth and decay across different regions. The OTCFM method with the global mass network successfully minimizes global mass error (0.0001), but as it keeps sample weights constant it fails to prevent misrouting, resulting in the highest mode leakage ($17.7\ \%$).

F.2 Single-Cell Dynamics Prediction

The benchmarks considered in the experiments range from a controlled synthetic setting to four real-world datasets of increasing dimensionality and complexity. All datasets share the common structure of temporal snapshots of single-cell gene expression data: each marginal ρ^m is a collection of cells measured at a given time point, represented as points in a high-dimensional feature space, with the total cell count varying across time points due to proliferation and death. As summarized in the dataset-specific tables below, the Wasserstein distance between successive marginals, which the learned transport network v_θ must replicate, is not proportional to the growth rate gradient $\nabla_x z^t$, systematically violating the geometric coupling assumption underlying WFR-based reconstruction methods.

Three-gene simulation (Zhang et al., 2025) (2D). This synthetic dataset simulates stochastic gene expression dynamics for a system of three genes incorporating self-activation, mutual inhibition, and external activation. The two-dimensional representation corresponds to the two leading principal genes of the gene expression trajectory. The initial distribution splits into two clusters, one of which remains quiescent while the other evolves through the state space, with a monotonically decreasing distance between successive marginals and an approximately constant growth rate gradient. The dataset comprises $M = 4$ observed marginals with three intermediate marginals included in training.

Table 12: Details on the three-gene simulated dataset (Zhang et al., 2025).

Time	Mass	Growth Rate z	∇z	N samples	$\mathcal{W}_2$ Distance	Distance %
0	1.000	10.5%	-	400	0.5185	33.9%
1	1.105	19.9%	9.4%	442	0.5117	33.5%
2	1.325	30.2%	10.3%	530	0.3193	20.9%
3	1.725	40.4%	10.2%	690	0.1783	11.7%
4	2.422	-	-	969	-	-

Mouse hematopoiesis (Weinreb et al., 2020) (2D). This dataset tracks blood stem cells in a mouse as they commit to one of two specialized blood cell lineages over time. The two-dimensional representation corresponds to the two principal axes of the differentiation process. The dataset comprises $M = 2$ observed marginals plus the initial condition, with a single intermediate marginal included in training.

Table 13: Details on the mouse blood hematopoiesis dataset (Weinreb et al., 2020).

Time	Mass	Growth Rate z	∇z	N samples	$\mathcal{W}_2$ Distance	Distance %
0	1.000	164.6%	-	1429	0.8041	61.11%
1	2.646	53.1%	-111.5%	3781	0.5117	38.89%
2	4.050	-	-	5788	-	-

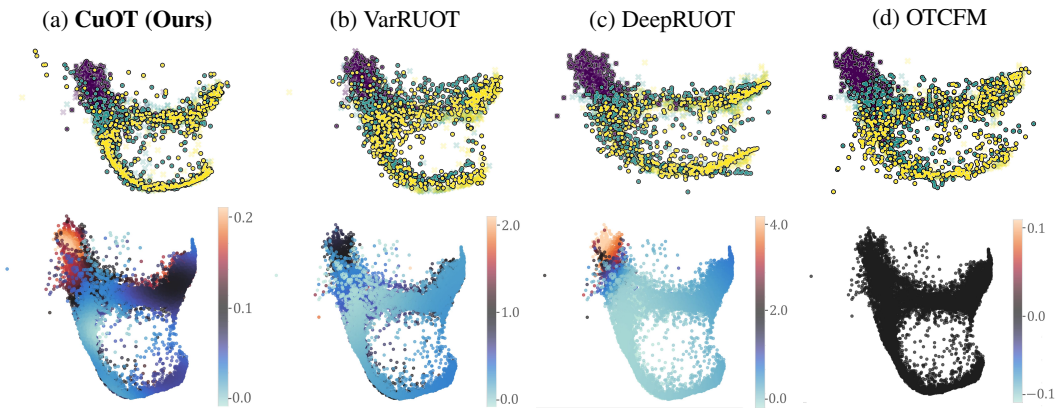

Figure 8: Qualitative analysis on the mouse blood hematopoiesis dataset (Weinreb et al., 2020). Predicted dynamics (t_0, t_1, t_2) (top) and growth rate (bottom).

Table 14: Quantitative analysis on the mouse blood hematopoiesis dataset (Weinreb et al., 2020).

	$t=1$			$t=2$		
Model	$\mathcal{W}_1$	$\mathcal{W}_2$	Mass	$\mathcal{W}_1$	$\mathcal{W}_2$	Mass
OTCFM (Tong et al., 2024a)	$0.1657_{\pm 0.0271}$	$0.1795_{\pm 0.0183}$	$2.7090_{\pm 0.0000}$	$0.1670_{\pm 0.0268}$	$0.1820_{\pm 0.0208}$	$9.3048_{\pm 0.0000}$
TIGON (Sha et al., 2024)	$0.3392_{\pm 0.0294}$	$0.3975_{\pm 0.0538}$	$1.3125_{\pm 0.3903}$	$0.4530_{\pm 0.0544}$	$0.5474_{\pm 0.0364}$	$5.1929_{\pm 0.7556}$
DeepRUOT (Zhang et al., 2025)	$0.1254_{\pm 0.0063}$	$0.1442_{\pm 0.0052}$	$0.1391_{\pm 0.0336}$	$0.1137_{\pm 0.0074}$	$0.1292_{\pm 0.0064}$	$0.6741_{\pm 0.1275}$
VarRUOT (Sun et al., 2025)	$0.1211_{\pm 0.0048}$	$0.1399_{\pm 0.0051}$	$0.0048_{\pm 0.0010}$	$0.1226_{\pm 0.0080}$	$0.1349_{\pm 0.0067}$	$0.0089_{\pm 0.0055}$
CuOT (Ours)	$\mathbf{0.0745}_{\pm 0.0095}$	$\mathbf{0.0765}_{\pm 0.0071}$	$\mathbf{0.0003}_{\pm 0.0003}$	$\mathbf{0.0561}_{\pm 0.0100}$	$\mathbf{0.0550}_{\pm 0.0071}$	$\mathbf{0.0019}_{\pm 0.0021}$

Epithelial–mesenchymal transition (EMT) (Cook and Vanderhyden, 2020) (10D). This dataset captures the process by which cancer cells switch from a stationary, tightly-packed epithelial state to a mobile, loosely-connected mesenchymal state, a key step in tumor metastasis and invasion. The 10-dimensional feature space corresponds to the leading principal components of the gene expression profile. The transition induces a pronounced shift in the distribution support across time points, accompanied by moderate mass variation, and the dataset includes $M = 3$ observed marginals with two intermediate marginals used in training.

Table 15: Details on the Emt dataset (Cook and Vanderhyden, 2020).

Time	Mass	Growth Rate z	∇z	N samples	$\mathcal{W}_2$ Distance	Distance %
0	1.000	53.4%	-	577	3.8210	61.19%
1	1.534	-11.0%	-64.4%	885	1.2108	19.39%
2	1.366	12.0%	23.0%	788	1.2122	19.41%
3	1.530	-	-	883	-	-

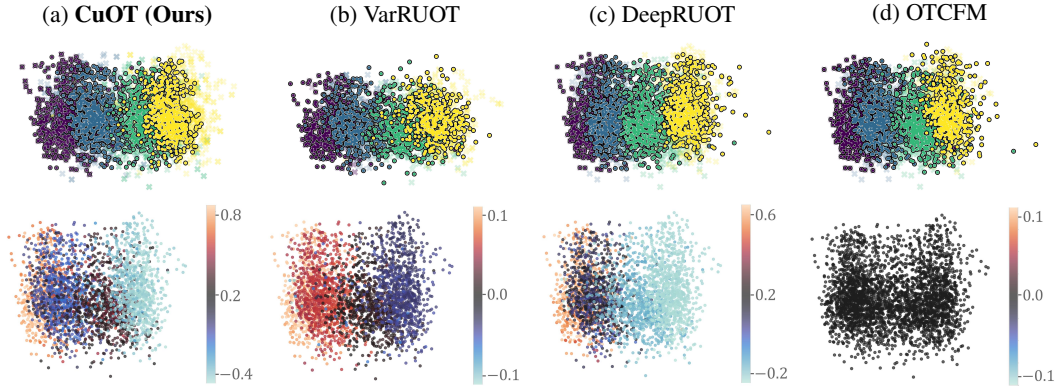

Figure 9: Qualitative analysis on the EMT dataset (Cook and Vanderhyden, 2020). Predicted dynamics (t_0, t_1, t_2, t_3) (top) and growth rate (bottom).

Table 16: Quantitative analysis on the EMT dataset (Cook and Vanderhyden, 2020).

Model	$t=1$			$t=2$			$t=3$		
	$\mathcal{W}_1$	$\mathcal{W}_2$	Mass	$\mathcal{W}_1$	$\mathcal{W}_2$	Mass	$\mathcal{W}_1$	$\mathcal{W}_2$	Mass
OTCFM (Tong et al., 2024a)	$0.1831_{\pm.0006}$	$0.2502_{\pm.0007}$	$0.2849_{\pm.0000}$	$0.1948_{\pm.0007}$	$0.2662_{\pm.0007}$	$0.1337_{\pm.0000}$	$0.2026_{\pm.0009}$	$0.2744_{\pm.0008}$	$0.2812_{\pm.0000}$
TIGON (Sha et al., 2024)	$0.2705_{\pm.0008}$	$0.3209_{\pm.0009}$	$0.2377_{\pm.0010}$	$0.3377_{\pm.0011}$	$0.3552_{\pm.0010}$	$0.2885_{\pm.010}$	$0.3751_{\pm.0012}$	$0.3935_{\pm.0011}$	$0.3126_{\pm.0012}$
DeepRUOT (Zhang et al., 2025)	$0.2664_{\pm.0023}$	$0.3654_{\pm.0022}$	$0.3632_{\pm.0029}$	$0.2367_{\pm.0024}$	$0.3260_{\pm.0025}$	$0.2777_{\pm.0032}$	$0.3901_{\pm.0028}$	$0.5513_{\pm.0030}$	$0.5533_{\pm.0034}$
VarRUOT (Sun et al., 2025)	$0.1846_{\pm.0018}$	$0.2533_{\pm.0021}$	$0.2234_{\pm.0016}$	$\mathbf{0.1901}_{\pm.0017}$	$0.2602_{\pm.0027}$	$0.0577_{\pm.0015}$	$0.1952_{\pm.0020}$	$0.2637_{\pm.0027}$	$0.1152_{\pm.0019}$
CuOT (Ours)	$\mathbf{0.1782}_{\pm.0016}$	$\mathbf{0.2424}_{\pm.0022}$	$\mathbf{0.0011}_{\pm.0005}$	$\mathbf{0.1889}_{\pm.0024}$	$\mathbf{0.2581}_{\pm.0033}$	$\mathbf{0.0017}_{\pm.0014}$	$\mathbf{0.1841}_{.0025}$	$\mathbf{0.2487}_{\pm.0031}$	$\mathbf{0.0015}_{\pm.0013}$

Pancreatic β-cell differentiation (Veres et al., 2019) (30D). This dataset describes the *in vitro*maturation of human stem cells into insulin-secreting pancreatic β-cells, a process of direct relevance to diabetes research. The 30-dimensional feature space is obtained via UMAP dimensionality reduction of the full gene expression profile. The dataset presents significant variability across time points, partly attributable to experimental sampling noise inherent to the *in vitro* protocol, and strong mass variation reflecting heterogeneous differentiation rates. It comprises $M = 7$ observed marginals with six intermediate marginals included in training, making it the most densely observed benchmark.

Table 17: Details on the pancreatic β-cell differentiation dataset (Veres et al., 2019).

Time	Mass	Growth Rate z	∇z	N samples	$\mathcal{W}_2$ Distance	Distance %
0	1.000	-0.2%	-	6190	5.8191	0.709%
1	0.998	-42.4%	-42.2%	6177	181.0527	22.062%
2	0.575	250.6%	293.0%	3562	341.0221	41.555%
3	2.016	-48.1%	-298.7%	12477	141.7822	17.277%
4	1.047	-1.1%	47.0%	6480	4.8370	0.589%
5	1.035	-6.2%	-5.1%	6404	17.1638	2.091%
6	0.971	-33.9%	-27.7%	6013	128.9814	15.717%
7	0.642	-	-	3971	-	-

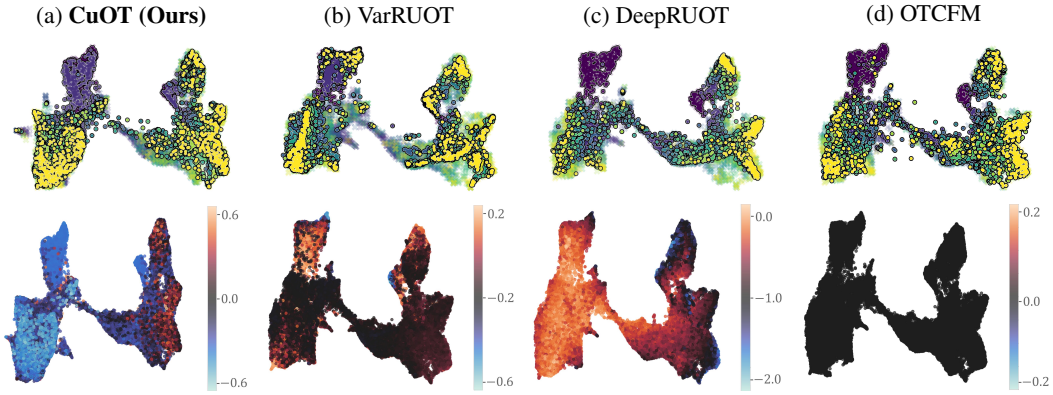

Figure 10: Qualitative analysis on the pancreatic β-cell differentiation dataset (Veres et al., 2019). Predicted dynamics $(t_0, t_1, t_2, t_3, t_4, t_5, t_6, t_7)$ (top) and growth rate (bottom).

Table 18: Quantitative analysis on the pancreatic β-cell differentiation dataset (Veres et al., 2019).

Method	$t = 1,2,3,4,5,6,7$ (average)		
	$\mathcal{W}_1$	$\mathcal{W}_2$	Mass
OTCFM (Tong et al., 2024a)	$7.3684_{\pm 1.3049}$	$5.5259_{\pm 1.0255}$	$0.1921_{\pm 0.3775}$
TIGON (Sha et al., 2024)	$127.3335_{\pm 39.4399}$	$16.6406_{\pm 2.4754}$	$1.0401_{\pm 0.8187}$
DeepRUOT (Zhang et al., 2025)	$11.2069_{\pm 13.3341}$	$5.1074_{\pm 5.3032}$	$0.5323_{\pm 0.6582}$
VarRUOT (Sun et al., 2025)	$10.2294_{\pm 12.2393}$	$4.6187_{\pm 4.8103}$	$0.1897_{\pm 0.3517}$
CuOT (Ours)	$\mathbf{4.1082}_{\pm \mathbf{5.6363}}$	$\mathbf{2.5572}_{\pm \mathbf{3.0217}}$	$\mathbf{0.0889}_{\pm \mathbf{0.1064}}$

Table 19: Quantitative analysis on the pancreatic β-cell differentiation dataset (Veres et al., 2019).

Model	Metric	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$
OTCFM (Tong et al., 2024a)	$\mathcal{W}_1$	$5.3684_{\pm.0916}$	$\mathbf{6.3219_{\pm.0936}}$	$\mathbf{6.7157_{\pm.1134}}$	$7.5628_{\pm.1133}$	$7.9461_{\pm.1123}$	$8.6920_{\pm.1120}$	$8.6920_{\pm.1152}$
	$\mathcal{W}_2$	$3.9929_{\pm.0172}$	$\mathbf{4.6971_{\pm.0513}}$	$\mathbf{4.9720_{\pm.0491}}$	$5.6603_{\pm.0408}$	$5.9798_{\pm.0436}$	$6.5607_{\pm.0343}$	$6.8187_{\pm.0478}$
	Mass	$\mathbf{0.0000_{\pm.0000}}$	$0.1802_{\pm.0000}$	$1.0316_{\pm.0000}$	$0.0022_{\pm.0000}$	$0.0012_{\pm.0000}$	$\mathbf{0.0008_{\pm.0000}}$	$0.1285_{\pm.0000}$
TIGON (Sha et al., 2024)	$\mathcal{W}_1$	$105.8188_{\pm.5371}$	$94.6576_{\pm.7118}$	$134.7391_{\pm.8179}$	$127.8763_{\pm.8631}$	$180.0941_{\pm.8922}$	$173.7854_{\pm.8194}$	$74.3633_{\pm.9463}$
	$\mathcal{W}_2$	$18.5314_{\pm.1283}$	$14.3638_{\pm.1343}$	$19.9549_{\pm.4356}$	$16.1346_{\pm.3465}$	$15.8342_{\pm.4329}$	$18.5540_{\pm.4133}$	$13.1114_{\pm.4520}$
	Mass	$0.0088_{\pm.0008}$	$0.46382_{\pm.0982}$	$0.5629_{\pm.0967}$	$0.6783_{\pm.00983}$	$1.4551_{\pm.0924}$	$1.9310_{\pm.0916}$	$2.1807_{\pm.1002}$
DeepRUOT (Zhang et al., 2025)	$\mathcal{W}_1$	$0.1757_{\pm.0001}$	$31.8596_{\pm.0320}$	$22.7228_{\pm.0525}$	$0.4565_{\pm.0011}$	$0.3750_{\pm.0030}$	$0.4328_{\pm.0005}$	$22.4263_{\pm.0545}$
	$\mathcal{W}_2$	$0.2583_{\pm.0001}$	$12.2500_{\pm.0122}$	$9.6641_{\pm.0229}$	$0.9371_{\pm.0033}$	$0.8706_{\pm.0062}$	$0.8383_{\pm.0043}$	$10.9334_{\pm.0110}$
	Mass	$0.0258_{\pm.0000}$	$\mathbf{0.0238_{\pm.0000}}$	$2.0108_{\pm.0015}$	$0.3471_{\pm.0016}$	$0.5010_{\pm.0029}$	$0.5659_{\pm.0037}$	$0.2518_{\pm.0027}$
VarRUOT (Sun et al., 2025)	$\mathcal{W}_1$	$0.2518_{\pm.0006}$	$30.3818_{\pm.4328}$	$21.6480_{\pm.1574}$	$0.3557_{\pm.0043}$	$0.2827_{\pm.0041}$	$0.2764_{\pm.0026}$	$18.4096_{\pm.3291}$
	$\mathcal{W}_2$	$0.3479_{\pm.0006}$	$11.6851_{\pm.1553}$	$9.0374_{\pm.0239}$	$0.8773_{\pm.0134}$	$0.6406_{\pm.0144}$	$0.5941_{\pm.0087}$	$9.1489_{\pm.0627}$
	Mass	$\mathbf{0.0000_{\pm.0000}}$	$0.1842_{\pm.0035}$	$1.0129_{\pm.0060}$	$\mathbf{0.0014_{\pm.0005}}$	$0.0009_{\pm.0007}$	$0.0014_{\pm.0017}$	$0.1270_{\pm.0267}$
CuOT (Ours)	$\mathcal{W}_1$	$\mathbf{0.1225_{\pm.0023}}$	$13.6947_{\pm.0247}$	$9.9932_{\pm.0361}$	$\mathbf{0.2247_{\pm.0055}}$	$\mathbf{0.2158_{\pm.0050}}$	$\mathbf{0.1552_{\pm.0042}}$	$\mathbf{8.3567_{\pm.0368}}$
	$\mathcal{W}_2$	$\mathbf{0.1196_{\pm.0011}}$	$7.1560_{\pm.0187}$	$5.4073_{\pm.0239}$	$\mathbf{0.6094_{\pm.0025}}$	$\mathbf{0.4372_{\pm.0028}}$	$\mathbf{0.5715_{\pm.0037}}$	$\mathbf{5.9001_{\pm.0102}}$
	Mass	$0.0148_{\pm.0009}$	$0.1644_{\pm.0084}$	$\mathbf{0.2840_{\pm.0088}}$	$0.1172_{\pm.0087}$	$\mathbf{0.0000_{\pm.0001}}$	$0.0424_{\pm.0032}$	$\mathbf{0.0000_{\pm.0001}}$

Zebrafish embryogenesis (Farrell et al., 2018) (50D). This dataset captures the early development of a zebrafish embryo, during which a population of initially undifferentiated cells diversifies into 25 distinct specialized cell types over approximately 9 hours. The 50-dimensional feature space is obtained via UMAP reduction of the full transcriptome. This is the highest-dimensional single-cell benchmark, characterized by rapid and highly heterogeneous proliferation across branching cell lineages, and presents the most challenging mass variation profile. The dataset comprises $M = 6$ observed marginals with five intermediate marginals included in training.

Table 20: Details on the zebrafish embryogenesis dataset Farrell et al. (2018).

Time	Mass	Growth Rate z	∇z	N samples	$\mathcal{W}_2$ Distance	Distance %
0	1.000	82.3%	-	3136	43944.6406	41.364%
1	1.823	-10.3%	-92.6%	5716	1106.7998	6.564%
2	1.635	20.5%	30.8%	5127	3699.7798	12.002%
3	1.970	-11.9%	-32.4%	6178	2166.1155	9.183%
4	1.735	30.8%	42.7%	5442	8879.2041	18.593%
5	2.269	-15.4%	-46.2%	7114	3879.7812	12.290%
6	1.919	-	-	6018	-	-

Table 21: Quantitative analysis on the zebrafish embryogenesis dataset (Farrell et al., 2018).

Model	Metric	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
OTCFM (Tong et al., 2024a)	$\mathcal{W}_1$	$4.0287_{\pm.0151}$	$4.9408_{\pm.0165}$	$5.7038_{\pm.0535}$	$7.1590_{\pm.2742}$	$8.4766_{\pm.3243}$	$10.7044_{\pm.3525}$
	$\mathcal{W}_2$	$3.4801_{\pm.0105}$	$3.9062_{\pm.0057}$	$4.4139_{\pm.0640}$	$5.3939_{\pm.2031}$	$6.2450_{\pm.3327}$	$7.1066_{\pm.3944}$
	Mass	$0.6768_{\pm0.0000}$	$0.4030_{\pm.0000}$	$0.9409_{\pm.0000}$	$0.5407_{\pm.0000}$	$1.6090_{\pm.0000}$	$0.8445_{\pm.0000}$
TIGON (Sha et al., 2024)	$\mathcal{W}_1$	$6.8247_{\pm.2426}$	$6.6419_{\pm.3035}$	$7.7523_{\pm.4264}$	$8.4695_{\pm.4572}$	$9.1028_{\pm.4876}$	$10.7916_{\pm.4237}$
	$\mathcal{W}_2$	$5.5471_{\pm.1314}$	$5.3589_{\pm.1298}$	$5.9637_{\pm.4412}$	$7.1284_{\pm.3519}$	$9.8475_{\pm.4371}$	$9.5468_{\pm.4186}$
	Mass	$0.5733_{\pm.0134}$	$0.3638_{\pm.0257}$	$0.6629_{\pm.0380}$	$0.4783_{\pm.0192}$	$0.4551_{\pm.0202}$	$0.6310_{\pm.0418}$
DeepRUOT (Zhang et al., 2025)	$\mathcal{W}_1$	$3.7562_{\pm.0430}$	$3.9570_{\pm.1302}$	$4.2795_{\pm.01279}$	$4.9962_{\pm.1367}$	$5.7769_{\pm.1698}$	$6.9029_{\pm.1804}$
	$\mathcal{W}_2$	$3.5913_{\pm.0332}$	$3.5195_{\pm.0916}$	$3.8945_{\pm.1382}$	$4.3707_{\pm.1471}$	$4.9568_{\pm.1653}$	$6.2108_{\pm.1690}$
	Mass	$0.4842_{\pm.0131}$	$0.0930_{\pm.0103}$	$0.1680_{\pm.0127}$	$0.0078_{\pm.0020}$	$0.0189_{\pm.0094}$	$0.3214_{\pm.0452}$
VarRUOT (Sun et al., 2025)	$\mathcal{W}_1$	$4.6608_{\pm.0712}$	$4.4166_{\pm.2412}$	$4.6337_{\pm.3522}$	$5.1508_{\pm.3319}$	$5.7901_{\pm.7662}$	$6.8333_{\pm.6958}$
	$\mathcal{W}_2$	$4.4311_{\pm.0290}$	$4.0036_{\pm.3274}$	$4.3000_{\pm.3541}$	$4.6924_{\pm.3099}$	$5.1680_{\pm.3128}$	$6.2095_{\pm.2757}$
	Mass	$0.4744_{\pm.0114}$	$0.1152_{\pm.0189}$	$0.3929_{\pm.0234}$	$0.1639_{\pm.0340}$	$0.9363_{\pm.0394}$	$0.4324_{\pm.0496}$
CuOT (Ours)	$\mathcal{W}_1$	$\mathbf{3.1730_{\pm.0356}}$	$\mathbf{2.6959_{\pm.0828}}$	$\mathbf{2.8354_{\pm.1037}}$	$\mathbf{3.2731_{\pm.1002}}$	$\mathbf{3.4931_{\pm.1696}}$	$\mathbf{4.1428_{\pm.1790}}$
	$\mathcal{W}_2$	$\mathbf{3.0379_{\pm.0296}}$	$\mathbf{2.3572_{\pm.0653}}$	$\mathbf{2.8226_{\pm.1378}}$	$\mathbf{3.2909_{\pm.1579}}$	$\mathbf{3.08173_{\pm.1473}}$	$\mathbf{3.6236_{\pm.1342}}$
	Mass	$\mathbf{0.0345_{\pm.0201}}$	$\mathbf{0.0373_{\pm.0298}}$	$\mathbf{0.0168_{\pm.0257}}$	$\mathbf{0.0107_{\pm.0063}}$	$\mathbf{0.0257_{\pm.0174}}$	$\mathbf{0.0203_{\pm.0148}}$

F.3 Unmatched Datasets Pairing

The Flickr-Faces-HQ (FFHQ) dataset is a high-resolution (1024×1024) collection of human faces that exhibits a remarkably wide range of visual variations (Karras et al., 2019). We follow the experimental protocol adopted by Gushchin et al. (2024), using their publicly available datasets from `https://github.com/ngushchin/LightSB`. We split the dataset into training (first $60K$ images) and testing (last $10K$ images) subsets. Each subset is further partitioned into age groups, specifically *adults* and *children*, using the per-image age labels provided by the FFHQ-Features dataset (Karras et al., 2019), which supplies facial attribute annotations for the full FFHQ collection. These two groups correspond to the reference marginal distributions ρ^m, yielding an unbalanced, unpaired matching problem: the training split contains approximately $45K$ adult samples and $15K$ child samples. The goal of this experiment is to synthesize age-transformed images of the same individual. Specifically, given an adult image, the model generates a realistic depiction of that person as a child, and conversely, given a child image, it produces a plausible adult version of the same individual. To train our solvers using the MLP architectures common across all baseline methods, we operate within the latent space of pre-trained autoencoders. To evaluate sensitivity to the choice of latent representation, we consider two encoder backbones: (i) the ALAE encoder (Pidhorskyi et al., 2020), which compresses a 1024×1024 image into a 512-dimensional vector, and (ii) the e4e encoder (Richardson et al., 2021), which maps images into the StyleGAN latent space by generating a distinct 512-dimensional style code for each of the 18 generator layers. This 18-fold increase in feature dimensionality (9216 dimensions in total) necessitates scaling up the baseline model capacities accordingly. During inference, an input image is mapped to the latent space, transformed via the learned transport map, and reconstructed through the corresponding decoder to produce the age-transformed result. Full hyperparameter configurations for both setups are detailed in Table 11.

The experiment defines two complementary transport directions. The Adult $\rightarrow$ Child task requires accurate early-stage dynamics in a high-density source distribution to ensure correct transport to the right mode in the lower-density target mode. Because the adult distribution is roughly three times larger than the child distribution, correctly learning the initial transport direction is fundamental to retrieve the correct equivalent mode in the target. Conversely, the Child $\rightarrow$ Adult task demands high-fidelity reconstructions and correct coverage of the broader adult distribution support, as the model must generate plausible, diverse adult appearances from a compact source.

We evaluate all methods using FID, and the latent-space Wasserstein distances $\mathcal{W}_1$ and $\mathcal{W}_2$, together with a mass discrepancy term. FID is computed directly in pixel space using the standard InceptionV3 feature extractor (2048-dimensional pool layer), following the formulation,

$$\mathrm{FID} = \|\mu_r - \mu_g\|^2 + \mathrm{Tr}\Big(\Sigma_r + \Sigma_g - 2\left(\Sigma_r \Sigma_g\right)^{1/2}\Big), \tag{58}$$

where (μ_r, Σ_r) and (μ_g, Σ_g) are the mean and covariance of the InceptionV3 activations computed over the real and generated test images, respectively. The matrix square root is computed via Schur decomposition, with the real part taken whenever the result contains negligible imaginary components due to numerical precision.

Quantitative results for both transport directions are reported separately in Tables 22 and 23 for the ALAE latent space, and in Tables 24 and 25 for the e2e latent space. Table 6 of the main paper reports the combined average across directions for the ALAE latent space. In this setting, while balanced OT approaches yield larger latent distances from the target distribution than WFR-based uOT methods, they achieve the second-best FID score, which better measures mode coverage in the reconstructed images. As shown in Figs. 11 and 5, higher latent distances correlate with lower image quality. Nevertheless, balanced OT methods still correctly capture target modes and provide full distribution coverage, producing images that broadly reflect the intended age group. In contrast, WFR-based uOT methods frequently fail to complete the transport effectively: despite achieving lower latent distances, they exhibit significantly higher FID scores, indicating that mass-growth bias causes distributional collapse in pixel space rather than faithful age transformation. Due to contact geometry, CuOT achieves state-of-the-art performance across all metrics, with improved target coverage and sample quality in both directions. Qualitative examples of the full transport bridge computed by our CuOT method and each baseline are presented in Figures 12–15. These figures also illustrate how CuOT allocates more integration steps in high-density regions of the source distribution, following the finer geodesic discretization induced by the contact Wasserstein metric $g^C = \sqrt{K}\, g^{W_2}$ (see Appendix E.6). A compact side-by-side qualitative comparison at the endpoint is shown in Figures 11

and 5. In the higher-dimensional latent space produced by the e2e encoder, CuOT maintains its competitive advantage, exhibiting only a modest performance drop relative to the lower-dimensional setup. Among the baselines, OTCFM is the sole method that scales reasonably well, whereas other WFR-based uOT frameworks suffer severe performance degradation. These results demonstrate both the adaptability of our framework to distinct latent architectures and its superior scalability compared to state-of-the-art baselines. Figure 16 visualize the bridge decoded from the the e2e late space.

Table 22: Adult → Child. FFHQ dataset (Karras et al., 2019), ALAE encoder (Pidhorskyi et al., 2020).

Method	FID	$\mathcal{W}_1$	$\mathcal{W}_2$	Mass
OTCFM (Tong et al., 2024a)	$6.3081_{\pm 0.0910}$	$14.3921_{\pm 0.1162}$	$10.2690_{\pm 0.1007}$	$0.6066_{\pm 0.0000}$
TIGON (Sha et al., 2024)	$23.1358_{\pm 0.7232}$	$16.3305_{\pm 0.1795}$	$11.2321_{\pm 0.1468}$	$0.8012_{\pm 0.0572}$
DeepRUOT (Zhang et al., 2025)	$6.9551_{\pm 0.3167}$	$13.7117_{\pm 0.1228}$	$9.7566_{\pm 0.1198}$	$0.0181_{\pm 0.0067}$
VarRUOT (Sun et al., 2025)	$30.0459_{\pm 0.5180}$	$14.0375_{\pm 0.1266}$	$10.0279_{\pm 0.1121}$	$0.9190_{\pm 0.2783}$
CuOT (Ours)	$\mathbf{5.9340_{\pm 0.1216}}$	$\mathbf{12.4156_{\pm 0.0989}}$	$\mathbf{8.8793_{\pm 0.0947}}$	$\mathbf{0.0086_{\pm 0.0035}}$

Table 23: Child → Adult. FFHQ dataset (Karras et al., 2019), ALAE encoder (Pidhorskyi et al., 2020).

Method	FID	$\mathcal{W}_1$	$\mathcal{W}_2$	Mass
OTCFM (Tong et al., 2024a)	$4.7533_{\pm 0.0721}$	$14.5779_{\pm 0.1012}$	$10.3942_{\pm 0.1186}$	$12.4037_{\pm 0.0000}$
TIGON (Sha et al., 2024)	$20.6358_{\pm 0.1834}$	$16.5892_{\pm 0.1795}$	$11.8064_{\pm 0.1510}$	$0.7877_{\pm 0.0819}$
DeepRUOT (Zhang et al., 2025)	$6.1756_{\pm 0.0435}$	$13.8212_{\pm 0.0952}$	$9.8473_{\pm 0.0917}$	$0.0023_{\pm 0.0096}$
VarRUOT (Sun et al., 2025)	$24.3997_{\pm 0.0093}$	$15.5000_{\pm 0.1263}$	$11.0904_{\pm 0.1220}$	$0.8507_{\pm 0.3332}$
CuOT (Ours)	$\mathbf{4.4188_{\pm 0.0306}}$	$\mathbf{12.3927_{\pm 0.1170}}$	$\mathbf{8.8354_{\pm 0.1063}}$	$\mathbf{0.0000_{\pm 0.0000}}$

Table 24: Adult → Child. FFHQ dataset (Karras et al., 2019), e2e encoder (Richardson et al., 2021).

Method	FID	$\mathcal{W}_1$	$\mathcal{W}_2$	Mass
OTCFM (Tong et al., 2024a)	$10.2854_{\pm 0.2088}$	$20.3649_{\pm 0.3954}$	$16.0835_{\pm 0.19013}$	$0.6066_{\pm 0.0000}$
TIGON (Sha et al., 2024)	$48.2374_{\pm 0.4570}$	$29.02231_{\pm 0.3294}$	$20.9152_{\pm 0.3492}$	$1.8263_{\pm 0.1899}$
DeepRUOT (Zhang et al., 2025)	$41.2531_{\pm 0.1992}$	$28.0340_{\pm 0.2265}$	$21.5633_{\pm 0.2068}$	$0.0839_{\pm 0.0092}$
VarRUOT (Sun et al., 2025)	$53.6016_{\pm 0.0724}$	$30.5293_{\pm 0.2545}$	$24.3990_{\pm 0.3351}$	$1.0479_{\pm 0.3780}$
CuOT (Ours)	$\mathbf{8.9612_{\pm 0.1035}}$	$\mathbf{17.0263_{\pm 0.0935}}$	$\mathbf{13.1258_{\pm 0.1593}}$	$\mathbf{0.01983_{\pm 0.0148}}$

Table 25: Child → Adult. FFHQ dataset (Karras et al., 2019), e2e encoder (Richardson et al., 2021).

Method	FID	$\mathcal{W}_1$	$\mathcal{W}_2$	Mass
OTCFM (Tong et al., 2024a)	$8.4863_{\pm 0.1204}$	$18.3142_{\pm 0.2071}$	$13.5799_{\pm 0.1860}$	$12.4032_{\pm 0.0000}$
TIGON (Sha et al., 2024)	$42.3115_{\pm 0.3207}$	$28.6456_{\pm 0.3737}$	$20.1813_{\pm 0.3662}$	$1.3548_{\pm 0.1207}$
DeepRUOT (Zhang et al., 2025)	$38.1583_{\pm 0.1911}$	$27.4958_{\pm 0.2104}$	$19.8897_{\pm 0.2186}$	$0.0631_{\pm 0.0114}$
VarRUOT (Sun et al., 2025)	$55.6621_{\pm 0.0875}$	$32.1010_{\pm 0.2637}$	$24.7603_{\pm 0.2958}$	$1.2267_{\pm 0.4270}$
CuOT (Ours)	$\mathbf{7.8332_{\pm 0.0909}}$	$\mathbf{15.1472_{\pm 0.1326}}$	$\mathbf{11.0913_{\pm 0.1222}}$	$\mathbf{0.0107_{\pm 0.0140}}$

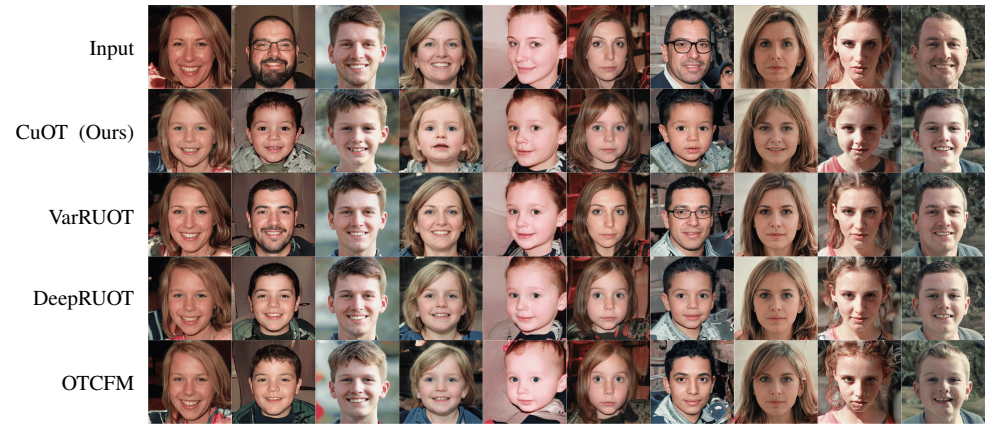

Figure 11: Adult → Child image generation experiment on the FFHQ dataset (Karras et al., 2019), ALAE encoder (Pidhorskyi et al., 2020).



Figure 12: Visualization of the bridge computed on the FFHQ dataset (Karras et al., 2019) by our CuOT method (ALAE encoder (Pidhorskyi et al., 2020)).



Figure 13: Visualization of the bridge computed on the FFHQ dataset (Karras et al., 2019) by the VarRUOT method Sun et al. (2025) (ALAE encoder (Pidhorskyi et al., 2020)).



Figure 14: Visualization of the bridge computed on the FFHQ dataset (Karras et al., 2019) by the DeepRUOT method Zhang et al. (2025) (ALAE encoder (Pidhorskyi et al., 2020)).



Figure 15: Visualization of the bridge computed on the FFHQ dataset (Karras et al., 2019) by the OTCFM method Tong et al. (2024a) (ALAE encoder (Pidhorskyi et al., 2020)).



Figure 16: Visualization of the bridge computed on the FFHQ dataset (Karras et al., 2019) by our CuOT method (e2e encoder (Richardson et al., 2021)).

F.4 Image-to-Video Experiment

We evaluate the scalability of CuOT on temporal video reconstruction using the UCF101 dataset (Soomro et al., 2012), a benchmark of 101 human action classes recorded at 320×240 pixel resolution. We select three action classes: *BreastStroke*, *BlowingCandles*, and *CuttingInKitchen*, each comprising respectively of 101, 105 and 104 videos. Within each class, the temporal sequence of frames defines the reference marginal distributions $\{\rho^m\}_{m=0}^M$, with $M = 60$ intermediate marginals corresponding to uniformly sampled discrete time steps along the video. Each marginal ρ^m is thus represented by the ensemble of frames $\{x_n^m\}$ at time step m across all videos in the class, providing the empirical distribution at that point in time. This formulation casts video reconstruction as a multi-marginal transport problem. Because the frame count per video remains constant across time steps, mass variation in this setting is negligible. Consequently, the primary objective of this experiment is not to test CuOT's mass-unbalancing capabilities, as in our other benchmark evaluations, but rather to demonstrate its scalability and stability when interpolating across a large number of marginal distributions in high-dimensional, structured data spaces.

To mitigate computational overhead, all solvers are trained directly within the latent space of pre-trained autoencoders. To evaluate method performance across varying latent space topologies, we consider two distinct encoder backbones: (i) the Deep Compression Autoencoder (Chen et al., 2025), which maps each 320×240 frame into a compact $32 \times 8 \times 8 = 2048$-dimensional feature map, and (ii) the TiTok encoder (Yu et al., 2024), which compresses frames into a 128×32 (4096-dimensional) sequence. Although TiTok yields a higher-dimensional representation, flattening the 2D spatial grid structure simplifies feature alignment for the transport networks, allowing us to maintain identical model capacities across both setups. During inference, an initial frame is encoded into the latent space, propagated along the predicted geodesic via the learned transport map, and decoded at each time step to reconstruct the video sequence. In both settings, we set the number of integration steps to $K = 180$, providing $K/M = 3$ discretization steps per observed marginal interval to ensure high temporal resolution. Full network and training configurations are detailed in Table 11.

This experiment poses two distinct challenges beyond those of the FFHQ pairing task. First, the high ambient dimensionality ($D = 2048$) stresses scalability, and the $32 \times 8 \times 8$ grid structure preserves spatial information, meaning the data do not lie in a simple vector space. Second, the large number of intermediate marginals ($M = 60$) requires the learned transport map to remain consistent across a long temporal horizon, preventing error accumulation along the predicted trajectory. These conditions expose the structural limitations of WFR-based methods, whose growth-rate bias compounds across marginals and leads to trajectory collapse in this high-dimensional regime.

We report three metrics. The FID is computed in the ambient pixel space after decoding, comparing the full set of reference frames against the full set of reconstructed frames across the entire video. We use the standard InceptionV3 feature extractor (2048-dimensional pool layer), following the formulation in equation (58) (Appendix F.3). This metric measures the ability of each method to recover the distributional support and diversity of the video content. The latent Wasserstein distances $\mathcal{W}_1$ and $\mathcal{W}_2$ are evaluated time-step by time-step in the latent space: for each $m \in \{0, \dots, M\}$, the empirical distribution of predicted latent codes is compared against the reference marginal ρ^m, and the reported value is averaged over all marginals. This per-timestep criterion directly probes whether each method correctly tracks the sequence of marginal distributions along the trajectory, penalising both early divergence and late-stage drift. The mass metric was intentionally omitted from the evaluation as the transport problem does not present mass variation.

Figures 17–18 display uniformly sampled frame sequences reconstructed from the CB latent space across each action class. Tables 26–28 detail the corresponding per-class quantitative results, while aggregate metrics are summarized in Table 8. Additionally, Table 29 presents the ablation results using the TiTok latent space, with corresponding decoded video trajectories across various initial frames visualized in Figure 20.

CuOT consistently achieves the best scores across all metrics, classes and latent representations, by a wide margin. Given the stringent requirement of aligning many intermediate marginals, this highlights the stability and effectiveness of the proposed geodesic-based resolution algorithm in high-dimensional multi-marginal settings. WFR-based baselines are typically not demonstrated on benchmarks with dimensionalities higher than 50, and they struggle significantly on this task. In this class of methods, Zhang et al. (2025) represents the strongest option, but still performs roughly an order of magnitude worse. This gap likely reflects the difficulty of jointly parameterizing multiple

variables and the need to enforce the HJB optimality equation in the loss (notoriously unstable (Buzun et al., 2026)), rather than directly simulating transport (provable stable in Appendix E.4). In this class we also find the method performing worst, Sun et al. (2025). Its strict enforcement of $v^t(x) = \nabla_x z^t(x)$ yields a noisy and unstable signal in the high-dimensional latent spaces, causing collapse to near-static trajectories where samples remain close to ρ^0 and exhibit minimal temporal evolution. Tong et al. (2024a) has been demonstrated to scale to higher-dimensional datasets like CIFAR-10, but it is fundamentally optimized for two marginals. Because it relies heavily on flow matching, adapting it to multiple marginals requires iterative reference refinement techniques to update the ground-truth matching along the learned path. This introduces significant computational overhead and drastically slows down convergence compared to our framework. While computationally efficient (similarly to Sha et al. (2024)), its reliance on density estimation rather than explicit OT matching does not scale: without re-anchoring to reference marginals, discretization errors accumulate over the $M = 60$ steps, leading to drift and higher variance.

CuOT consistently achieves the best performance across all metrics, action classes, and latent representations by a wide margin. Given the stringent requirement of aligning numerous intermediate marginals, these results highlight the stability and effectiveness of our proposed geodesic-based resolution algorithm in high-dimensional multi-marginal settings. Existing WFR-based uOT baselines are typically evaluated only on low-dimensional benchmarks (≤ 50 dimensions) and struggle significantly on this task. Among these, DeepRUOT (Zhang et al., 2025) represents the strongest option, yet it performs roughly an order of magnitude worse than CuOT. This performance gap stems from the inherent difficulty of jointly parameterizing multiple field variables and enforcing the HJB optimality loss, a process known to be notoriously unstable (Buzun et al., 2026), as opposed to directly simulating transport maps (provably stable; see Appendix E.4). Within the same family, VarRUOT (Sun et al., 2025) exhibits the poorest performance: its strict constraint $v^t(x) = \nabla_x z^t(x)$ produces noisy gradient signals in high-dimensional latent spaces, causing trajectories to collapse into near-static states that remain anchored close to ρ^0 with minimal temporal evolution. Flow-matching baselines like OTCFM (Tong et al., 2024a) have been shown to scale to higher-dimensional image datasets (e.g., CIFAR-10), but they are fundamentally optimized for two-marginal problems. Adapting OTCFM to multi-marginal trajectories requires iterative reference refinement to update ground-truth matchings along the learned path, introducing heavy computational overhead and substantially slowing convergence.

Table 26: *BreastStroke* class, UCF101 (Soomro et al., 2012), CB encoder (Chen et al., 2025).

Method	FID	$\mathcal{W}_1$	$\mathcal{W}_2$
OTCFM	$186.4736_{\pm 2.1483}$	$21.6347_{\pm 1.5824}$	$14.8247_{\pm 1.2436}$
TIGON	$318.4247_{\pm 2.8364}$	$26.8347_{\pm 1.9628}$	$17.4236_{\pm 1.4827}$
DeepRUOT	$\underline{132.9642}_{\pm 1.4471}$	$\underline{17.9528}_{\pm 1.0326}$	$\underline{12.5841}_{\pm 0.8735}$
VarRUOT	$310.7648_{\pm 2.9365}$	$24.1739_{\pm 1.8421}$	$15.8127_{\pm 1.2638}$
CuOT (Ours)	$\mathbf{16.4875_{\pm 0.6842}}$	$\mathbf{9.5832_{\pm 0.5217}}$	$\mathbf{7.3124_{\pm 0.4631}}$

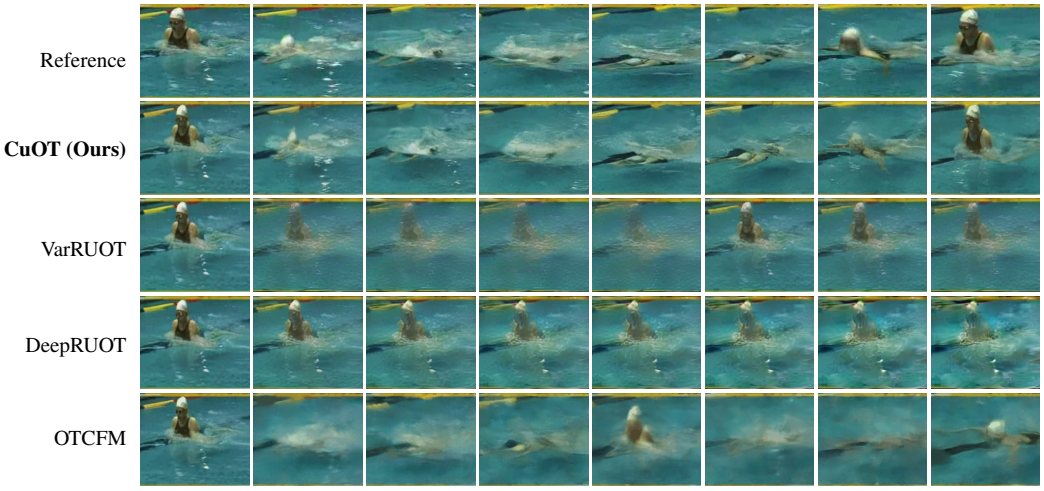

Figure 17: *BreastStroke* class, UCF101 (Soomro et al., 2012), CB encoder (Chen et al., 2025).

Table 27: *BlowingCandles* class, UCF101 (Soomro et al., 2012), CB encoder (Chen et al., 2025)..

Method	FID	$\mathcal{W}_1$	$\mathcal{W}_2$
OTCFM	$224.3847_{\pm 2.7624}$	$28.4736_{\pm 2.0847}$	$19.3628_{\pm 1.6247}$
TIGON	$352.1736_{\pm 3.9248}$	$36.4736_{\pm 2.5847}$	$23.1847_{\pm 1.8436}$
DeepRUOT	$\underline{143.9126}_{\pm 1.9637}$	$\underline{23.4632}_{\pm 1.5426}$	$\underline{16.7315}_{\pm 1.2084}$
VarRUOT	$340.6289_{\pm 3.8427}$	$33.7814_{\pm 2.4365}$	$21.9438_{\pm 1.7421}$
CuOT (Ours)	$\mathbf{20.4368_{\pm 0.8135}}$	$\mathbf{13.2147_{\pm 0.7426}}$	$\mathbf{9.3826_{\pm 0.5184}}$

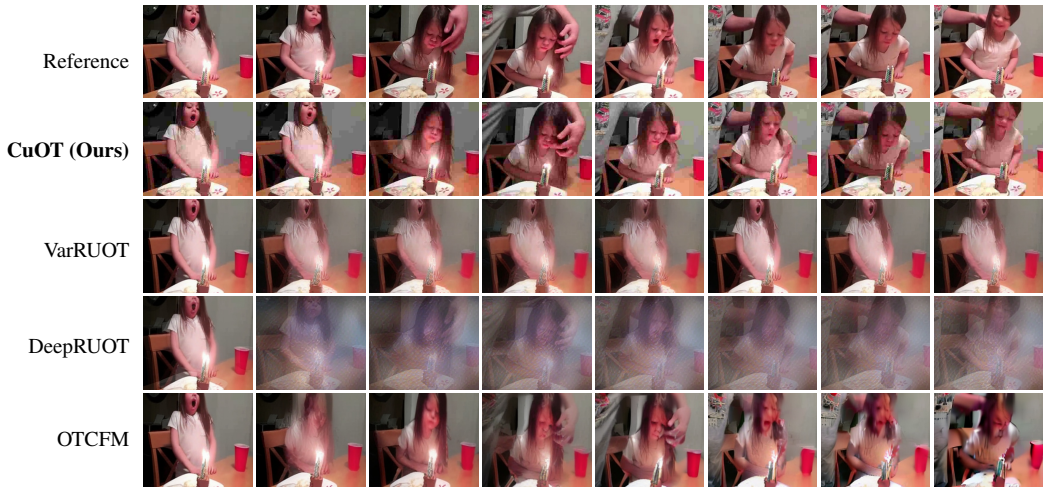

Figure 18: *BlowingCandles* class, UCF101 (Soomro et al., 2012), CB encoder (Chen et al., 2025).

Table 28: *CuttingInKitchen* class, UCF101 (Soomro et al., 2012), CB encoder (Chen et al., 2025)..

Method	FID	$\mathcal{W}_1$	$\mathcal{W}_2$
OTCFM	$214.7362_{\pm 2.5824}$	$26.8436_{\pm 1.9473}$	$17.9473_{\pm 1.4836}$
TIGON	$338.9362_{\pm 3.2647}$	$29.7647_{\pm 2.1436}$	$19.8347_{\pm 1.6827}$
DeepRUOT	$\underline{137.9231}_{\pm 1.7326}$	$\underline{20.5843}_{\pm 1.2874}$	$\underline{13.8417}_{\pm 1.0635}$
VarRUOT	$325.7062_{\pm 3.2148}$	$26.3447_{\pm 2.1046}$	$17.3431_{\pm 1.5627}$
CuOT (Ours)	$\mathbf{18.2106_{\pm 0.8043}}$	$\mathbf{10.8039_{\pm 0.6235}}$	$\mathbf{7.8058_{\pm 0.5172}}$

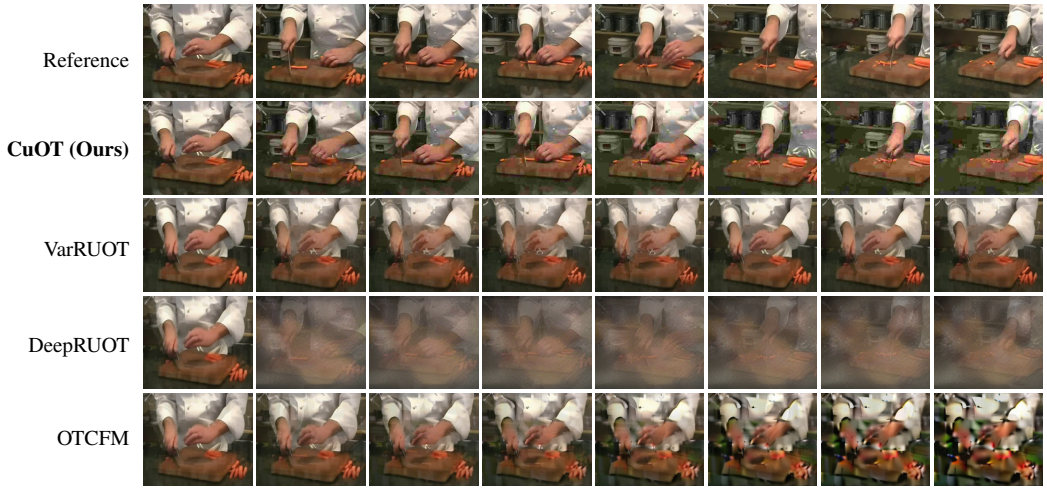

Figure 19: *CuttingInKitchen* class, UCF101 (Soomro et al., 2012), CB encoder (Chen et al., 2025).

Table 29: *BreastStroke* class, UCF101 (Soomro et al., 2012), TiTok encoder (Yu et al., 2024).

Method	FID	$\mathcal{W}_1$	$\mathcal{W}_2$
OTCFM	$92.4136_{\pm 2.9927}$	$18.2705_{\pm 1.8718}$	$15.7947_{\pm 1.3327}$
TIGON	$361.0779_{\pm 3.0314}$	$39.5178_{\pm 2.0123}$	$25.0515_{\pm 2.0329}$
DeepRUOT	$103.87_{\pm 2.1189}$	$19.3748_{\pm 1.9142}$	$17.7740_{\pm 1.0970}$
VarRUOT	$312.1575_{\pm 3.4569}$	$28.2140_{\pm 2.5769}$	$23.5365_{\pm 1.5311}$
CuOT (Ours)	$\mathbf{25.5301_{\pm 1.0747}}$	$\mathbf{11.4517_{\pm 0.6784}}$	$\mathbf{7.1934_{\pm 0.5704}}$

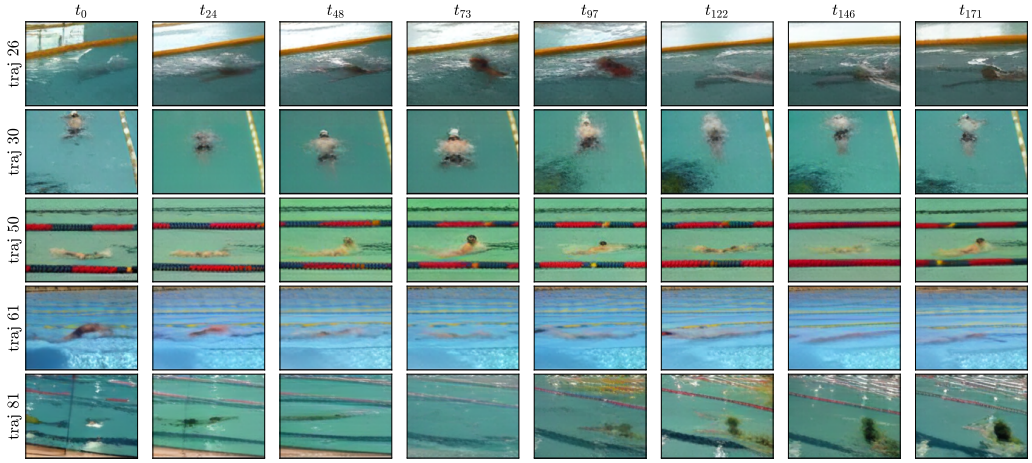

Figure 20: *BreastStroke* class, UCF101 (Soomro et al., 2012), TiTok encoder (Yu et al., 2024).

G Broader Impacts

CuOT introduces contact geometry as a unifying framework for optimal transport, bridging physical transport (described by Wasserstein geometry) and the information-theoretic structure of data distributions (described by Fisher–Rao geometry). This constitutes a significant theoretical advance in the study of the geometry of generative processes, as prior Wasserstein–Fisher–Rao approaches typically interpolate between these distinct regimes, whereas our method provides a generalized lift that avoids restrictive assumptions.

Our framework is particularly suited for modeling stochastic dynamics in the natural sciences. By removing the need for expert-specified growth priors (Chizat et al., 2022; Lavenant et al., 2024; Pariset et al., 2023; Ventre et al., 2023; Schiebinger et al., 2019), CuOT facilitates more accurate reconstruction of cell-fate dynamics and disease progression from sparse, static observations. This reduces reliance on costly longitudinal single-cell experiments and supports faster progress in areas such as regenerative medicine and oncology by extracting high-resolution developmental information from standard datasets. The proposed framework improves unpaired domain adaptation, particularly in translating synthetic data to real clinical imagery (Wang et al., 2024a; Bortoli et al., 2021). This enhances the robustness and generalizability of diagnostic models across varying acquisition conditions and clinical environments. CuOT also ensures that underrepresented data modes are preserved, potentially reducing bias in generative models used for decision-making or synthetic data augmentation. While CuOT improves reliability, deployment in medical settings requires careful validation to ensure that the decoupling of mass growth does not introduce artifacts that could affect clinical decisions.

On the implementation front, CuOT efficiency reduces the total computational budget, and the associated carbon footprint required for large-scale distribution alignment tasks compared to more resource-intensive quadratic-complexity alternatives, summarized in Table 1.

H Limitations and Future Directions

The experimental validation of CuOT has been conducted exclusively using MLP architectures for both the transport network v_θ^t and the growth network z_θ^t. MLPs treat input coordinates as mutually independent, and are therefore well-suited to the vector-valued data encountered in the single-cell experiments and in highly-compressed image latent spaces. However, they do not explicitly capture inter-element dependencies, such as spatial correlations in images or sequential structure in time series, that are critical in domains where the data geometry cannot be faithfully represented by a coordinate-wise factorization. Architectures such as Transformers and U-Nets represent the state of the art in structured sequence modeling. While the contact-geometric formulation of CuOT is architecture-agnostic and does not preclude the use of such backbones, their integration with the geodesic discretization scheme of Eqs. (6)–(8) has not been investigated, and the scalability and reconstruction quality of CuOT in these high-structure regimes remains to be studied.

As an Optimal Transport (OT) formulation designed for interpolating between distribution pairs and multiple marginals, CuOT is not directly comparable to specialized image-to-video models (Ni et al., 2023; Wang et al., 2024b). Interpolation-based approaches, which treat each video frame as a discrete sample, are inherently less efficient than architectures that represent an entire video as a single latent point. By utilizing the initial image as a conditioning factor, these specialized models optimize higher-dimensional but less constrained generative dynamics, resulting in a more scalable optimization framework.

CuOT enables the adaptation of generative sampling steps to the varying data density of the interpolated marginals, mitigating overshooting effects along the transport path. An interesting structural parallel arises with recent work investigating the geometry of diffusion-based generation (Karczewski et al., 2026), which operates in a more restricted setting by connecting data points within a single trained model via geodesics under the Fisher–Rao metric, rather than bridging arbitrary unpaired marginals. Both frameworks share the underlying principle that step sizes should reflect the local geometry of the data distribution rather than ambient Euclidean structure. Nevertheless, an explicit characterization of the link between the contact geometry introduced here and the Fisher–Rao geometry, and more broadly, the relationship between contact dynamics on the probability manifold and information-geometric metrics on latent spaces, remains an open direction for future work.