DYNAMIC GENERALIZED GROMOV-WASSERSTEIN OPTIMAL TRANSPORT

Anonymous authors
Paper under double-blind review

ABSTRACT

Gromov–Wasserstein optimal transport (GW-OT) extends classical optimal transport by introducing structure-aware transport cost. This is particularly relevant for spatial transcriptomics, where dynamical reconstruction should preserve tissue structure in addition to matching expression patterns. While static formulations have been widely used for such structure-aware alignment, a general dynamic formulation for reconstructing continuous trajectories is still missing. We introduce **T**ravelling **P**air **D**ynamical **A**lignment and **T**rajectory **E**stimation (TP-DATE), a theoretical and computational framework to generalize GW-OT dynamically in a simulation-free manner. We formulate a broad class of static and dynamic Quadratic-form OT (QOT) through path actions and prove the static dynamic equivalence. We further develop travelling-pair flow matching, which allows interacting conditional paths and marginalizes their interactions into a single vector field. On synthetic and real spatial transcriptomics data, TP-DATE better preserves spatial structure and improves continuous 3D dynamics reconstruction.

1 INTRODUCTION

Optimal transport (OT) (Kantorovich, 1958) provides a principled way to match probability distributions by minimizing transport cost, while its dynamic formulation (Benamou & Brenier, 2000) lifts this endpoint matching into a continuous time evolution of measures. This perspective has made OT a natural tool for reconstructing dynamics from population snapshots, including single-cell systems where individual cells cannot be tracked longitudinally. Static OT aligns snapshots by infering couplings across time points, whereas dynamic OT enables continuous interpolation and trajectory inference between observed snapshots (Schiebinger et al., 2019; Tong et al., 2020; 2024a; Sha et al., 2024; Zhang et al., 2025a; Peng et al., 2026b). Extensions such as Schrödinger bridges (Schrödinger, 1932; Léonard, 2014) and Wasserstein–Fisher–Rao (Chizat et al., 2018b;a; Liero et al., 2018) further broaden this framework to stochastic and unbalanced dynamics.

However, pointwise transport cost alone may be insufficient when the data carry meaningful internal structure such as spatial structure. Gromov–Wasserstein OT (GW-OT) compares pairwise relations within distributions and are therefore well suited to structure-aware matching tasks Mémoli (2011). Fused GW-OT (FGW-OT) further combines feature and structural information (Vayer et al., 2020; Klein et al., 2025). Such methods have become useful for structured data including graphs, heterogeneous domains, and spatial omics. However, most GW-like OTs still lack a dynamic formulation, therefore can only align snapshots instead of reconstructing a continuous time dynamics.

Though a dynamic GW-like formulation may not be meaningful when aligning snapshots from different modalities, it becomes important when the snapshots represent different states of the same structured system. Spatial transcriptomics is a canonical example. Snapshots collected at different times describe the same tissue evolving over time, and a meaningful interpolation should recover not only gene expression changes but also the continuous evolution of spatial organization (Chen et al., 2022; Wei et al., 2022). The same principle applies to serial tissue sections, where the interpolation axis is tissue depth rather than time and unseen intermediate sections correspond to physically meaningful states (Zhang et al., 2026a).

Recent work has begun to generalize GW-OT from complementary directions. On the static side, (Wang & Zhang, 2025) defines a family of static Quadratic-form optimal transport (QOT) including GW-OT as a special case. On the dynamic side, inner-product GW-OT (IGW-OT) (Zhang et al.,

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

2026b) develops and studies a dynamic formulation for a particular GW-like OT based on gradient flow and Riemannian geometry. However, a general dynamic QOT formulation and its theory is still missing, as is an efficient simulation-free algorithm for solving them.

To address these limitations, We introduce **T**ravelling **P**air **D**ynamical **A**lignment and **T**rajectory **E**stimation (TP-DATE), a general framework for dynamic QOT together with a simulation-free travelling-pair flow matching solver. Our contributions are summarized as follows.

- We developed a mathematical theory of dynamic QOT and included standard OT, GW-OT, and IGW-OT as special cases of our QOT framework.
- We developed a simulation-free travelling pair flow matching framework for learning dynamics with interacting conditional paths and solving the dynamic QOT problems.
- We proposed a dynamic fusion of OT and GW-OT, and demonstrate the ability of it in several biological tasks including spatiotemporal dynamics and 3D structure reconstruction.

2 Related works

Optimal transport and extensions. Optimal transport (Kantorovich, 1958) and its dynamic formulations (Benamou & Brenier, 2000) have been widely used to reconstruct dynamics from snapshot data. Stochastic counterparts (Schrödinger, 1932; Léonard, 2014) and unbalanced extensions (Chizat et al., 2018a; Liero et al., 2018) have further been applied to model more complex dynamics. Another important line of extension is QOT, represented by formulations such as GW-OT and FGW-OT (Mémoli, 2011; Vayer et al., 2020; Klein et al., 2025), which can model global structure preservation during transport. Recently, (Wang & Zhang, 2025) introduced a static QOT framework that unifies a broad class of such extensions, while (Zhang et al., 2026b) developed IGW-OT as a dynamic formulation for a particular QOT problem. We propose TP-DATE, a unified dynamic formulation for a broad class of QOT problems which is compatible with these existing formulations.

Flow matching based optimal transport solvers. Flow matching (Lipman et al., 2023) is an efficient simulation-free generative modeling framework that has been used to solve OT (Tong et al., 2024a), Schrödinger bridge (Tong et al., 2024b), WFR (Peng et al., 2026b), and a variety of OT-based dynamics reconstruction problems (Eyring et al., 2024; Rathod et al., 2026; Klein et al., 2024; Ying et al., 2026). Inspired by the conditional path technique and the travelling Dirac in optimal transport (Chizat et al., 2018a), we formulate dynamic QOT from a path action viewpoint and develop a flow matching solver for this class of problems.

OT-based spatial transcriptomics dynamics reconstruction. Several recent methods have been developed for spatiotemporal dynamics inference from spatial transcriptomics. On the static alignment side, (Klein et al., 2025; Zeira et al., 2022; Liu et al., 2023; Bryan et al., 2026) formulate cross-time or spatial alignment via GW-OT and FGW-OT or related structure-aware OT objectives, while (Halmos et al., 2025) encodes temporal and spatial structural information into static transport objectives. On the dynamic side, (Rathod et al., 2026) incorporates structure-aware static couplings into flow matching. (Peng et al., 2026a) introduces structure-preserving terms directly into the dynamic transport objective, and (Zhang et al., 2025b) explicitly models interactions within the learned dynamics, both by simulation-based NeuralODE (Chen et al., 2018). This leaves a clear gap for TP-DATE, a simulation-free flow-matching framework that directly models structure-aware transport dynamics.

3 Preliminaries

Static optimal transport. Let $\mathcal{P}(\mathcal{X})$ be the set of all probability densities supported on $\mathcal{X}$ for some $\mathcal{X} \subset \mathbb{R}^d$. μ_0, μ_1 are probability densities in $\mathcal{P}(\mathcal{X})$. $\Pi(\mu_0, \mu_1)$ denotes the set of all couplings.

$$\Pi(\mu_0, \mu_1) = \left\{ \gamma \in \mathcal{P}(\mathcal{X}^2) | \int_{\mathcal{X}} \gamma(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{y} = \mu_0(\boldsymbol{x}),\ \int_{\mathcal{X}} \gamma(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{x} = \mu_1(\boldsymbol{y}) \right\}$$

The static optimal transport (OT), also known as the Kantorovich form (Kantorovich, 1958), is defined as

$$\mathrm{OT}(\mu_0, \mu_1) = \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\mathcal{X}^2} C(\boldsymbol{x}, \boldsymbol{y}) \gamma(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{y} \tag{1}$$

where $C(\boldsymbol{x}, \boldsymbol{y})$ represents the cost of transporting unit mass from $\boldsymbol{x}$ to $\boldsymbol{y}$, and $\gamma(\boldsymbol{x}, \boldsymbol{y}) \in \Pi(\mu_0, \mu_1)$ is called the coupling. From an optimization perspective, (1) can be viewed as a linear programming w.r.t the coupling γ. As a well known example, when choosing $C(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|^2$, which is the square of the Euclidean distance, the infimum is called the square of 2-Wasserstein distance ($\mathcal{W}_2$).

Static quadratic-form optimal transport. Recently, a static quadratic-from optimal transport (QOT) problem is defined and studied mathematically (Wang & Zhang, 2025). Given two spaces $\mathcal{X}, \mathcal{Y}$, and a cost function $C : (\mathcal{X} \times \mathcal{Y})^2 \to \mathbb{R}$, the static QOT is defined as

$$\text{QOT}(\mu_0, \mu_1) = \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{(\mathcal{X}\times\mathcal{Y})^2} C(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x}', \boldsymbol{y}')\gamma(\boldsymbol{x}, \boldsymbol{y})\gamma(\boldsymbol{x}', \boldsymbol{y}')\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{x}'\mathrm{d}\boldsymbol{y}' \tag{2}$$

If we choose $C(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x}', \boldsymbol{y}') = |d_{\mathcal{X}}(\boldsymbol{x}, \boldsymbol{x}') - d_{\mathcal{Y}}(\boldsymbol{y}, \boldsymbol{y}')|^2$ where $(\mathcal{X}, d_{\mathcal{X}}), (\mathcal{Y}, d_{\mathcal{Y}})$ are two metric spaces, (2) recovers the Gromov-Wasserstein OT (GW-OT). This type of QOT has been widely used in single-cell trajectory inference (Mémoli, 2011; Klein et al., 2025). Intuitively, GW-OT preserves the local structure after transport. If we choose $C(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x}', \boldsymbol{y}') = f(\boldsymbol{x}, \boldsymbol{y}) + g(\boldsymbol{x}', \boldsymbol{y}')$, (2) just reduces to the standard static OT with cost $f + g$. Different to static OT, static QOT is a quadratic programming w.r.t the coupling γ.

Dynamic optimal transport. For static OT, (Benamou & Brenier, 2000) established a dynamic form for $\mathcal{W}_2$ case, also known as the BB-form.

$$\begin{aligned} &\text{OT}(\mu_0, \mu_1) = \inf_{\rho, \boldsymbol{u}} \int_0^1 \int_{\mathcal{X}} \frac{1}{2}\|\boldsymbol{u}(\boldsymbol{x}, t)\|_2^2 \rho_t(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t \\ &\text{s.t.} \qquad \partial_t \rho + \nabla_{\boldsymbol{x}} \cdot (\rho \boldsymbol{u}) = 0,\ \rho_0 = \mu_0,\ \rho_1 = \mu_1 \end{aligned} \tag{3}$$

They proved the equivalence between (3) and (1) when $C(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|^2$. Intuitively, it aims to find a continuous probability flow connecting μ_0, μ_1 which also minimizes the total kinetic energy. With nice fluid dynamics interpretation, it has also been widely used in single-cell trajectory inference (Tong et al., 2020; 2024a; Klein et al., 2024).

Travelling Dirac. To solve dynamic OT, one can first consider the dynamic OT between two Dirac measures

$$\text{OT}(\delta_{\boldsymbol{x}_0}, \delta_{\boldsymbol{x}_1}) = \inf_{\boldsymbol{x}_t} \frac{1}{2} \int_0^1 \|\dot{\boldsymbol{x}}_t\|_2^2 \mathrm{d}t \quad \text{s.t.} \quad \boldsymbol{x}_0 = \boldsymbol{x}_0, \quad \boldsymbol{x}_1 = \boldsymbol{x}_1 \tag{4}$$

which yields the displacement interpolation $\boldsymbol{x}_t = (1 - t)\boldsymbol{x}_0 + t\boldsymbol{x}_1$, also known as the travelling Dirac. More complex travelling Dirac can also be derived for different type of OT, such as unbalanced OT (Chizat et al., 2018b;a). For a specific dynamic OT, previous works obtained its solution by integrating these travelling Diracs over the static coupling γ (Tong et al., 2024a; Peng et al., 2026b). Therefore, the dynamic OT can decoupled to two parts: the travelling Dirac and the static coupling. Integration can be realized by flow matching.

Flow matching. Flow matching is a simulation-free generative framework for learning a continuous probability flow from data (Lipman et al., 2023). Given μ_0, μ_1, it aims to learn a marginal velocity field $\boldsymbol{u}_t(\boldsymbol{x})$, such that $\partial_t \rho + \nabla_{\boldsymbol{x}} \cdot (\rho \boldsymbol{u}) = 0$ and $\rho_0 = \mu_0, \rho_1 = \mu_1$, which means $\boldsymbol{u}_t(\boldsymbol{x})$ transports μ_0 to μ_1. They parameterized a neural network $\boldsymbol{u}_{\boldsymbol{\theta}}$ to approximate the true $\boldsymbol{u}$. The neural networks are trained to minimize the marginal regression loss.

$$\mathcal{L}_{\text{FM}}(\boldsymbol{\theta}) = \int_0^1 \int_{\mathcal{X}} \|\boldsymbol{u}_{\boldsymbol{\theta}}(\boldsymbol{x}, t) - \boldsymbol{u}_t(\boldsymbol{x})\|_2^2 \, \rho_t(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t \tag{5}$$

Although the true $\rho, \boldsymbol{u}$ are intractable, they proved that minimizing the loss above is equivalent to minimize the conditional regression loss

$$\mathcal{L}_{\text{CFM}}(\boldsymbol{\theta}) = \mathbb{E}_{t\sim\mathcal{U}[0,1], \boldsymbol{z}\sim q(\boldsymbol{z}), \boldsymbol{x}\sim\rho_t(\boldsymbol{x}|\boldsymbol{z})} \|\boldsymbol{u}_{\boldsymbol{\theta}}(\boldsymbol{x}, t) - \boldsymbol{u}_t(\boldsymbol{x}|\boldsymbol{z})\|_2^2 \tag{6}$$

where $\boldsymbol{z} \sim q(\boldsymbol{z})$ is some conditional variable and $\partial_t \rho_t(\boldsymbol{x}|\boldsymbol{z}) + \nabla_{\boldsymbol{x}} \cdot (\rho_t(\boldsymbol{x}|\boldsymbol{z})\boldsymbol{u}_t(\boldsymbol{x}|\boldsymbol{z})) = 0$ is called the conditional continuity equation, which the conditional probability flow $\rho_t(\boldsymbol{x}|\boldsymbol{z})$ and the conditional velocity $\boldsymbol{u}_t(\boldsymbol{x}|\boldsymbol{z})$ satisfy. The marginal probability flow satisfies $\rho_t(\boldsymbol{x}) = \int \rho_t(\boldsymbol{x}|\boldsymbol{z})q(\boldsymbol{z})\mathrm{d}\boldsymbol{z}$. The marginalization theorem states that the marginal velocity $\boldsymbol{u}_t(\boldsymbol{x}) = \int \boldsymbol{u}_t(\boldsymbol{x}|\boldsymbol{z}) \frac{\rho_t(\boldsymbol{x}|\boldsymbol{z})q(\boldsymbol{z})}{\rho_t(\boldsymbol{x})} \mathrm{d}\boldsymbol{z}$ transports μ_0 to μ_1. Therefore, one can learn a admissible $\boldsymbol{u}_t(\boldsymbol{x})$ by designing tractable $q(\boldsymbol{z})$ and conditional path $\rho_t(\boldsymbol{x}|\boldsymbol{z})$ with tractable conditional velocity $\boldsymbol{u}_t(\boldsymbol{x}|\boldsymbol{z})$, and applying flow matching.

By careful design, flow matching can be used for solving dynamic OT. Following (Tong et al., 2024a), one can choose $\boldsymbol{z} = (\boldsymbol{x}_0, \boldsymbol{x}_1) \sim \gamma(\boldsymbol{x}_0, \boldsymbol{x}_1)$ drawn from the static OT coupling, set the travelling Dirac as conditional path, and derive the conditional velocity from it. The resulting flow are proved to recover the dynamic OT flow. Under this framework, travelling Diracs are integrated over the static OT coupling independently, which means particles actually move independently. In this work, we develop a flow matching framework which allows conditional paths to interact.

4 DYNAMIC QOT

In this section, we consider one metric space $\mathcal{X} \subset \mathbb{R}^d$, i.e. $\mathcal{Y} = \mathcal{X}$, and develop a dynamic formulation for QOT under this condition. Since the flow have to live in some space, the meaning of a dynamic flow connecting two totally different spaces needs further study. All the proofs are left to A.

4.1 STATIC AND DYNAMIC FORM

Path action. To model conditional paths with interaction, we study how a pair is transported to a pair rather than travelling Dirac. We call the corresponding path **travelling pair**. Let $\boldsymbol{z} = (\boldsymbol{x}_0, \boldsymbol{x}_1), \boldsymbol{z}' = (\boldsymbol{y}_0, \boldsymbol{y}_1)$, we define the action of a pair path for some Lagrangian $\mathcal{L}$ with sufficient regularity

$$\mathcal{A}(\boldsymbol{z}, \boldsymbol{z}') = \inf_{\boldsymbol{x}_t, \boldsymbol{y}_t} \int_0^1 \mathcal{L}(t, \boldsymbol{x}_t, \boldsymbol{y}_t, \dot{\boldsymbol{x}}_t, \dot{\boldsymbol{y}}_t) \mathrm{d}t \tag{7}$$

The minimizer yields a conditional velocity pair $\dot{\boldsymbol{x}}_t = \boldsymbol{u}_t^1(\boldsymbol{x}_t, \boldsymbol{y}_t | \boldsymbol{z}, \boldsymbol{z}'), \dot{\boldsymbol{y}}_t = \boldsymbol{u}_t^2(\boldsymbol{x}_t, \boldsymbol{y}_t | \boldsymbol{z}, \boldsymbol{z}')$. It can also be viewed as over the travelling Pair measure path $\pi_t(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}, \boldsymbol{z}')$

$$\begin{aligned}
&\mathcal{A}(\boldsymbol{z}, \boldsymbol{z}') = \inf_{\pi, \boldsymbol{u}^1, \boldsymbol{u}^2} \int_0^1 \int_{\mathcal{X}^2} \mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, u_t^1(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}, \boldsymbol{z}'), u_t^2(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}, \boldsymbol{z}')) \pi_t(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}, \boldsymbol{z}') \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{y} \mathrm{d}t \\
&\partial_t \pi_t(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}, \boldsymbol{z}') + \nabla_{\boldsymbol{x}} \cdot (\pi_t(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}, \boldsymbol{z}') \boldsymbol{u}_t^1(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}, \boldsymbol{z}')) + \nabla_{\boldsymbol{y}} \cdot (\pi_t(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}, \boldsymbol{z}') \boldsymbol{u}_t^2(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}, \boldsymbol{z}')) = 0
\end{aligned} \tag{8}$$

with boundary condition $\pi_0(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}, \boldsymbol{z}') = \delta_{(\boldsymbol{x}_0, \boldsymbol{y}_0)}(\boldsymbol{x}, \boldsymbol{y}), \pi_1(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}, \boldsymbol{z}') = \delta_{(\boldsymbol{x}_1, \boldsymbol{y}_1)}(\boldsymbol{x}, \boldsymbol{y})$.

Static form. Based on the path action $\mathcal{A}$, we can define the corresponding static QOT as

$$\mathrm{QOT}_S(\mu_0, \mu_1) = \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\mathcal{X}^4} \mathcal{A}(\boldsymbol{z}, \boldsymbol{z}') \gamma(\boldsymbol{z}) \gamma(\boldsymbol{z}') \mathrm{d}\boldsymbol{z} \mathrm{d}\boldsymbol{z}' \tag{9}$$

The optimal coupling can be solved by standard quadratic programming algorithms, such as Frank-Wolfe (Kerdoncuff et al., 2021; Flamary et al., 2021). We left the details to B.8.

Dynamic form. Inspired by the travelling Dirac technique in standard OT, we define the dynamic QOT on the travelling pair path level. The corresponding dynamic probability flow is $\pi_t(\boldsymbol{x}, \boldsymbol{y}) = \int \pi_t(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}, \boldsymbol{z}') \gamma(\boldsymbol{z}) \gamma(\boldsymbol{z}') \mathrm{d}\boldsymbol{z} \mathrm{d}\boldsymbol{z}'$.

$$\begin{aligned}
&\mathrm{QOT}_D(\mu_0, \mu_1) = \\
&\inf \int_0^1 \int \int_{\mathcal{X}^2} \mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, u_t^1(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}, \boldsymbol{z}'), u_t^2(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}, \boldsymbol{z}')) \pi_t(\boldsymbol{x}, \boldsymbol{y} | \boldsymbol{z}, \boldsymbol{z}') \gamma(\boldsymbol{z}) \gamma(\boldsymbol{z}') \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{y} \mathrm{d}\boldsymbol{z} \mathrm{d}\boldsymbol{z}' \mathrm{d}t
\end{aligned} \tag{10}$$

The infimum is taken over the coupling γ and all travelling pairs. A fluid dynamics form, or say BB-form, analogous to OT (Benamou & Brenier, 2000), is also defined as

$$\begin{aligned}
&\mathrm{QOT}_{BB}(\mu_0, \mu_1) = \inf_{\pi, \boldsymbol{u}^1, \boldsymbol{u}^2} \int_0^1 \int_{\mathcal{X}^2} \mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, u_t^1(\boldsymbol{x}, \boldsymbol{y}), u_t^2(\boldsymbol{x}, \boldsymbol{y})) \pi_t(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{y} \mathrm{d}t \\
&\partial_t \pi_t(\boldsymbol{x}, \boldsymbol{y}) + \nabla_{\boldsymbol{x}} \cdot (\pi_t(\boldsymbol{x}, \boldsymbol{y}) u_t^1(\boldsymbol{x}, \boldsymbol{y})) + \nabla_{\boldsymbol{y}} \cdot (\pi_t(\boldsymbol{x}, \boldsymbol{y}) u_t^2(\boldsymbol{x}, \boldsymbol{y})) = 0
\end{aligned} \tag{11}$$

where π_t is a continuous probability flow on the two-particle space $\mathcal{X}^2$, with boundary value $\mu_0 \otimes \mu_0, \mu_1 \otimes \mu_1$. The marginal velocity pair $\boldsymbol{u}_t^1, \boldsymbol{u}_t^2$ are particle velocities with interaction effect. The formulation intuitively seeks for a two-particle flow minimizing the total action. Our first result is the equivalence between the static and dynamic form. However, these forms are upper bound of the BB-form but not equal to it, hence a surrogate. We discuss the details and difficulties in F. Fortunately, since both the static and dynamic formulations provide upper bounds on the BB-form, solving dynamic QOT also implicitly minimizes the total energy in the sense of the BB-form.

Theorem 4.1. *If $\mathcal{L}$ is convex w.r.t $(\dot{\boldsymbol{x}}, \dot{\boldsymbol{y}})$, then $QOT_S(\mu_0, \mu_1) = QOT_D(\mu_0, \mu_1) \geq QOT_{BB}(\mu_0, \mu_1)$.*

4.2 LAGRANGIAN WITH TRACTABLE CONDITIONAL VELOCITY PAIR

In the next section, we established a flow matching framework for solving dynamic QOT flows. In order to do that, we need to access to the conditional velocity pair. Therefore, we then consider what kind of $\mathcal{L}$ leads to tractable conditional velocity. Note that TP-DATE is not restricted to the choices introduced below. As long as the conditional velocity pair can be obtained in some way, TP-DATE remains applicable. To take kinetic energy into account, we consider a broad class of $\mathcal{L}$ with the form

$$\mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, \dot{\boldsymbol{x}}, \dot{\boldsymbol{y}}) = \frac{1}{2}\|\dot{\boldsymbol{x}}\|^2 + \frac{1}{2}\|\dot{\boldsymbol{y}}\|^2 + \lambda\Phi(\boldsymbol{x} - \boldsymbol{y}, \dot{\boldsymbol{x}} - \dot{\boldsymbol{y}}) \tag{12}$$

where Φ represents a convex interaction term, making it different to standard OT. To simplify the pair dynamics, we change the variables to $\boldsymbol{c}_t = \frac{\boldsymbol{x}_t + \boldsymbol{y}_t}{2}$, $\boldsymbol{q}_t = \boldsymbol{x}_t - \boldsymbol{y}_t$, which are the barycenter and the relative displacement of the pair. The Lagrangian becomes

$$\mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, \dot{\boldsymbol{x}}, \dot{\boldsymbol{y}}) = \|\dot{\boldsymbol{c}}_t\|^2 + \frac{1}{4}\|\dot{\boldsymbol{q}}_t\|^2 + \lambda\Phi(\boldsymbol{q}, \dot{\boldsymbol{q}}) \tag{13}$$

Since the convexity preserves under affine transformation, it is sufficient to choose Φ to be convex w.r.t $\dot{\boldsymbol{q}}$. One natural choice is $\Phi(\boldsymbol{q})$, which is independent of $\dot{\boldsymbol{q}}$. If Φ is further radial, then the travelling pair has a low dimensional structure, hence learnable without the curse of dimensionality.

Proposition 4.2. *The optimal $\boldsymbol{c}_t$ is $\boldsymbol{c}_t = (1-t)\boldsymbol{c}_0 + t\boldsymbol{c}_1$. If $\Phi = \Phi(\|\boldsymbol{q}\|)$, $\boldsymbol{q}_0 \nparallel \boldsymbol{q}_1$, then $\boldsymbol{q}_t \in$ span$\{\boldsymbol{q}_0, \boldsymbol{q}_1\}$.*

GW-OT as a limit case. Another family of meaningful Φ is $\Phi = |\frac{\mathrm{d}}{\mathrm{d}t}\phi(\boldsymbol{q}_t)|^2$. It penalizes the variation of the relative displacement. The most explicit choice is $\Phi = |\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{q}_t\||^2$. Intuitively, this seeks for a transport not only saving kinetic energy, but also trying to minimize the distortion. This Φ leads to analytic conditional velocity.

Proposition 4.3. *If $\Phi = |\frac{\mathrm{d}}{\mathrm{d}t}\phi(\boldsymbol{q}_t)|^2$, Φ is convex w.r.t $\dot{\boldsymbol{q}}_t$. If $\Phi = |\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{q}_t\||^2$, $\boldsymbol{q}_t$ has analytic solution.*

A notable property of this choice of $\Phi = |\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{q}_t\||^2$ is the following theorem.

Theorem 4.4. *$\Phi = |\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{q}_t\||^2$. When $\lambda = 0$ and $d \geq 2$, the corresponding static QOT reduces to standard OT, while when $\lambda \to +\infty$, it reduces to standard GW-OT: λ^{-1}QOT$_S(\mu_0, \mu_1) \to$ GW-OT(μ_0, μ_1).*

We therefore refer to this choice of Lagrangian as a **dynamic fusion of OT and GW-OT**. Of note, this is different from the existing FGW-OT formulation (Vayer et al., 2020; Klein et al., 2025). FGW-OT is obtained by taking a weighted combination directly at the static level, whereas dynamic fusion introduces the weighting at the level of the dynamic formulation. Consequently, its induced static formulation is not FGW-OT. Also, by choosing proper Φ, the recently considered static IGW-OT (Zhang et al., 2026b) can also be realized as a special case of our dynamic QOT framework. We discuss the relation between these formulations in D.1 D.2.

Modality separation. In some settings, the coordinates may consist of multiple modalities, and we may wish to impose different interaction terms on different modalities. We illustrate the formulation using the two modality case, the extension to multiple modalities follows naturally. Let $\boldsymbol{x} = (\boldsymbol{x}^1, \boldsymbol{x}^2), \boldsymbol{y} = (\boldsymbol{y}^1, \boldsymbol{y}^2)$, and the Lagrangian is separated to the two modalities

$$\mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, \dot{\boldsymbol{x}}, \dot{\boldsymbol{y}}) = \mathcal{L}_1(t, \boldsymbol{x}^1, \boldsymbol{y}^1, \dot{\boldsymbol{x}}^1, \dot{\boldsymbol{y}}^1) + \mathcal{L}_2(t, \boldsymbol{x}^2, \boldsymbol{y}^2, \dot{\boldsymbol{x}}^2, \dot{\boldsymbol{y}}^2). \tag{14}$$

Then one can easily check that the conditional velocities of the two modalities can also be computed separately. The conditional velocity pair of $(\boldsymbol{x}, \boldsymbol{y})$ is exactly the concatenation of the conditional velocity pairs induced by $\mathcal{L}_i$.

5 TRAVELLING PAIR FLOW MATCHING

In this section, we develop a flow matching framework for solving the dynamic QOT problem. As OT-CFM (Tong et al., 2024a) averages travelling Diracs over OT coupling, our framework allows to average travelling pairs over QOT coupling. All the proofs are left to A.

5.1 MARGINALIZATION THEORY

Pair marginalization. Given conditional probability paths $\pi_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')$ and conditional velocity pairs $\boldsymbol{u}_t^1(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}'), \boldsymbol{u}_t^2(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')$ satisfying the conditional continuity equation, let the conditional variables $(\boldsymbol{z}, \boldsymbol{z}') \sim q(\boldsymbol{z}, \boldsymbol{z}')$, and define the marginal probability flow as $\pi_t(\boldsymbol{x}, \boldsymbol{y}) = \int \pi_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')q(\boldsymbol{z}, \boldsymbol{z}')\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'$, the marginalization theorem on the two-particle space holds.

Theorem 5.1. *Define the marginal velocity as* $\boldsymbol{u}_t^i(\boldsymbol{x}, \boldsymbol{y}) = \int \boldsymbol{u}_t^i(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')\frac{\pi_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')q(\boldsymbol{z}, \boldsymbol{z}')}{\pi_t(\boldsymbol{x}, \boldsymbol{y})}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'$, *the marginals satisfy the marginal continuity equation*

$$\partial_t \pi_t + \nabla_{\boldsymbol{x}} \cdot (\pi_t \boldsymbol{u}_t^1) + \nabla_{\boldsymbol{y}} \cdot (\pi_t \boldsymbol{u}_t^2) = 0 \tag{15}$$

If $q = \gamma \otimes \gamma$ *for some* $\gamma \in \Pi(\mu_0, \mu_1)$, *the marginal velocity transports* $\mu_0 \otimes \mu_0$ *to* $\mu_1 \otimes \mu_1$.

Particle marginalization. The marginal velocity $\boldsymbol{u}_t^1(\boldsymbol{x}, \boldsymbol{y})$ can be interpreted as how particle $\boldsymbol{x}$ travels given the influence of $\boldsymbol{y}$. From this perspective, one can obtain a marginal velocity $\boldsymbol{v}_t(\boldsymbol{x})$ only depends on $\boldsymbol{x}$ by averaging the influences of all possible $\boldsymbol{y}$.

$$\boldsymbol{v}_t(\boldsymbol{x}) = \mathbb{E}_{\pi_t}[\boldsymbol{u}_t^1(\boldsymbol{X}_t, \boldsymbol{Y}_t)|\boldsymbol{X}_t = \boldsymbol{x}] = \int_{\mathcal{X}} \boldsymbol{u}_t^1(\boldsymbol{x}, \boldsymbol{y})\pi_t(\boldsymbol{y}|\boldsymbol{x})\mathrm{d}\boldsymbol{y} \tag{16}$$

This velocity only depends on the particle itself, but it already contains all the interaction information by taking conditional expectation. We further define the $\boldsymbol{x}$ marginal $\rho_t(\boldsymbol{x}) = \int_{\mathcal{X}} \pi_t(\boldsymbol{x}, \boldsymbol{y})\mathrm{d}\boldsymbol{y}$. For these single particle marginals, we also have a marginalization theorem.

Theorem 5.2. *The single particle marginals satisfy the continuity equation*

$$\partial_t \rho_t + \nabla_{\boldsymbol{x}} \cdot (\rho_t \boldsymbol{v}_t) = 0 \tag{17}$$

Equivalently, it means $\boldsymbol{v}_t(\boldsymbol{x})$ transports μ_0 to μ_1. It can be viewed as the projection of the pair flow on the single particle space. However, it is worth clarifying that the two-particle dynamics generally cannot be recovered from the single-particle marginal dynamics. They are not equivalent.

Mean-field approximation. We further decouple the velocity $\boldsymbol{u}_t^1$ into two parts $\boldsymbol{u}_t^1(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{b}_t(\boldsymbol{x}) + \boldsymbol{f}_t(\boldsymbol{x}, \boldsymbol{y})$. $\boldsymbol{b}_t$ represents the self-driven term, and $\boldsymbol{f}_t$ represents the interaction term. Under this decomposition, we have

$$\boldsymbol{v}_t(\boldsymbol{x}) = \boldsymbol{b}_t(\boldsymbol{x}) + \int_{\mathcal{X}} \boldsymbol{f}_t(\boldsymbol{x}, \boldsymbol{y})\pi_t(\boldsymbol{y}|\boldsymbol{x})\mathrm{d}\boldsymbol{y} \tag{18}$$

The particle continuity equation (17) becomes

$$\partial_t \rho_t(\boldsymbol{x}) + \nabla_{\boldsymbol{x}} \cdot [\rho_t(\boldsymbol{x})\big(\boldsymbol{b}_t(\boldsymbol{x}) + \int_{\mathcal{X}} \boldsymbol{f}_t(\boldsymbol{x}, \boldsymbol{y})\pi_t(\boldsymbol{y}|\boldsymbol{x})\mathrm{d}\boldsymbol{y}\big)] = 0 \tag{19}$$

Under independence approximation $\pi_t(\boldsymbol{x}, \boldsymbol{y}) \approx \rho_t(\boldsymbol{x})\rho_t(\boldsymbol{y})$, (19) reduces to the mean-field flow.

$$\partial_t \rho_t(\boldsymbol{x}) + \nabla_{\boldsymbol{x}} \cdot [\rho_t(\boldsymbol{x})\big(\boldsymbol{b}_t(\boldsymbol{x}) + \int_{\mathcal{X}} \boldsymbol{f}_t(\boldsymbol{x}, \boldsymbol{y})\rho_t(\boldsymbol{y})\mathrm{d}\boldsymbol{y}\big)] = 0 \tag{20}$$

Though the independence may not hold accurately due to the particle interactions, the difference between the continuous measure path generated by the mean-field approximation (20) and the true dynamics (19) can be controlled by the difference between π_t and $\rho_t \otimes \rho_t$.

Proposition 5.3. *Let the true probability flow be* ρ_t *and the mean-field approximation be* $\hat{\rho}_t$. *Assume* $\boldsymbol{b}_t, \boldsymbol{f}_t$ *are both Lipschitz, and* $M = \max_{0<t\leq 1} \max_{\boldsymbol{x}} \mathcal{W}_1(\rho_t, \pi_t(\cdot|\boldsymbol{x}))$, *then* $\forall 0 < t \leq 1, \exists C_t > 0$ *such that* $\mathcal{W}_1(\rho_t, \hat{\rho}_t) \leq C_t M$.

5.2 TRAINING LOSS DESIGN

Marginal velocity pair. Analog to conditional flow matching (Lipman et al., 2023), when travelling pairs are tractable, we can parameterize two neural networks $\boldsymbol{u}_{\boldsymbol{\theta}}^i(\boldsymbol{x}, \boldsymbol{y}, t), i = 1, 2$ and regress the true marginal velocity pair by minimizing the conditional flow matching loss. This further leads us to the solution to dynamic QOT problem (10).

$$\mathcal{L}_{\text{pair}}(\boldsymbol{\theta}) = \mathbb{E}_{t\sim\mathcal{U}[0,1],(\boldsymbol{z},\boldsymbol{z}')\sim q(\boldsymbol{z},\boldsymbol{z}'),(\boldsymbol{x},\boldsymbol{y})\sim\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')} \sum_{i=1}^{2} \left\|\boldsymbol{u}_{\boldsymbol{\theta}}^i(\boldsymbol{x}, \boldsymbol{y}, t) - \boldsymbol{u}_t^i(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')\right\|_2^2 \tag{21}$$

Theorem 5.4. *The minimizer is $\boldsymbol{u}^{\boldsymbol{i}}_{\boldsymbol{\theta}}(\boldsymbol{x}, \boldsymbol{y}, t) = \boldsymbol{u}^i_t(\boldsymbol{x}, \boldsymbol{y})$. When $q = \gamma \otimes \gamma$ where γ is the optimal QOT coupling of (9), and the travelling pair used is the minimizer of (7), then the learned velocity pair generates the probability flow of the dynamic QOT problem (10).*

Particle velocity. A new design is that we can also obtain the particle velocity $\boldsymbol{v}_t(\boldsymbol{x})$ (16) by flow matching. We can parameterize a single neural network $\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x}, t)$ and try to minimize the intractable marginal regression loss $\mathcal{L}_M(\boldsymbol{\theta}) = \mathbb{E}_{t\sim\mathcal{U}[0,1],\boldsymbol{x}\sim\rho_t(\boldsymbol{x})}\|\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x}, t) - \boldsymbol{v}_t(\boldsymbol{x})\|^2$. The minimizer is $\boldsymbol{v}_t(\boldsymbol{x})$. A tractable conditional version is

$$\mathcal{L}_C(\boldsymbol{\theta}) = \mathbb{E}_{t\sim\mathcal{U}[0,1],(\boldsymbol{z},\boldsymbol{z}')\sim q(\boldsymbol{z},\boldsymbol{z}'),(\boldsymbol{x},\boldsymbol{y})\sim\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')}\|\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x}, t) - \boldsymbol{u}^1_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')\|^2 \tag{22}$$

Theorem 5.5. $\nabla_{\boldsymbol{\theta}}\mathcal{L}_M(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\mathcal{L}_C(\boldsymbol{\theta})$.

In practice, if the symmetry condition holds, i.e. $q = \gamma \otimes \gamma$ and $\mathcal{L}$ is symmetric to $(\boldsymbol{x}, \dot{\boldsymbol{x}}), (\boldsymbol{y}, \dot{\boldsymbol{y}})$, then it is easy to show that $\boldsymbol{u}^1_t(\boldsymbol{y}, \boldsymbol{x}|\boldsymbol{z}, \boldsymbol{z}') = \boldsymbol{u}^2_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')$. In this case, the conditional loss (22) can be replaced by a convex combination of itself and $\|\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{y}, t) - \boldsymbol{u}^2_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')\|^2$ to make the training more efficient. In most QOT settings, this condition obviously holds.

Velocity decomposition. If there are some gauges of velocity decomposition $\boldsymbol{u}^1_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}') = \boldsymbol{b}_t(\boldsymbol{x}) + \boldsymbol{f}_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')$, it is also possible to regress the interaction part $\boldsymbol{f}$ via flow matching. Let's parameterize a neural network $\boldsymbol{f}_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}, t)$ and minimize the conditional loss below.

$$\mathcal{L}_f(\boldsymbol{\phi}) = \mathbb{E}_{t\sim\mathcal{U}[0,1],(\boldsymbol{z},\boldsymbol{z}')\sim q(\boldsymbol{z},\boldsymbol{z}'),(\boldsymbol{x},\boldsymbol{y})\sim\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')}\|\boldsymbol{f}_{\boldsymbol{\phi}}(\boldsymbol{x}, \boldsymbol{y}, t) - \boldsymbol{f}_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')\|_2^2 \tag{23}$$

Theorem 5.6. *The minimizer of (23) is $\boldsymbol{f}_t(\boldsymbol{x}, \boldsymbol{y})$.*

An interesting point of the decomposed velocity is that one can use it to simulate the mean-field interaction dynamics. Consider N particles $X^i_0, i = 1, 2\cdots, N$ independently sampled from $\rho_0 = \mu_0$ following the ODE system $\mathrm{d}X^i_t = \Big(\boldsymbol{b}_t(X^i_t) + \frac{1}{N-1}\sum_{j\neq i}\boldsymbol{f}_t(X^i_t, X^j_t)\Big)\mathrm{d}t, i = 1, 2, \cdots, N$. When $N \to \infty$, this dynamics approximates the mean-field dynamics (20). Note that the velocity decomposition is not naturally unique, hence some gauges must be given. One possible decomposition is to first train a single-particle velocity field as $\boldsymbol{b}$ using TP-DATE, OT-CFM, or other methods, or to specify $\boldsymbol{b}$ directly based on domain specific prior knowledge. When only a conditional form of $\boldsymbol{b}$ is available, additional conditions are required to make the regression well defined; we discuss this case in E. Based on domain specific decomposition gauges, users may further exploit such a decomposition to provide more interpretable meanings for the components of the learned velocity.

6 Experiments

As a direct application of QOT, we demonstrate on both synthetic and real datasets that TP-DATE outperforms existing methods on spatiotemporal transcriptomics slice interpolation and continuous 3D reconstruction tasks. In this section, we use $\Phi = |\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{q}_t\||^2$ for space and $\Phi = 0$ for expression. Details of this Lagrangian choice can be found in B.3. We train TP-DATE under single-particle version (22). We use (F)GW-OT coupling for flow matching training, i.e. (F)GW-CFM, as baselines to further demonstrate the necessity of dynamic QOT against static QOT. See B.7 for details. Ablation and scaling study of TP-DATE are left to C.1, C.2.

TP-DATE preserves spatial structure during transport. We first use two synthetic datasets to verify that TP-DATE better preserves spatial structure during the transport. Each synthetic dataset consists of three cell types, and the ground truth spatial dynamics are defined by a 90° counterclockwise rigid rotation, so that the relative spatial structure of the data remains unchanged throughout the entire process. In the Rotation data, the ground truth expression dynamics keep gene expression unchanged, whereas in the Rotation + expression distractor setting (Distractor), a type dependent expression change is applied to each cell type. See B.4 for further details.

We conducted the standard hold one out experiment commonly used in the field to evaluate the dynamics reconstruction performance of different methods. That is, we trained the models on the initial and final time points and compared the model predictions at the intermediate unseen time point with the ground truth. As shown in Figure 1, on Rotation data, OT-CFM follows nearly straight paths to minimize the kinetic energy and shrinks the global structure, whereas TP-DATE recovers the more faithful curved rotation while better preserving the spatial structure.

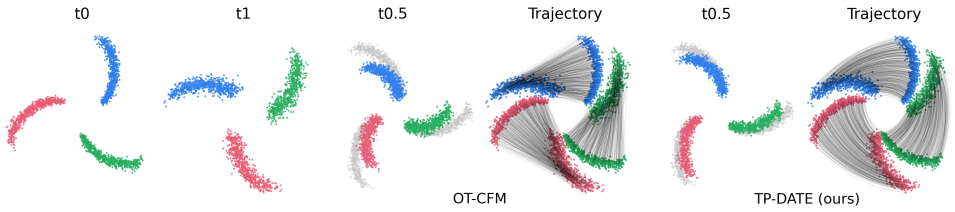


Figure 1: Trajectories on Rotation

Table 1: Hold one out experiments on synthetic data. The best performing results are in bold. We report mean and standard deviation over 5 random seeds.

Dataset	Method	Spatial MSE (↓)	Pair distortion (↓)	Balanced fused W_2 (↓)
Rotation	OT-CFM (Tong et al., 2024a)	0.10601±0.01460	0.33787±0.03021	11.34260±0.81057
	GW-CFM	0.57642±0.07041	0.62157±0.07289	23.21205±1.14792
	FGW-CFM	0.10579±0.02591	0.33472±0.04125	11.30984±1.50067
	stVCR (Peng et al., 2026a)	0.54023±0.01521	0.17179±0.01119	23.82269±0.23244
	CytoBridge (Zhang et al., 2025b)	0.09268±0.00318	0.33237±0.01992	10.62306±0.18430
	ContextFlow (Rathod et al., 2026)	0.23476±0.00584	0.28606±0.01139	17.01128±0.23455
	TP-DATE (ours)	**0.02565±0.00480**	**0.17174±0.01483**	**5.63983±0.54717**
Rotation + expr. distractor	OT-CFM	0.11015±0.00911	0.32524±0.01367	11.63383±0.50811
	GW-CFM	0.65267±0.06988	0.64477±0.04815	23.34253±0.74481
	FGW-CFM	0.10654±0.01524	0.31233±0.01226	11.43368±0.83093
	stVCR	0.52214±0.03327	**0.15765±0.01677**	23.50474±0.58836
	CytoBridge	0.09085±0.00343	0.31848±0.00244	10.54004±0.19780
	ContextFlow	0.24025±0.01084	0.29867±0.01964	17.22919±0.38236
	TP-DATE (ours)	**0.02269±0.00176**	0.16102±0.00654	**5.35946±0.20573**

Quantitatively, we use the Spatial MSE and pairwise distortion to assess how well the model recovers the intermediate spatial structure, and the fused Wasserstein distance to evaluate whether the model can recover the joint pattern of gene expression and spatial coordinates. See B.6 for further details. As shown in Table 1, TP-DATE gives the lowest spatial and fused errors, and low pair distortions on both datasets. These results show that TP-DATE can better preserve the spatial structure during transport. For these rotation datasets, to maintain the task non-trivial, we implemented stVCR (Peng et al., 2026a) without performing rigid body transformation only on them.

TP-DATE improves spatiotemporal dynamics reconstruction. We further evaluate TP-DATE by hold one out experiments on two real spatiotemporal transcriptomics datasets, Mouse brain (Chen et al., 2022) and ARTISTA (Wei et al., 2022). We use the Wasserstein distances in spatial coordinate and expression space, respectively, to evaluate whether the model prediction can accurately recover the spatial structure and gene expression pattern of the intermediate slice. To jointly account for spatial and expression information, we use the fused Wasserstein distance introduced above. In addition, we introduce the spatially coupled expression MSE (SC-eMSE) to evaluate the recovery of spatial gene expression patterns in the intermediate slice. Specifically, SC-eMSE first computes a spatial optimal transport coupling between the predicted and ground truth slice using squared distance in spatial coordinate space, and then evaluates the transport cost under this coupling using squared distance in expression space. Intuitively, this metric measures how different the gene expression profiles are between spatially corresponding locations in the two slices. See B.5 for dataset details, and B.6 for metric details.

On ARTISTA, although TP-DATE is less effective than the current state-of-the-art method stVCR (Peng et al., 2026a) in recovering the overall spatial shape of the tissue and performs comparably to the recently proposed ContextFlow (Rathod et al., 2026), it outperforms existing baselines on the other evaluation metrics and on the Mouse Brain dataset (Table 2). We emphasize that spatiotemporal interpolation requires not only recovering the overall spatial morphology of the tissue, but also reconstructing the spatial patterns of gene expression. TP-DATE performs better on the latter aspect. Therefore, these results suggest that TP-DATE improves spatiotemporal dynamics reconstruction.

TP-DATE allows continuous 3D reconstruction. Recently, Zhang et al. (2026a) released a spatial transcriptomics dataset containing slices collected from the same tumor tissue at the same time

Table 2: Hold one out experiments on spatiotemporal transcriptomics data. The best performing results are in bold. We report mean and standard deviation over 5 random seeds.

Dataset	Method	Spatial W_2 (↓)	Expression W_2 (↓)	SC-eMSE (↓)	Balanced fused W_2 (↓)
Mouse brain	OT-CFM	0.09190 ± 0.00495	6.52705 ± 0.04751	1.52747 ± 0.00994	9.91431 ± 0.19724
	GW-CFM	0.10666 ± 0.01233	6.55846 ± 0.04810	1.55450 ± 0.01966	10.58390 ± 0.51047
	FGW-CFM	0.09470 ± 0.00504	6.53585 ± 0.05995	1.53293 ± 0.02363	10.04189 ± 0.19298
	stVCR	0.09053 ± 0.00358	7.55942 ± 0.12372	1.96876 ± 0.03023	10.76074 ± 0.11473
	CytoBridge	0.09006 ± 0.00519	6.70320 ± 0.05889	1.61689 ± 0.02489	9.99210 ± 0.17831
	ContextFlow	0.09141 ± 0.00410	7.25501 ± 0.07919	1.87498 ± 0.03245	10.56965 ± 0.11675
	TP-DATE (ours)	$\mathbf{0.08678 \pm 0.00291}$	$\mathbf{6.40989 \pm 0.04644}$	$\mathbf{1.47693 \pm 0.02312}$	$\mathbf{9.59301 \pm 0.09255}$
ARTISTA	OT-CFM	0.10137 ± 0.00965	5.94656 ± 0.30203	1.03487 ± 0.08668	8.80314 ± 0.38687
	GW-CFM	0.10770 ± 0.00734	5.78681 ± 0.39641	0.98707 ± 0.10308	8.90625 ± 0.24148
	FGW-CFM	0.10001 ± 0.00739	5.97909 ± 0.31567	1.04710 ± 0.08661	8.78789 ± 0.16616
	stVCR	$\mathbf{0.08582 \pm 0.00509}$	7.42768 ± 0.49295	1.46830 ± 0.16836	9.42630 ± 0.39355
	CytoBridge	0.09833 ± 0.00827	7.00844 ± 0.54146	1.34401 ± 0.17710	9.49561 ± 0.35947
	ContextFlow	0.09174 ± 0.00663	7.32022 ± 0.49145	1.44565 ± 0.16502	9.53524 ± 0.38818
	TP-DATE (ours)	0.09196 ± 0.00500	$\mathbf{5.39310 \pm 0.28347}$	$\mathbf{0.87824 \pm 0.07587}$	$\mathbf{8.01158 \pm 0.27289}$

Table 3: Hold one out experiments on 3D reconstruction. The best performing results are in bold. We report mean and standard deviation over 5 random seeds.

Method	Spatial W_2 (↓)	Expression W_2 (↓)	SC-eMSE (↓)	Balanced fused W_2 (↓)
OT-CFM	0.07042 ± 0.00144	5.23581 ± 0.41570	1.15154 ± 0.08720	7.74659 ± 0.29841
GW-CFM	0.07731 ± 0.00789	5.20556 ± 0.27124	1.15201 ± 0.05978	7.90406 ± 0.35317
FGW-CFM	0.07459 ± 0.00802	5.05869 ± 0.14095	1.10336 ± 0.03139	7.70859 ± 0.27401
stVCR	0.07059 ± 0.00188	5.24894 ± 0.44683	1.13438 ± 0.09702	7.63873 ± 0.32859
CytoBridge	0.08103 ± 0.00191	5.08726 ± 0.39788	1.08467 ± 0.08165	7.76704 ± 0.22224
ContextFlow	$\mathbf{0.06768 \pm 0.00166}$	5.23844 ± 0.43624	1.13347 ± 0.09420	7.57907 ± 0.35204
TP-DATE (ours)	0.06959 ± 0.00151	$\mathbf{4.81955 \pm 0.13874}$	$\mathbf{1.04595 \pm 0.02881}$	$\mathbf{7.37480 \pm 0.08766}$

point but at different depths. The slices are ordered along the depth axis rather than time. Since high-throughput volumetric spatial transcriptomics remains experimentally challenging and is not routinely available for many tissues and platforms, continuous reconstruction of 3D spatial structure from serial tissue sections remains an important computational task. If we assume that the spatial structure and gene expression patterns of the same tissue vary smoothly across adjacent depths, the QOT framework can likewise be used to interpolate unseen depths, thereby reconstructing the three dimensional tissue structure from a small number of sections sampled at different depths. TP-DATE achieves performance comparable to the best baseline in terms of spatial Wasserstein distance, while outperforming all baselines on the other evaluation metrics (Table 3), demonstrating its superior performance on this new and important computational task.

7 Conclusion and discussion

We proposed TP-DATE, a dynamic QOT theory with a travelling pair flow matching framework for solving a family of dynamic QOT problems. Theoretically, we formulated dynamic QOT and included GW-OT, IGW-OT as special cases. Algorithmically, we developed a flow matching method that allows conditional paths to interact and used it to solve the dynamic QOT problem. Empirically, we quantitatively demonstrated the advantages of dynamic QOT on spatiotemporal dynamics reconstruction and continuous 3D reconstruction. We also theoretically offered a possibility to decompose the learned velocity into interpretable parts when specific domain knowledge is given.

QOT is a relatively new concept, yet some of its special cases, such as GW-OT, have already demonstrated substantial potential for applications (Mémoli, 2011; Vayer et al., 2020; Klein et al., 2025). Beyond the dynamic formulation studied in this work, an important direction is to investigate whether other extensions of OT can also be introduced to the QOT setting, such as stochastic or unbalanced QOT formulations. Future work can also try to combine the velocity decomposition paradigm with specific field knowledge for more interpretable dynamics modeling. Another promising direction is to consider more general interaction terms, or even to learn these interactions directly from data.

AI USE DISCLOSURE

In this work, we used generative AI tools to assist with translation and language polishing, and to help verify the correctness and technical details of the mathematical arguments and algorithms. We did not use generative AI tools to develop the theoretical framework, formulate the main mathematical results, construct the proofs, design or implement the algorithms, or write the scientific content of the paper. All AI-assisted content was independently checked by the authors, and all final scientific and editorial decisions were made by the authors. We take full responsibility for the final content of this work.

REFERENCES

Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*. Lectures in Mathematics. ETH Zürich. Birkhäuser Basel, 2 edition, 2008. ISBN 978-3-7643-8721-1. doi: 10.1007/978-3-7643-8722-8.

Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84:375–393, 2000. doi: 10.1007/s002110050002.

John P. Bryan, Samouil L. Farhi, and Brian Cleary. Accurate trajectory inference in time-series spatial transcriptomics with structurally-constrained optimal transport. *Nature Communications*, 17:8075, 2026. doi: 10.1038/s41467-026-74927-8.

Ao Chen, Sha Liao, Mengnan Cheng, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using dna nanoball-patterned arrays. *Cell*, 185(10):1777–1792, 2022. ISSN 0092-8674. doi: https://doi.org/10.1016/j.cell.2022.04.003. URL https://www.sciencedirect.com/science/article/pii/S0092867422003993.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. An interpolating distance between optimal transport and fisher–rao metrics. *Foundations of Computational Mathematics*, 18(1):1–44, 2018a.

Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Unbalanced optimal transport: Dynamic and kantorovich formulations. *Journal of Functional Analysis*, 274(11): 3090–3123, 2018b.

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 17695–17709. Curran Associates, Inc., 2021.

Stefan Elfwing, Eiji Uchibe, and Kenji Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks*, 107:3–11, 2018. ISSN 0893-6080. Special issue on deep reinforcement learning.

Luca Eyring, Dominik Klein, Théo Uscidda, Giovanni Palla, Niki Kilbertus, Zeynep Akata, and Fabian J Theis. Unbalancedness in neural monge maps improves unpaired domain translation. In *The Twelfth International Conference on Learning Representations*, 2024.

Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, et al. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.

Peter Halmos, Xinhao Liu, Julian Gold, Feng Chen, Li Ding, and Benjamin J. Raphael. Destot: Alignment of spatiotemporal transcriptomics data. *Cell Systems*, 16(2):101160, 2025. doi: 10.1016/j.cels.2024.12.001.

Leonid V Kantorovich. On the translocation of masses. *Management Science*, 5(1):1–4, 1958. ISSN 00251909, 15265501.

Tanguy Kerdoncuff, Rémi Emonet, and Marc Sebban. Sampled gromov wasserstein. *Machine Learning*, 110:2151–2186, 2021. doi: 10.1007/s10994-021-06035-1.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. URL https://arxiv.org/abs/1412.6980.

Dominik Klein, Théo Uscidda, Fabian J Theis, and Marco cuturi. GENOT: Entropic (gromov) wasserstein flow matching with applications to single-cell genomics. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=hjspWd7jvg.

Dominik Klein, Giovanni Palla, Marius Lange, et al. Mapping cells through time and space with moscot. *Nature*, 638:1065–1075, 2025. doi: 10.1038/s41586-024-08453-2.

Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal entropy-transport problems and a new hellinger–kantorovich distance between positive measures. *Inventiones mathematicae*, 211: 969–1117, 2018.

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

Stefano Lisini. Characterization of absolutely continuous curves in wasserstein spaces. *Calculus of Variations and Partial Differential Equations*, 28:85–120, 2007. doi: 10.1007/s00526-006-0032-2.

Xinhao Liu, Ron Zeira, and Benjamin J. Raphael. Partial alignment of multislice spatially resolved transcriptomics data. *Genome Research*, 33(7):1124–1132, 2023. doi: 10.1101/gr.277670.123.

Christian Léonard. A survey of the schrödinger problem and some of its connections with optimal transport. *Discrete and Continuous Dynamical Systems*, 34(4):1533–1574, 2014. ISSN 1078-0947. doi: 10.3934/dcds.2014.34.1533.

Facundo Mémoli. Gromov–wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11:417–487, 2011. doi: 10.1007/s10208-011-9093-5.

Q. Peng, P. Zhou, and T. Li. stvcr: spatiotemporal dynamics of single cells. *Nature Methods*, 23: 542–553, 2026a. doi: 10.1038/s41592-026-03010-3.

Qiangwei Peng, Zihan Wang, Junda Ying, Yuhao Sun, Qing Nie, Lei Zhang, Tiejun Li, and Peijie Zhou. WFR-FM: Simulation-free dynamic unbalanced optimal transport. In *The Fourteenth International Conference on Learning Representations*, 2026b. URL https://openreview.net/forum?id=1nqu7bK1mm.

Santanu Subhash Rathod, Francesco Ceccarelli, Sean B. Holden, Pietro Liò, Xiao Zhang, and Jovan Tanevski. Contextflow: Context-aware flow matching for trajectory inference from spatial omics data, 2026. URL https://arxiv.org/abs/2510.02952.

Kristiyan Sakalyan, Alessandro Palma, Filippo Guerranti, Fabian J Theis, and Stephan Günnemann. Modeling microenvironment trajectories on spatial transcriptomics with nicheflow. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=5ofJyjgrth.

Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176:928–943, 2019. doi: 10.1016/j.cell.2019.01.006.

Erwin Schrödinger. Sur la théorie relativiste de l'électron et l'interprétation de la mécanique quantique. *Annales de l'Institut Henri Poincaré*, 2(4):269–310, 1932.

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

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schrödinger bridge matching. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 62183–62223. Curran Associates, Inc., 2023.

Alexander Tong, Jessie Huang, Guy Wolf, David Van Dijk, and Smita Krishnaswamy. TrajectoryNet: A dynamic optimal transport network for modeling cellular dynamics. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 9526–9536. PMLR, 13–18 Jul 2020.

Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*, pp. 1–34, 2024a. ISSN 2835-8856.

Alexander Y. Tong, Nikolay Malkin, Kilian Fatras, Lazar Atanackovic, Yanlei Zhang, Guillaume Huguet, Guy Wolf, and Yoshua Bengio. Simulation-free Schrödinger bridges via score and flow matching. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 1279–1287. PMLR, 02–04 May 2024b.

V. A. Traag, L. Waltman, and N. J. van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific Reports*, 9:5233, 2019. doi: 10.1038/s41598-019-41695-z.

Titouan Vayer, Laetitia Chapel, Remi Flamary, Romain Tavenard, and Nicolas Courty. Fused gromov-wasserstein distance for structured objects. *Algorithms*, 13(9), 2020. ISSN 1999-4893. doi: 10.3390/a13090212. URL https://www.mdpi.com/1999-4893/13/9/212.

Ruodu Wang and Zhenyuan Zhang. Quadratic-form optimal transport. *Mathematical Programming*, 2025. doi: 10.1007/s10107-025-02282-5.

Xiaoyu Wei, Sulei Fu, Hanbo Li, et al. Single-cell stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science*, 377(6610):eabp9444, 2022. doi: 10.1126/science.abp9444. URL https://www.science.org/doi/abs/10.1126/science.abp9444.

Junda Ying, Yuxuan Wang, Bowen Yang, Peijie Zhou, and Lei Zhang. Beyond continuity: Simulation-free reconstruction of discrete branching dynamics from single-cell snapshots. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=IABjbJlinz.

Ron Zeira, Max Land, Alexander Strzalkowski, and Benjamin J. Raphael. Alignment and integration of spatial transcriptomics data. *Nature Methods*, 19:567–575, 2022. doi: 10.1038/s41592-022-01459-6.

H. Zhang, Z. Zhang, P. Wang, T. Xu, X. Chen, Y. Zhao, S. Lin, W. Cai, P. Ren, C. Luo, P. Zhang, Y. Wang, S. Hou, Y. Zhao, H. Zeng, Z. Liu, C. Wang, Z. Gao, Y. Feng, D. Pan, and Z. Zeng. Uncovering spatially resolved functional genomics with crispr screen sequencing. *Cell*, 189: 4594–4618, 2026a. doi: 10.1016/j.cell.2026.04.049.

Zhengxin Zhang, Ziv Goldfeld, Kristjan Greenewald, et al. Gradient flows and riemannian structure in the gromov-wasserstein geometry. *Foundations of Computational Mathematics*, 26:1911–2003, 2026b. doi: 10.1007/s10208-025-09722-w.

Zhenyi Zhang, Tiejun Li, and Peijie Zhou. Learning stochastic dynamics from snapshots through regularized unbalanced optimal transport. In *The Thirteenth International Conference on Learning Representations*, 2025a. URL https://openreview.net/forum?id=gQlxd3Mtru.

Zhenyi Zhang, Zihan Wang, Yuhao Sun, Tiejun Li, and Peijie Zhou. Modeling cell dynamics and interactions with unbalanced mean field schrödinger bridge. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025b. URL https://openreview.net/forum?id=Z6DJJIN8IJ.

A PROOFS

Proofs of theorems and propositions.

A.1 PROOF OF THEOREM 4.1

Theorem 4.1. *If $\mathcal{L}$ is convex w.r.t $(\dot{\boldsymbol{x}}, \dot{\boldsymbol{y}})$, then $QOT_S(\mu_0, \mu_1) = QOT_D(\mu_0, \mu_1) \geq QOT_{BB}(\mu_0, \mu_1)$.*

Proof. For convenience, we restate the three forms below. The static form is

$$\mathrm{QOT}_S(\mu_0, \mu_1) = \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\mathcal{X}^4} \mathcal{A}(\boldsymbol{z}, \boldsymbol{z}')\gamma(\boldsymbol{z})\gamma(\boldsymbol{z}')\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'. \tag{24}$$

The dynamic form is

$$\begin{aligned}&\mathrm{QOT}_D(\mu_0, \mu_1) = \\ &\inf \int_0^1 \int \int_{\mathcal{X}^2} \mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, u_t^1(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}'), u_t^2(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}'))\pi_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')\gamma(\boldsymbol{z})\gamma(\boldsymbol{z}')\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'\mathrm{d}t.\end{aligned} \tag{25}$$

The BB-form is

$$\begin{aligned}&\mathrm{QOT}_{BB}(\mu_0, \mu_1) = \inf_{\pi, \boldsymbol{u}^1, \boldsymbol{u}^2} \int_0^1 \int_{\mathcal{X}^2} \mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, u_t^1(\boldsymbol{x}, \boldsymbol{y}), u_t^2(\boldsymbol{x}, \boldsymbol{y}))\pi_t(\boldsymbol{x}, \boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}t \\ &\partial_t \pi_t(\boldsymbol{x}, \boldsymbol{y}) + \nabla_{\boldsymbol{x}} \cdot (\pi_t(\boldsymbol{x}, \boldsymbol{y})u_t^1(\boldsymbol{x}, \boldsymbol{y})) + \nabla_{\boldsymbol{y}} \cdot (\pi_t(\boldsymbol{x}, \boldsymbol{y})u_t^2(\boldsymbol{x}, \boldsymbol{y})) = 0. \\ &\pi_0 = \mu_0 \otimes \mu_0, \qquad \pi_1 = \mu_1 \otimes \mu_1\end{aligned} \tag{26}$$

For any $(\boldsymbol{z}, \boldsymbol{z}')$, the optimal path action $\mathcal{A}(\boldsymbol{z}, \boldsymbol{z}')$ is attained by the travelling pair $\dot{\boldsymbol{x}}_t = \boldsymbol{u}_t^1(\boldsymbol{x}_t, \boldsymbol{y}_t|\boldsymbol{z}, \boldsymbol{z}'), \dot{\boldsymbol{y}}_t = \boldsymbol{u}_t^2(\boldsymbol{x}_t, \boldsymbol{y}_t|\boldsymbol{z}, \boldsymbol{z}')$.

$$\begin{aligned}\mathcal{A}(\boldsymbol{z}, \boldsymbol{z}') &= \inf_{\boldsymbol{x}_t, \boldsymbol{y}_t} \int_0^1 \mathcal{L}(t, \boldsymbol{x}_t, \boldsymbol{y}_t, \dot{\boldsymbol{x}}_t, \dot{\boldsymbol{y}}_t)\mathrm{d}t \\ &= \int_0^1 \mathcal{L}(t, \boldsymbol{x}_t, \boldsymbol{y}_t, \boldsymbol{u}_t^1(\boldsymbol{x}_t, \boldsymbol{y}_t|\boldsymbol{z}, \boldsymbol{z}'), \boldsymbol{u}_t^2(\boldsymbol{x}_t, \boldsymbol{y}_t|\boldsymbol{z}, \boldsymbol{z}'))\mathrm{d}t \\ &= \int_0^1 \int_{\mathcal{X}^2} \mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}_t^1(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}'), \boldsymbol{u}_t^2(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}'))\pi_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}t\end{aligned} \tag{27}$$

where $\pi_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')$ is the conditional Dirac probability path generated by the travelling pair velocity, which is

$$\begin{aligned}&\partial_t \pi_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}') + \nabla_{\boldsymbol{x}} \cdot (\pi_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')\boldsymbol{u}_t^1(\boldsymbol{x}_t, \boldsymbol{y}_t|\boldsymbol{z}, \boldsymbol{z}')) + \nabla_{\boldsymbol{y}} \cdot (\pi_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')\boldsymbol{u}_t^2(\boldsymbol{x}_t, \boldsymbol{y}_t|\boldsymbol{z}, \boldsymbol{z}')) = 0 \\ &\pi_0(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}') = \delta_{(\boldsymbol{x}_0, \boldsymbol{y}_0)}(\boldsymbol{x}, \boldsymbol{y}), \quad \pi_1(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}') = \delta_{(\boldsymbol{x}_1, \boldsymbol{y}_1)}(\boldsymbol{x}, \boldsymbol{y})\end{aligned} \tag{28}$$

Therefore, $\mathrm{QOT}_D(\mu_0, \mu_1) = \mathrm{QOT}_S(\mu_0, \mu_1)$.

For any admissible coupling γ,

$$\begin{aligned}&\int_{\mathcal{X}^4} \mathcal{A}(\boldsymbol{z}, \boldsymbol{z}')\gamma(\boldsymbol{z})\gamma(\boldsymbol{z}')\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}' \\ &= \int \mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}_t^1(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}'), \boldsymbol{u}_t^2(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}'))\pi_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')\gamma(\boldsymbol{z})\gamma(\boldsymbol{z}')\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}t\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}' \\ &= \int \Big(\int \mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}_t^1(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}'), \boldsymbol{u}_t^2(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}'))\frac{\pi_t(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}')\gamma(\boldsymbol{z})\gamma(\boldsymbol{z}')}{\pi_t(\boldsymbol{x}, \boldsymbol{y})}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'\Big)\pi_t(\boldsymbol{x}, \boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}t \\ &\geq \int \mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}_t^1(\boldsymbol{x}, \boldsymbol{y}), \boldsymbol{u}_t^2(\boldsymbol{x}, \boldsymbol{y}))\pi_t(\boldsymbol{x}, \boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}t \qquad \text{(Jensen's inequality)} \\ &\geq \mathrm{QOT}_{BB}(\mu_0, \mu_1) \qquad \text{(Take infimum)}\end{aligned} \tag{29}$$

By taking infimum w.r.t γ, we have $\mathrm{QOT}_S(\mu_0, \mu_1) \geq \mathrm{QOT}_{BB}(\mu_0, \mu_1)$. □

A.2 Proof of Proposition 4.2

Proposition 4.2. *The optimal $\boldsymbol{c}_t$ is $\boldsymbol{c}_t = (1-t)\boldsymbol{c}_0 + t\boldsymbol{c}_1$. If $\Phi = \Phi(\|\boldsymbol{q}\|)$, $\boldsymbol{q}_0 \nparallel \boldsymbol{q}_1$, then $\boldsymbol{q}_t \in$ span$\{\boldsymbol{q}_0, \boldsymbol{q}_1\}$.*

Proof. Recall the Lagrangian

$$\mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, \dot{\boldsymbol{x}}, \dot{\boldsymbol{y}}) = \|\dot{\boldsymbol{c}}_t\|^2 + \frac{1}{4}\|\dot{\boldsymbol{q}}_t\|^2 + \lambda\Phi(\boldsymbol{q}, \dot{\boldsymbol{q}}). \tag{30}$$

The Euler-Lagrange equation is

$$\begin{cases} 2\ddot{\boldsymbol{c}}_t = \boldsymbol{0} & (\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathcal{L}}{\partial\dot{\boldsymbol{c}}} = \frac{\partial\mathcal{L}}{\partial\boldsymbol{c}}) \\ \frac{1}{2}\ddot{\boldsymbol{q}}_t + \lambda\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\Phi}{\partial\dot{\boldsymbol{q}}} = \lambda\frac{\partial\Phi}{\partial\boldsymbol{q}} & (\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\mathcal{L}}{\partial\dot{\boldsymbol{q}}} = \frac{\partial\mathcal{L}}{\partial\boldsymbol{q}}) \end{cases} \tag{31}$$

The first equation yields $\boldsymbol{c}_t = (1-t)\boldsymbol{c}_0 + t\boldsymbol{c}_1$. When Φ is a radial function of $\boldsymbol{q}$, the second equation reduces to

$$\frac{1}{2}\ddot{\boldsymbol{q}}_t = \lambda\frac{\partial\Phi}{\partial\boldsymbol{q}} = \lambda\Phi'(\|\boldsymbol{q}_t\|)\frac{\boldsymbol{q}_t}{\|\boldsymbol{q}_t\|} \tag{32}$$

That means the acceleration $\ddot{\boldsymbol{q}}_t$ is parallel to $\boldsymbol{q}_t$, and hence the angular momentum is conserved. Therefore, $\boldsymbol{q}_t$ is constrained on a plane. When $\boldsymbol{q}_0 \nparallel \boldsymbol{q}_1$, we have $\boldsymbol{q}_t \in \text{span}\{\boldsymbol{q}_0, \boldsymbol{q}_1\}$. □

A.3 Proof of Proposition 4.3

Proposition 4.3. *If $\Phi = |\frac{\mathrm{d}}{\mathrm{d}t}\phi(\boldsymbol{q}_t)|^2$, Φ is convex w.r.t $\dot{\boldsymbol{q}}_t$. If $\Phi = |\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{q}_t\||^2$, $\boldsymbol{q}_t$ has analytic solution.*

Proof. By direct calculation,

$$\Phi = |\frac{\mathrm{d}}{\mathrm{d}t}\phi(\boldsymbol{q}_t)|^2 = |\nabla_{\boldsymbol{q}}\phi(\boldsymbol{q}_t)^{\mathrm{T}}\dot{\boldsymbol{q}}_t|^2 = \dot{\boldsymbol{q}}_t^{\mathrm{T}}[\nabla_{\boldsymbol{q}}\phi(\boldsymbol{q}_t)\nabla_{\boldsymbol{q}}\phi(\boldsymbol{q}_t)^{\mathrm{T}}]\dot{\boldsymbol{q}}_t \tag{33}$$

Therefore, it is convex w.r.t $\dot{\boldsymbol{q}}_t$. When $\Phi = |\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{q}_t\||^2$, the action of $(\boldsymbol{q}, \dot{\boldsymbol{q}})$ is

$$\int_0^1 \frac{1}{4}\|\dot{\boldsymbol{q}}_t\|^2 + \lambda|\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{q}_t\||^2\mathrm{d}t. \tag{34}$$

The corresponding Euler-Lagrange equation is

$$\frac{1}{2}\ddot{\boldsymbol{q}}_t + \lambda\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\Phi}{\partial\dot{\boldsymbol{q}}} = \lambda\frac{\partial\Phi}{\partial\boldsymbol{q}}. \tag{35}$$

By direct calculation, we have

$$\frac{\partial\Phi}{\partial\dot{\boldsymbol{q}}} = \frac{2\boldsymbol{q}_t^{\mathrm{T}}\dot{\boldsymbol{q}}_t}{\|\boldsymbol{q}_t\|^2}\boldsymbol{q}_t, \qquad \frac{\partial\Phi}{\partial\boldsymbol{q}} = \frac{2\boldsymbol{q}_t^{\mathrm{T}}\dot{\boldsymbol{q}}_t}{\|\boldsymbol{q}_t\|^2}\dot{\boldsymbol{q}}_t - \frac{2(\boldsymbol{q}_t^{\mathrm{T}}\dot{\boldsymbol{q}}_t)^2}{\|\boldsymbol{q}_t\|^4}\boldsymbol{q}_t. \tag{36}$$

Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\Phi}{\partial\dot{\boldsymbol{q}}} = 2\Big(\frac{\|\dot{\boldsymbol{q}}_t\|^2 + \boldsymbol{q}_t^{\mathrm{T}}\ddot{\boldsymbol{q}}_t}{\|\boldsymbol{q}_t\|^2} - 2\frac{(\boldsymbol{q}_t^{\mathrm{T}}\dot{\boldsymbol{q}}_t)^2}{\|\boldsymbol{q}_t\|^4}\Big)\boldsymbol{q}_t + \frac{2\boldsymbol{q}_t^{\mathrm{T}}\dot{\boldsymbol{q}}_t}{\|\boldsymbol{q}_t\|^2}\dot{\boldsymbol{q}}_t. \tag{37}$$

The Euler-Lagrange equation reduces to

$$\frac{1}{2}\ddot{\boldsymbol{q}}_t + 2\lambda\Big(\frac{\|\dot{\boldsymbol{q}}_t\|^2 + \boldsymbol{q}_t^{\mathrm{T}}\ddot{\boldsymbol{q}}_t}{\|\boldsymbol{q}_t\|^2} - \frac{(\boldsymbol{q}_t^{\mathrm{T}}\dot{\boldsymbol{q}}_t)^2}{\|\boldsymbol{q}_t\|^4}\Big)\boldsymbol{q}_t = \boldsymbol{0}. \tag{38}$$

This leads to the conservation of angular momentum, therefore $\boldsymbol{q}_t$ is again constrained on the plane spanned by $\boldsymbol{q}_0, \boldsymbol{q}_1$. For convenience, we define $r_t = \|\boldsymbol{q}_t\|, \alpha = \arccos\frac{\langle\boldsymbol{q}_0, \boldsymbol{q}_1\rangle}{\|\boldsymbol{q}_0\|\|\boldsymbol{q}_1\|}$, and the polar coordinates on that plane.

$$\boldsymbol{q}_0 = r_0(1, 0), \qquad \boldsymbol{q}_1 = r_1(\cos\alpha, \sin\alpha), \qquad \boldsymbol{q}_t = r_t(\cos\theta_t, \sin\theta_t) \tag{39}$$

Under this representation, $\dot{\boldsymbol{q}}_t = \dot{r}_t(\cos\theta_t, \sin\theta_t) + r_t\dot{\theta}_t(-\sin\theta_t, \cos\theta_t)$, $|\dot{\boldsymbol{q}}_t|^2 = \dot{r}_t^2 + r_t^2\dot{\theta}_t^2$. The action (34) becomes

$$\int_0^1 \frac{1}{4} r_t^2\dot{\theta}_t^2 + (\frac{1}{4} + \lambda)\dot{r}_t^2 \mathrm{d}t. \tag{40}$$

Write $a = \frac{1}{4} + \lambda, k = \sqrt{\frac{1}{1+4\lambda}} \in (0, 1]$, and define $\boldsymbol{w}_t = \sqrt{a} r_t(\cos k\theta_t, \sin k\theta_t)$, we have

$$\|\dot{\boldsymbol{w}}_t\|^2 = a(\dot{r}_t^2 + k^2 r_t^2\dot{\theta}_t^2) = \frac{1}{4} r_t^2\dot{\theta}_t^2 + (\frac{1}{4} + \lambda)\dot{r}_t^2. \tag{41}$$

Therefore, the minimization problem reduces to

$$\inf_{\boldsymbol{w}_t} \int_0^1 \|\dot{\boldsymbol{w}}_t\|^2 \mathrm{d}t \quad \text{s.t.} \quad \boldsymbol{w}_0 = \sqrt{a} r_0(1, 0), \boldsymbol{w}_1 = \sqrt{a} r_1(\cos k\alpha, \sin k\alpha). \tag{42}$$

Since $\alpha \in (0, \pi)$ and $k \in (0, 1]$, the optimal $\boldsymbol{w}_t$ is the displacement interpolation

$$\boldsymbol{w}_t = (1-t)\boldsymbol{w}_0 + t\boldsymbol{w}_1 = \sqrt{a}((1-t)r_0 + tr_1\cos k\alpha, tr_1 \sin k\alpha). \tag{43}$$

By direct calculation, we have

$$\|\dot{\boldsymbol{w}}_t\|^2 = a(r_0^2 + r_1^2 - 2r_0 r_1 \cos k\alpha). \tag{44}$$

Therefore, given the conditional variables $\boldsymbol{z} = (\boldsymbol{x}_0, \boldsymbol{x}_1), \boldsymbol{z}' = (\boldsymbol{y}_0, \boldsymbol{y}_1)$, we can obtain the analytic solution of the travelling pair path and the path action via (43,44). We first calculate $\boldsymbol{c}_0 = \frac{\boldsymbol{x}_0 + \boldsymbol{y}_0}{2}, \boldsymbol{c}_1 = \frac{\boldsymbol{x}_1 + \boldsymbol{y}_1}{2}, \boldsymbol{q}_0 = \boldsymbol{x}_0 - \boldsymbol{y}_0, \boldsymbol{q}_1 = \boldsymbol{x}_1 - \boldsymbol{y}_1$ and $r_0 = \|\boldsymbol{q}_0\|, r_1 = \|\boldsymbol{q}_1\|, \alpha = \arccos \frac{\langle \boldsymbol{q}_0, \boldsymbol{q}_1\rangle}{\|\boldsymbol{q}_0\|\|\boldsymbol{q}_1\|}$. Then we get the analytic path action

$$\mathcal{A}(\boldsymbol{z}, \boldsymbol{z}') = \|\boldsymbol{c}_1 - \boldsymbol{c}_0\|^2 + (\frac{1}{4} + \lambda)(r_0^2 + r_1^2 - 2r_0 r_1 \cos k\alpha). \tag{45}$$

Then, we calculate the analytic solution of the polar coordinates.

$$\begin{cases} r_t = \|\boldsymbol{w}_t\|/\sqrt{a} = \sqrt{(1-t)^2 r_0^2 + t^2 r_1^2 + 2t(1-t)r_0 r_1 \cos k\alpha} \\ \theta_t = \mathrm{Arg}(\boldsymbol{w}_t)/k = \sqrt{1+4\lambda} \cdot \mathrm{Arg}\Big((1-t)r_0 + tr_1\cos k\alpha + i \cdot tr_1 \sin k\alpha\Big) \end{cases} \tag{46}$$

The analytic solution of C-Q decomposition can be further calculated.

$$\begin{cases} \boldsymbol{c}_t = (1-t)\boldsymbol{c}_0 + t\boldsymbol{c}_1 \\ \boldsymbol{q}_t = r_t(\cos\theta_t, \sin\theta_t) \end{cases} \tag{47}$$

The analytic solution of the travelling pair is

$$\begin{cases} \boldsymbol{x}_t = \boldsymbol{c}_t + \frac{1}{2}\boldsymbol{q}_t \\ \boldsymbol{y}_t = \boldsymbol{c}_t - \frac{1}{2}\boldsymbol{q}_t \end{cases}, \qquad \begin{cases} \dot{\boldsymbol{x}}_t = \boldsymbol{c}_1 - \boldsymbol{c}_0 + \frac{1}{2}\dot{\boldsymbol{q}}_t \\ \dot{\boldsymbol{y}}_t = \boldsymbol{c}_1 - \boldsymbol{c}_0 - \frac{1}{2}\dot{\boldsymbol{q}}_t \end{cases} \tag{48}$$

□

A.4 Proof of Theorem 4.4

Theorem 4.4 *$\Phi = |\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{q}_t\||^2$. When $\lambda = 0$ and $d \geq 2$, the corresponding static QOT reduces to standard OT, while when $\lambda \to +\infty$, it reduces to standard GW-OT: λ^{-1}QOT$_S(\mu_0, \mu_1) \to$ GW-OT(μ_0, μ_1).*

Proof. See D.1 for details. □

A.5 Proof of Theorem 5.1

Theorem 5.1. *Define the marginal velocity as* $\boldsymbol{u}_t^i(\boldsymbol{x},\boldsymbol{y}) = \int \boldsymbol{u}_t^i(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')\frac{\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')q(\boldsymbol{z},\boldsymbol{z}')}{\pi_t(\boldsymbol{x},\boldsymbol{y})}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'$, *the marginals satisfy the marginal continuity equation.*

$$\partial_t \pi_t + \nabla_{\boldsymbol{x}} \cdot (\pi_t \boldsymbol{u}_t^1) + \nabla_{\boldsymbol{y}} \cdot (\pi_t \boldsymbol{u}_t^2) = 0 \tag{49}$$

If $q = \gamma \otimes \gamma$ *for some* $\gamma \in \Pi(\mu_0,\mu_1)$*, the marginal velocity transports* $\mu_0 \otimes \mu_0$ *to* $\mu_1 \otimes \mu_1$.

Proof. Consider the probability path $\pi_t(\boldsymbol{x},\boldsymbol{y})$ on the two-particle augmented space $\mathcal{X}^2$, the corresponding velocity is $\boldsymbol{u}_t(\boldsymbol{x},\boldsymbol{y}) = \begin{pmatrix}\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y})\\ \boldsymbol{u}_t^2(\boldsymbol{x},\boldsymbol{y})\end{pmatrix}$. Applying the marginalization theorem of standard conditional flow matching (Lipman et al., 2023; Tong et al., 2024a), we have

$$\partial_t \pi_t + \nabla_{(\boldsymbol{x},\boldsymbol{y})} \cdot (\pi_t \boldsymbol{u}_t) = 0 \tag{50}$$

By definition, $\nabla_{(\boldsymbol{x},\boldsymbol{y})} \cdot (\pi_t \boldsymbol{u}_t) = \nabla_{\boldsymbol{x}} \cdot (\pi_t \boldsymbol{u}_t^1) + \nabla_{\boldsymbol{y}} \cdot (\pi_t \boldsymbol{u}_t^2)$. Therefore, identity (49) holds. If we further have $q = \gamma \otimes \gamma$ for some $\gamma \in \Pi(\mu_0,\mu_1)$, then by definition, we have

$$\begin{aligned}
\pi_0(\boldsymbol{x},\boldsymbol{y}) &= \int \pi_0(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')q(\boldsymbol{z},\boldsymbol{z}')\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}' \\
&= \int \delta_{(\boldsymbol{x}_0,\boldsymbol{y}_0)}(\boldsymbol{x},\boldsymbol{y})\gamma(\boldsymbol{x}_0,\boldsymbol{x}_1)\gamma(\boldsymbol{y}_0,\boldsymbol{y}_1)\mathrm{d}\boldsymbol{x}_0\mathrm{d}\boldsymbol{x}_1\mathrm{d}\boldsymbol{y}_0\mathrm{d}\boldsymbol{y}_1 \\
&= \int \delta_{\boldsymbol{x}_0}(\boldsymbol{x})\gamma(\boldsymbol{x}_0,\boldsymbol{x}_1)\mathrm{d}\boldsymbol{x}_0\mathrm{d}\boldsymbol{x}_1 \int \delta_{\boldsymbol{y}_0}(\boldsymbol{y})\gamma(\boldsymbol{y}_0,\boldsymbol{y}_1)\mathrm{d}\boldsymbol{y}_0\mathrm{d}\boldsymbol{y}_1 \\
&= \mu_0(\boldsymbol{x})\mu_0(\boldsymbol{y})
\end{aligned} \tag{51}$$

where $\boldsymbol{z} = (\boldsymbol{x}_0,\boldsymbol{x_1}), \boldsymbol{z}' = (\boldsymbol{y}_0,\boldsymbol{y}_1)$. Similarly, $\pi_1(\boldsymbol{x},\boldsymbol{y}) = \mu_1(\boldsymbol{x})\mu_1(\boldsymbol{y})$. Therefore, the marginal velocity transports $\mu_0 \otimes \mu_0$ to $\mu_1 \otimes \mu_1$. □

A.6 Proof of Theorem 5.2

Theorem 5.2. *The single particle marginals satisfy the continuity equation*

$$\partial_t \rho_t + \nabla_{\boldsymbol{x}} \cdot (\rho_t \boldsymbol{v}_t) = 0 \tag{52}$$

Proof. Recall the definitions

$$\rho_t(\boldsymbol{x}) = \int_{\mathcal{X}} \pi_t(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{y}, \quad \boldsymbol{v}_t(\boldsymbol{x}) = \int_{\mathcal{X}} \boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y})\pi_t(\boldsymbol{y}|\boldsymbol{x})\mathrm{d}\boldsymbol{y} = \int_{\mathcal{X}} \boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y})\frac{\pi_t(\boldsymbol{x},\boldsymbol{y})}{\rho_t(\boldsymbol{x})}\mathrm{d}\boldsymbol{y}. \tag{53}$$

By direct calculation, we have

$$\begin{aligned}
\partial_t \rho_t(\boldsymbol{x}) &= \partial_t \int_{\mathcal{X}} \pi_t(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{y} = \int_{\mathcal{X}} \partial_t \pi_t(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{y} \\
&= -\int_{\mathcal{X}} \nabla_{\boldsymbol{x}} \cdot (\pi_t(\boldsymbol{x},\boldsymbol{y})\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y})) + \nabla_{\boldsymbol{y}} \cdot (\pi_t(\boldsymbol{x},\boldsymbol{y})\boldsymbol{u}_t^2(\boldsymbol{x},\boldsymbol{y}))\mathrm{d}\boldsymbol{y} \\
&= -\nabla_{\boldsymbol{x}} \cdot \int_{\mathcal{X}} \pi_t(\boldsymbol{x},\boldsymbol{y})\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{y} - \rho_t(\boldsymbol{x}) \int_{\mathcal{X}} \nabla_{\boldsymbol{y}} \cdot \Big(\pi_t(\boldsymbol{y}|\boldsymbol{x})\boldsymbol{u}_t^2(\boldsymbol{x},\boldsymbol{y})\Big)\mathrm{d}\boldsymbol{y} \\
&= -\nabla_{\boldsymbol{x}} \cdot \int_{\mathcal{X}} \pi_t(\boldsymbol{x},\boldsymbol{y})\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{y} \\
&= -\nabla_{\boldsymbol{x}} \cdot \Big(\rho_t(\boldsymbol{x}) \int_{\mathcal{X}} \frac{\pi_t(\boldsymbol{x},\boldsymbol{y})}{\rho_t(\boldsymbol{x})}\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{y}\Big) \\
&= -\nabla_{\boldsymbol{x}} \cdot (\rho_t(\boldsymbol{x})\boldsymbol{v}_t(\boldsymbol{x}))
\end{aligned} \tag{54}$$

□

Remark. If we further have $\int q(\boldsymbol{z},\boldsymbol{z}')\mathrm{d}\boldsymbol{z}' = \gamma(\boldsymbol{z})$ for some $\gamma \in \Pi(\mu_0,\mu_1)$, then it is easy to check $\rho_0 = \mu_0, \rho_1 = \mu_1$. In this case, $\boldsymbol{v}_t$ transports μ_0 to μ_1. We point out that this condition is slightly weaker than $q = \gamma \otimes \gamma$ for some $\gamma \in \Pi(\mu_0,\mu_1)$, which we used in theorem 5.1.

A.7 Proof of Proposition 5.3

Proposition 5.3. *Let the true probability flow be ρ_t and the mean-field approximation be $\hat{\rho}_t$. Assume $\boldsymbol{b}_t, \boldsymbol{f}_t$ are both Lipschitz, and $M = \max_{0<t\leq 1} \max_{\boldsymbol{x}} \mathcal{W}_1(\rho_t, \pi_t(\cdot|\boldsymbol{x}))$, then $\forall 0 < t \leq 1, \exists C_t > 0$ such that $\mathcal{W}_1(\rho_t, \hat{\rho}_t) \leq C_t M$.*

Proof. Let L_b, L_f be the Lipschitz constant of $\boldsymbol{b}_t, \boldsymbol{f}_t$, and $L \geq L_b, L_f$. The true dynamics is

$$\partial_t \rho_t(\boldsymbol{x}) + \nabla_{\boldsymbol{x}} \cdot [\rho_t(\boldsymbol{x})\big(\boldsymbol{b}_t(\boldsymbol{x}) + \int_{\mathcal{X}} \boldsymbol{f}_t(\boldsymbol{x}, \boldsymbol{y})\pi_t(\boldsymbol{y}|\boldsymbol{x})\mathrm{d}\boldsymbol{y}\big)] = 0. \tag{55}$$

The mean-field approximation is

$$\partial_t \hat{\rho}_t(\boldsymbol{x}) + \nabla_{\boldsymbol{x}} \cdot [\hat{\rho}_t(\boldsymbol{x})\big(\boldsymbol{b}_t(\boldsymbol{x}) + \int_{\mathcal{X}} \boldsymbol{f}_t(\boldsymbol{x}, \boldsymbol{y})\hat{\rho}_t(\boldsymbol{y})\mathrm{d}\boldsymbol{y}\big)] = 0. \tag{56}$$

Let $R_t(\boldsymbol{x}) = \int_{\mathcal{X}} \boldsymbol{f}_t(\boldsymbol{x}, \boldsymbol{y})[\pi_t(\boldsymbol{y}|\boldsymbol{x}) - \hat{\rho}_t(\boldsymbol{y})]\mathrm{d}\boldsymbol{y}$ represent the difference between the two velocity fields. By the Lipschitz condition, we have

$$\begin{aligned}
\|R_t(\boldsymbol{x})\| &= \| \int_{\mathcal{X}} \boldsymbol{f}_t(\boldsymbol{x}, \boldsymbol{y})[\pi_t(\boldsymbol{y}|\boldsymbol{x}) - \hat{\rho}_t(\boldsymbol{y})]\mathrm{d}\boldsymbol{y}\| \\
&\leq \| \int_{\mathcal{X}} \boldsymbol{f}_t(\boldsymbol{x}, \boldsymbol{y})[\pi_t(\boldsymbol{y}|\boldsymbol{x}) - \rho_t(\boldsymbol{y})]\mathrm{d}\boldsymbol{y}\| + \| \int_{\mathcal{X}} \boldsymbol{f}_t(\boldsymbol{x}, \boldsymbol{y})[\rho_t(\boldsymbol{y}) - \hat{\rho}_t(\boldsymbol{y})]\mathrm{d}\boldsymbol{y}\| \\
&\leq L\Big(\mathcal{W}_1(\pi_t(\cdot|\boldsymbol{x}), \rho_t) + \mathcal{W}_1(\rho_t, \hat{\rho}_t)\Big) \\
&\leq LM + L\mathcal{W}_1(\rho_t, \hat{\rho}_t).
\end{aligned} \tag{57}$$

Since $\hat{\rho}_0 = \rho_0$, the difference between $\rho_t, \hat{\rho}_t$ can be fully described by the difference between the velocity fields. For any starting point $X_0 = Y_0 = \boldsymbol{x}$, the true dynamics is

$$\frac{\mathrm{d}}{\mathrm{d}t} X_t = \boldsymbol{b}_t(X_t) + \int_{\mathcal{X}} \boldsymbol{f}_t(X_t, \boldsymbol{y})\pi_t(\boldsymbol{y}|X_t)\mathrm{d}\boldsymbol{y} \triangleq A(X_t). \tag{58}$$

The mean-field dynamics is

$$\frac{\mathrm{d}}{\mathrm{d}t} Y_t = \boldsymbol{b}_t(Y_t) + \int_{\mathcal{X}} \boldsymbol{f}_t(Y_t, \boldsymbol{y})\hat{\rho}_t(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \triangleq B(Y_t). \tag{59}$$

Let $Z_t = \|X_t - Y_t\|$, we have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}(X_t - Y_t) &= A(X_t) - B(Y_t) = A(X_t) - B(X_t) + B(X_t) - B(Y_t) \\
&= R_t(X_t) + [\boldsymbol{b}_t(X_t) - \boldsymbol{b}_t(Y_t) + \int_{\mathcal{X}} \Big(\boldsymbol{f}_t(X_t, \boldsymbol{y}) - \boldsymbol{f}_t(Y_t, \boldsymbol{y})\Big)\hat{\rho}_t(\boldsymbol{y})\mathrm{d}\boldsymbol{y}]
\end{aligned} \tag{60}$$

Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t} Z_t \leq \|\frac{\mathrm{d}}{\mathrm{d}t}(X_t - Y_t)\| \leq LM + L\mathcal{W}_1(\rho_t, \hat{\rho}_t) + 2LZ_t. \tag{61}$$

By coupling X_t, Y_t generated by the same starting point $\boldsymbol{x}$ together, we obtain an inequality of $\mathcal{W}_1(\rho_t, \hat{\rho}_t)$.

$$\begin{aligned}
\mathcal{W}_1(\rho_t, \hat{\rho}_t) &= \inf_{\gamma} \int_{\mathcal{X}^2} \|\boldsymbol{x} - \boldsymbol{y}\|\gamma(\boldsymbol{x}, \boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y} \\
&\leq \int_{\mathcal{X}} Z_t(\boldsymbol{x})\rho_0(\boldsymbol{x})\mathrm{d}\boldsymbol{x}.
\end{aligned} \tag{62}$$

Therefore, it is sufficient to estimate the upper bound of Z_t. By Gronwall's inequality, we have

$$\begin{aligned}
Z_t &\leq e^{2Lt} L \int_0^t e^{-2Ls}(M + \mathcal{W}_1(\rho_s, \hat{\rho}_s))\mathrm{d}s \\
&= \frac{e^{2Lt} - 1}{2} M + e^{2Lt} L \int_0^t e^{-2Ls}\mathcal{W}_1(\rho_s, \hat{\rho}_s)\mathrm{d}s.
\end{aligned} \tag{63}$$

Therefore, we get another inequality of $\mathcal{W}_1(\rho_t, \hat{\rho}_t)$.

$$\begin{aligned}\mathcal{W}_1(\rho_t, \hat{\rho}_t) &\leq \int_{\mathcal{X}} \Big(\frac{e^{2Lt}-1}{2} M + e^{2Lt} L \int_0^t e^{-2Ls} \mathcal{W}_1(\rho_s, \hat{\rho}_s) \mathrm{d}s\Big) \rho_0(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &= \frac{e^{2Lt}-1}{2} M + e^{2Lt} L \int_0^t e^{-2Ls} \mathcal{W}_1(\rho_s, \hat{\rho}_s) \mathrm{d}s.\end{aligned} \tag{64}$$

Or equivalently, we have

$$e^{-2Lt} \mathcal{W}_1(\rho_t, \hat{\rho}_t) \leq \frac{1-e^{-2Lt}}{2} M + L \int_0^t e^{-2Ls} \mathcal{W}_1(\rho_s, \hat{\rho}_s) \mathrm{d}s. \tag{65}$$

Let $F(t) = \int_0^t e^{-2Ls} \mathcal{W}_1(\rho_s, \hat{\rho}_s) \mathrm{d}s$, the inequality can be reformulated as

$$F'(t) \leq \frac{1-e^{-2Lt}}{2} M + L F(t). \tag{66}$$

By Gronwall's inequality again, we have

$$F(t) \leq \frac{M}{6L}(2e^{Lt} - 3 + e^{-2Lt}). \tag{67}$$

Plug in (66), we have

$$\mathcal{W}_1(\rho_t, \hat{\rho}_t) \leq \frac{e^{3Lt}-1}{3} M. \tag{68}$$

□

A.8 Proof of Theorem 5.4

Theorem 5.4. *The minimizer is $\boldsymbol{u}_{\boldsymbol{\theta}}^{\boldsymbol{i}}(\boldsymbol{x}, \boldsymbol{y}, t) = \boldsymbol{u}_t^i(\boldsymbol{x}, \boldsymbol{y})$. When $q = \gamma \otimes \gamma$ where γ is the optimal QOT coupling of (9), and the travelling pair used is the minimizer of (7), then the learned velocity pair generates the probability flow of the dynamic QOT problem (10).*

Proof. Recall the conditional loss

$$\mathcal{L}_{\text{pair}}(\boldsymbol{\theta}) = \mathbb{E}_{t\sim\mathcal{U}[0,1],(\boldsymbol{z},\boldsymbol{z}')\sim q(\boldsymbol{z},\boldsymbol{z}'),(\boldsymbol{x},\boldsymbol{y})\sim\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')} \sum_{i=1}^{2} \left\| \boldsymbol{u}_{\boldsymbol{\theta}}^{\boldsymbol{i}}(\boldsymbol{x}, \boldsymbol{y}, t) - \boldsymbol{u}_t^i(\boldsymbol{x}, \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{z}') \right\|_2^2 \tag{69}$$

Similarly to what we have done in the proof of theorem 5.1, on the two-particle augmented space, the standard equivalence between the marginal flow matching loss and the conditional flow matching loss (Lipman et al., 2023; Tong et al., 2024a) tells us that minimizing the conditional loss (69) is equivalent to minimizing the marginal loss

$$\mathcal{L}_{\text{pair-marginal}}(\boldsymbol{\theta}) = \mathbb{E}_{t\sim\mathcal{U}[0,1],(\boldsymbol{x},\boldsymbol{y})\sim\pi_t(\boldsymbol{x},\boldsymbol{y})} \sum_{i=1}^{2} \left\| \boldsymbol{u}_{\boldsymbol{\theta}}^{\boldsymbol{i}}(\boldsymbol{x}, \boldsymbol{y}, t) - \boldsymbol{u}_t^i(\boldsymbol{x}, \boldsymbol{y}) \right\|_2^2 \tag{70}$$

Therefore, the minimizer is $\boldsymbol{u}_{\boldsymbol{\theta}}^{\boldsymbol{i}}(\boldsymbol{x}, \boldsymbol{y}, t) = \boldsymbol{u}_t^i(\boldsymbol{x}, \boldsymbol{y})$. By the pair marginalization theorem 5.1, we know that the minimizer generates the probability flow of the dynamic QOT. □

Remark. Although the learned velocity induces the same probability flow as the dynamic QOT construction, the mechanism by which it generates this flow is not exactly the same as in the definition of dynamic QOT. In dynamic QOT, the coupling first determines where each particle is transported, and the travelling-pair path then specifies how each pairwise transport is carried out. The resulting process is therefore generally non-Markovian, since its evolution depends on the prescribed endpoint. In realistic dynamics, however, future states are typically not known in advance. Instead, the evolution is determined by the current state, possibly together with its history. A natural approximation is therefore to replace the original process by a Markov process. The velocity pair learned by flow matching defines a pair of time-dependent velocity fields on the two-particle space, so particles evolve according to their current positions and time, making the resulting dynamics Markovian. The two processes share the same two-particle marginal distribution $\pi_t(\boldsymbol{x}, \boldsymbol{y})$ at every time t. In this sense, the learned velocity pair can be viewed as providing a Markovian projection of the original endpoint-conditioned process.

A.9 Proof of Theorem 5.5

Theorem 5.5. $\nabla_{\boldsymbol{\theta}}\mathcal{L}_M(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\mathcal{L}_C(\boldsymbol{\theta})$.

Proof. We first recall the two losses.

$$\begin{aligned}\mathcal{L}_M(\boldsymbol{\theta}) &= \mathbb{E}_{t\sim\mathcal{U}[0,1],\boldsymbol{x}\sim\rho_t(\boldsymbol{x})}\|\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t)-\boldsymbol{v}_t(\boldsymbol{x})\|^2\\ \mathcal{L}_C(\boldsymbol{\theta}) &= \mathbb{E}_{t\sim\mathcal{U}[0,1],(\boldsymbol{z},\boldsymbol{z}')\sim q(\boldsymbol{z},\boldsymbol{z}'),(\boldsymbol{x},\boldsymbol{y})\sim\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')}\|\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t)-\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')\|^2\end{aligned} \tag{71}$$

By direct calculation, we have

$$\begin{aligned}\mathcal{L}_C(\boldsymbol{\theta}) &= \mathbb{E}_{t\sim\mathcal{U}[0,1],(\boldsymbol{z},\boldsymbol{z}')\sim q(\boldsymbol{z},\boldsymbol{z}'),(\boldsymbol{x},\boldsymbol{y})\sim\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')}\|\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t)-\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')\|^2\\
&= \int_0^1\int\int_{\mathcal{X}^2}\|\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t)-\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')\|^2\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')q(\boldsymbol{z},\boldsymbol{z}')\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'\mathrm{d}t\\
&= \int_0^1\int\int_{\mathcal{X}^2}\|\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t)\|^2\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')q(\boldsymbol{z},\boldsymbol{z}')\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'\mathrm{d}t\\
&\quad+\int_0^1\int\int_{\mathcal{X}^2}\|\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')\|^2\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')q(\boldsymbol{z},\boldsymbol{z}')\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'\mathrm{d}t\\
&\quad-2\int_0^1\int\int_{\mathcal{X}^2}\langle\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t),\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')\rangle\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')q(\boldsymbol{z},\boldsymbol{z}')\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'\mathrm{d}t\\
&= \int_0^1\int_{\mathcal{X}^2}\|\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t)\|^2\pi_t(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}t+C\\
&\quad-2\int_0^1\int_{\mathcal{X}^2}\langle\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t),\int\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')\frac{\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')q(\boldsymbol{z},\boldsymbol{z}')}{\pi_t(\boldsymbol{x},\boldsymbol{y})}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'\rangle\pi_t(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}t\\
&= \int_0^1\int_{\mathcal{X}^2}\|\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t)\|^2\pi_t(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}t+C\\
&\quad-2\int_0^1\int_{\mathcal{X}^2}\langle\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t),\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y})\rangle\pi_t(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}t\\
&= \int_0^1\int_{\mathcal{X}^2}\|\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t)\|^2\pi_t(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}t+C\\
&\quad-2\int_0^1\int_{\mathcal{X}}\langle\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t),\int_{\mathcal{X}}\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y})\frac{\pi_t(\boldsymbol{x},\boldsymbol{y})}{\rho_t(\boldsymbol{x})}\mathrm{d}\boldsymbol{y}\rangle\rho_t(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t\\
&= \int_0^1\int_{\mathcal{X}^2}\|\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t)\|^2\pi_t(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}t+C-2\int_0^1\int_{\mathcal{X}}\langle\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t),\boldsymbol{v}_t(\boldsymbol{x})\rangle\rho_t(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t\\
&= \int_0^1\int_{\mathcal{X}}\|\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t)\|^2\rho_t(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t-2\int_0^1\int_{\mathcal{X}}\langle\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t),\boldsymbol{v}_t(\boldsymbol{x})\rangle\rho_t(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t\\
&\quad+\int_0^1\int_{\mathcal{X}}\|\boldsymbol{v}_t(\boldsymbol{x})\|^2\rho_t(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t+C\\
&= \int_0^1\int_{\mathcal{X}}\|\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x},t)-\boldsymbol{v}_t(\boldsymbol{x})\|^2\rho_t(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t+C\\
&= \mathcal{L}_M(\boldsymbol{\theta})+C\end{aligned} \tag{72}$$

where C is some constant independent of $\boldsymbol{\theta}$. Therefore, $\nabla_{\boldsymbol{\theta}}\mathcal{L}_M(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\mathcal{L}_C(\boldsymbol{\theta})$. □

A.10 Proof of Theorem 5.6

Theorem 5.6. *The minimizer of (23) is* $\boldsymbol{f}_t(\boldsymbol{x},\boldsymbol{y})$.

Proof. We first recall the loss.

$$\mathcal{L}_f(\boldsymbol{\phi}) = \mathbb{E}_{t\sim\mathcal{U}[0,1],(\boldsymbol{z},\boldsymbol{z}')\sim q(\boldsymbol{z},\boldsymbol{z}'),(\boldsymbol{x},\boldsymbol{y})\sim\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')}\|\boldsymbol{f}_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{y},t)-\boldsymbol{f}_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')\|_2^2 \tag{73}$$

Similarly to theorem 5.4, we have

$$\mathcal{L}_f(\boldsymbol{\phi}) = \mathbb{E}_{t\sim\mathcal{U}[0,1],(\boldsymbol{x},\boldsymbol{y})\sim\pi_t(\boldsymbol{x},\boldsymbol{y})} \left\|\boldsymbol{f}_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{y},t) - \boldsymbol{f}_t(\boldsymbol{x},\boldsymbol{y})\right\|_2^2 + C \tag{74}$$

where C is a constant independent of $\boldsymbol{\phi}$. Therefore, the minimizer is $\boldsymbol{f}_t(\boldsymbol{x},\boldsymbol{y})$. □

B Implementation details

B.1 Computation resources

All experiments were performed on a personal computer with NVIDIA 5070 Ti GPU.

B.2 Network architecture

The velocity net $\boldsymbol{v_\theta}$ was parameterized as a MLP with 128 hidden channels, 5 hidden layers and SiLU activations Elfwing et al. (2018). It receives the scalar time t concatenated with expression and spatial state $(\boldsymbol{x},\boldsymbol{s})$, and outputs the velocity on both expression and spatial space. Across all experiments, the networks were optimized by Adam (Kingma & Ba, 2017) with learning rate 2×10^{-3}, weight decay 10^{-5}, and gradient norm clipping at 5.

B.3 Lagrangian choice for experiments

As introduced in 4.2, we used a modality separated Lagrangian for numerical experiments. On both synthetic data and real data, we only added interaction term on spatial coordinates, and used standard kinetic energy on expression spaces. That is, let two cells represented by $(\boldsymbol{x},\boldsymbol{s}),(\boldsymbol{y},\boldsymbol{h})$ where $\boldsymbol{x},\boldsymbol{y}$ are expressions, $\boldsymbol{s},\boldsymbol{h}$ are spatial coordinates, we choose the Lagrangian as

$$\begin{aligned}\mathcal{L} =&\eta_{\text{gene}}\left(\frac{1}{2}\|\dot{\boldsymbol{x}}\|^2+\frac{1}{2}\|\dot{\boldsymbol{y}}\|^2\right)+\lambda_{\text{gene}}\left|\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{x}-\boldsymbol{y}\|\right|^2\\ &+\eta_{\text{spatial}}\left(\frac{1}{2}\|\dot{\boldsymbol{s}}\|^2+\frac{1}{2}\|\dot{\boldsymbol{h}}\|^2\right)+\lambda_{\text{spatial}}\left|\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{s}-\boldsymbol{h}\|\right|^2\end{aligned} \tag{75}$$

Since the corresponding QOT problem remains the same under global scaling, we set $\eta_{\text{gene}}=1$. The interaction term is only added to spatial Lagrangian, hence $\lambda_{\text{gene}}=0$. The Lagrangian reduces to

$$\mathcal{L} =\frac{1}{2}\|\dot{\boldsymbol{x}}\|^2+\frac{1}{2}\|\dot{\boldsymbol{y}}\|^2+\eta_{\text{spatial}}\left(\frac{1}{2}\|\dot{\boldsymbol{s}}\|^2+\frac{1}{2}\|\dot{\boldsymbol{h}}\|^2\right)+\lambda_{\text{spatial}}\left|\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{s}-\boldsymbol{h}\|\right|^2 \tag{76}$$

There are only two hyperparameters $\eta_{\text{spatial}},\lambda_{\text{spatial}}$. We conducted an ablation study in C.1.

B.4 Synthetic rotation data details

In 6, we evaluated TP-DATE on two synthetic rotation datasets. Here, we introduce the generation details. Each synthetic rotation dataset contains 2,000 paired cells from three cell types, with two spatial coordinates and 50 expression features. The target time point is a 90° rotation of the source time point. Therefore, the true midpoint is the corresponding 45° rotation.

The source spatial coordinates were sampled from three noisy curved arms. For a cell of type c, we sampled

$$r\sim\mathcal{U}(0.18,1),\qquad \epsilon_\theta\sim\mathcal{N}(0,0.055^2),\qquad \epsilon_r,\epsilon_q\sim\mathcal{N}(0,0.035^2),$$

and set

$$\theta=\theta_c+1.05r+\epsilon_\theta,\qquad \boldsymbol{s}_0=(0.35+2.25r+\epsilon_r)\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix}+\epsilon_q\begin{bmatrix}-\sin\theta\\ \cos\theta\end{bmatrix},$$

where $(\theta_0,\theta_1,\theta_2)=(15^\circ,140^\circ,260^\circ)$.

After concatenating the three cell types, we randomly permuted the cells, subtracted the global spatial centroid, and divided all coordinates by $\sqrt{n^{-1}\sum_i \|\boldsymbol{s}_{0,i}\|_2^2}$. The spatial trajectory was a rigid counterclockwise rotation,

$$\boldsymbol{s}_{t,i} = R(90^\circ t)\boldsymbol{s}_{0,i}, \qquad t \in [0,1],$$

where $R(\cdot)$ is the two-dimensional rotation matrix. Consequently, the observed source, held out midpoint, and target correspond to rotations of 0°, 45°, and 90°, respectively, and all pairwise spatial distances are exactly preserved.

Source expression was initialized independently sampled from $\mathcal{N}(0, 0.18^2)$. To generate cell-type markers, we added 2.2 to genes $1-12$, $13-24$, and $25-36$ for cell types 1, 2, and 3, respectively. Genes $37-50$ additionally received 0.75 times the following 14 dimensional spatial signals, evaluated using the normalized source coordinates (x_i, y_i), radius $\rho_i = \sqrt{x_i^2 + y_i^2}$, and angle $\phi_i = \text{atan2}\frac{y_i}{x_i}$:

$$(x_i, y_i, \rho_i, \sin\phi_i, \cos\phi_i, x_i y_i, x_i^2 - y_i^2, \sin 2\phi_i, \cos 2\phi_i,$$
$$\rho_i \sin(\phi_i + 0.7), \rho_i \cos(\phi_i - 0.4), \exp(-0.7\rho_i^2), \tanh(1.5x_i), \tanh(1.5y_i)).$$

Each expression feature was then centered and divided by its population standard deviation. For Rotation, expression remained unchanged over time. For Distractor, we constructed a drift matrix $\boldsymbol{\Delta}$ by adding 0.45 to genes $37-41$, $42-46$, and $47-50$ for cell types 1, 2, and 3, respectively, followed by independent noise sampled from $\mathcal{N}(0, 0.16^2)$ added for each cell, all 50 features. Denoted the expression matrix at time t as $\boldsymbol{e}_t$, we set

$$\boldsymbol{e}_{0.5} = \boldsymbol{e}_0 + \tfrac{1}{2}\boldsymbol{\Delta}, \qquad \boldsymbol{e}_1 = \boldsymbol{e}_0 + \boldsymbol{\Delta}.$$

Finally, each feature was standardized using its mean and population standard deviation computed jointly over the concatenated source, midpoint, and target expression matrices. Therefore, the Distractor dataset contains a cell type dependent temporal expression shift, while its spatial dynamics remain the same rigid rotation as in Rotation.

For the TP-DATE experiments in the main text, the hyperparameter was set to $(\eta_{\text{spatial}}, \lambda_{\text{spatial}}) = (1, 16)$.

B.5 Real data details

B.5.1 Developing Mouse brain

The developing mouse brain data was adopted from (Chen et al., 2022). This dataset contains spatial transcriptomics sections from mouse embryonic development at eight different time points. We selected brain cells according to the cell annotations provided with the dataset and performed Leiden clustering (Traag et al., 2019) on gene expression space for visualization purpose only. We use the snapshots from day 14.5, 15.5, and 16.5, denoted E14.5, E15.5, and E16.5 with 17591, 17031, 17296 cells, respectively. We reduced the dimension by first selecting 1500 highly variable genes, then PCA to 50D. The spatial coordinates were linearly scaled to $[-1, 1]$. The hold one out experiment was trained on E14.5, E16.5, and evaluated on E15.5. The hyperparameters are set to $(\eta_{\text{spatial}}, \lambda_{\text{spatial}}) = (100, 0.01)$. An overview of the data is shown in figure 2. Different colors stand for different Leiden clusters.

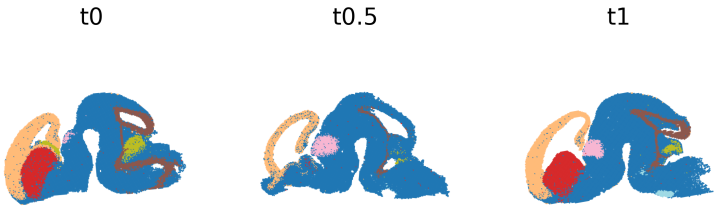


Figure 2: Mouse Brain overview

B.5.2 ARTISTA

The ARTISTA data was adopted from (Wei et al., 2022). This dataset describes the brain regeneration process in axolotl following injury and contains spatial transcriptomics sections collected at seven different time points from Day 2 to Day 60 after injury. In the main text, we use snapshots at Day 2, Day 5, and Day 10, and subsample 7,500 cells per snapshot. The expression vectors were also dimension reduced to 50D by PCA. The spatial coordinates were linearly scaled to $[-1, 1]$. For hold one out experiment, we trained on Day 2 and Day 10, and evaluated on Day 5. The hyperparameters are set to $(\eta_{\text{spatial}}, \lambda_{\text{spatial}}) = (100, 0.01)$. The normalized evaluation time is therefore 0.375 instead of 0.5. An overview of the data is shown in figure 3. Different colors stand for Niche annotations provided with the dataset.

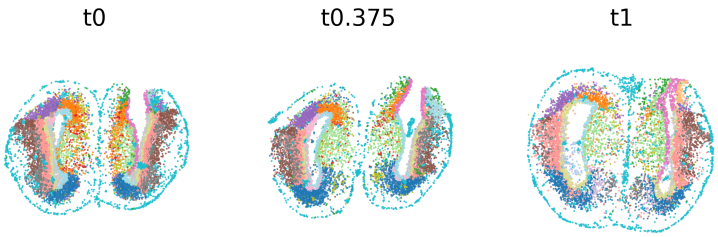


Figure 3: ARTISTA overview

B.5.3 TUMOR

The Tumor data was adopted from (Zhang et al., 2026a). This dataset contains five spatial transcriptomics sections of the same MC38 tumor collected at different depths (subQ-1 to subQ-5). We use subQ-3, subQ-4, and subQ-5 with 10,000 spatially sampled cells per snapshot. The expression vectors are also reduced to 50D via PCA. The spatial coordinates were linearly scaled to $[-1, 1]$. For hold one out experiments, models are trained on subQ-3 and subQ-5, and evaluated on subQ-4. The hyperparameters are set to $(\eta_{\text{spatial}}, \lambda_{\text{spatial}}) = (0.01, 1)$. An overview of the data is shown in figure 4. Different colors stand for Niche annotations provided with the dataset.

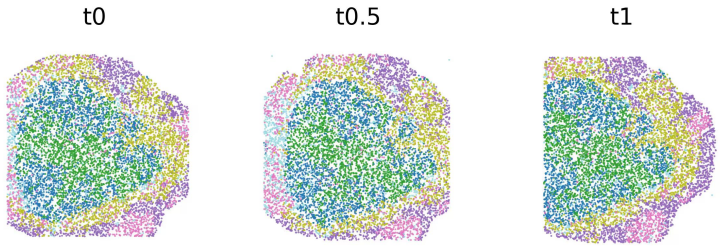


Figure 4: Tumor overview

B.6 EVALUATION METRICS

In the hold one out experiments on real data, we train the model on the two endpoint time points, infer the distribution at the intermediate time point, and compare the prediction with the ground truth. In the spatial transcriptomics setting, each cell is jointly characterized by its gene expression $\boldsymbol{x}$ and spatial coordinate $\boldsymbol{s}$, so that each snapshot can be regarded as a joint distribution $p(\boldsymbol{x}, \boldsymbol{s})$. At the intermediate time point, we evaluate the discrepancy between the ground truth distribution p and the predicted distribution $\hat{p}$.

B.6.1 MARGINAL WASSERSTEINS

To measure the discrepancy between two joint distribution, one can first measure the discrepancy between the marginal distributions. Let $p_x(\boldsymbol{x}) = \int p(\boldsymbol{x}, \boldsymbol{s}) \mathrm{d}\boldsymbol{s}, p_s(\boldsymbol{s}) = \int p(\boldsymbol{x}, \boldsymbol{s}) \mathrm{d}\boldsymbol{x}$ be the marginals

on expression space and spatial space, the spatial Wasserstein distance and expression Wasserstein distance are defined as

$$\begin{aligned}\text{Spatial } W_2^2(p,\hat{p}) &= \inf_{\gamma\in\Pi(p_s,\hat{p}_s)} \int \|\boldsymbol{s}-\boldsymbol{s}'\|^2\gamma(\boldsymbol{s},\boldsymbol{s}')\mathrm{d}\boldsymbol{s}\mathrm{d}\boldsymbol{s}' \\ \text{Expression } W_2^2(p,\hat{p}) &= \inf_{\gamma\in\Pi(p_x,\hat{p}_x)} \int \|\boldsymbol{x}-\boldsymbol{x}'\|^2\gamma(\boldsymbol{x},\boldsymbol{x}')\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{x}'\end{aligned} \tag{77}$$

Intuitively, these marginal discrepancies measure how well can the model reconstruct the spatial or expression information solely. To further measure how well can the model reconstruct the spatial and expression information jointly, we introduce fused Wasserstein distance and spatially coupled expression MSE (SC-eMSE).

B.6.2 Fused Wasserstein

Fused W_2^2 is defined as

$$\text{Fused } W_2^2(p,\hat{p}) = \inf_{\gamma\in\Pi(p,\hat{p})} \int \|\boldsymbol{z}-\boldsymbol{z}'\|^2\gamma(\boldsymbol{z},\boldsymbol{z}')\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}' \tag{78}$$

where $\boldsymbol{z}=(\boldsymbol{x},\alpha\boldsymbol{s})$. The expression vector $\boldsymbol{x}$ was concatenated with a scaled spatial vector $\alpha\boldsymbol{s}$, and the standard Wasserstein distance is calculated based on the concatenated vector. α was introduced to balance the importance of $\boldsymbol{x}$ and $\boldsymbol{s}$, since they may have different dimension and scale. For all datasets, we choose α with a common rule

$$\alpha = \frac{\sum_{d=1}^{D_x}\mathrm{Std}(\boldsymbol{x}^d)}{\sum_{d=1}^{D_s}\mathrm{Std}(\boldsymbol{s}^d)}. \tag{79}$$

where D_x, D_s are the dimensions of expression space and spatial space, $\boldsymbol{x}^d, \boldsymbol{s}^d$ are the d-th components of $\boldsymbol{x}, \boldsymbol{s}$. Intuitively, this selection of α balances the total standard deviation of expression and space. In hold one out experiments, the standard deviations are calculated only on the training time points.

B.6.3 SC-eMSE

To test whether expression is reconstructed at the correct spatial location, we additionally report spatially coupled expression MSE (SC-eMSE). Let γ be the optimal coupling between $\rho, \hat{\rho}$, obtained using spatial coordinates $\boldsymbol{s}$ alone with the spatial cost $\|\boldsymbol{s}-\boldsymbol{s}'\|^2$, we define

$$\text{SC-eMSE}(\rho,\hat{\rho}) = \frac{1}{D_x}\int \|\boldsymbol{x}-\boldsymbol{x}'\|^2\gamma(\boldsymbol{z},\boldsymbol{z}')\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'$$

where $\boldsymbol{z}=(\boldsymbol{x},\boldsymbol{s})$. The spatial coupling γ is fitted without expression information, while the transport cost $\|\boldsymbol{x}-\boldsymbol{x}'\|^2$ is calculated without spatial information. Intuitively, SC-eMSE measures whether the predicted snapshot shares similar spatial expression patterns with the ground truth.

B.6.4 Toydata metrics

On rotation toys, the ground truth dynamics are known rigid rotations. Therefore, we can directly calculate the spatial MSE to measure the spatial level accuracy, instead of calculating distribution-level Wasserstein distance. At the unseen time point t, let the ground truth spatial states be $\boldsymbol{s}^\star$, the predicted spatial states be $\hat{\boldsymbol{s}}$. The spatial MSE is defined as

$$\text{MSE}_s = \frac{1}{n}\sum_{i=1}^{n}\|\hat{\boldsymbol{s}}_i-\boldsymbol{s}_i^\star\|^2. \tag{80}$$

We also calculated the pair distortion to measure how well can the predicted dynamics preserve the spatial structure. We uniformly sampled $M=6000$ pairs $\mathcal{P}=\{(i_r,k_r)\}_{r=1}^M$ where $i_t\neq k_r$. The pair distortion is defined as

$$\text{PD} = \frac{1}{M}\sum_{r=1}^{M}\frac{\left|\|\hat{s}_{i_r}-\hat{s}_{k_r}\|_2-\|s_{i_r}^\star-s_{k_r}^\star\|_2\right|}{\|s_{i_r}^\star-s_{k_r}^\star\|_2}. \tag{81}$$

Intuitively, it measures the average relative error of pairwise spatial distance. A lower PD means the predicted dynamics preserves the spatial structure better. In practice, $\|s^\star_{i_r} - s^\star_{k_r}\|_2$ can be easily calculated at time 0

$$\|s^\star_{i_r} - s^\star_{k_r}\|_2 = \|s^0_{i_r} - s^0_{k_r}\|_2$$

since the true dynamics are rigid transformations. Note that, the ground truth pair distortions are zero on toy datasets, so lower values are meaningful here. But it is not the case of real datasets where nonzero biological deformation may be correct, and the held out cells have no source-cell correspondence. Therefore, we only report pair distortions on toy datasets.

B.7 GW-CFM AND FGW-CFM BASELINES

In the main text, we compared TP-DATE with GW-CFM and FGW-CFM. These baselines are introduced as a naive dynamic extension of static GW-OT and FGW-OT. We used GW-OT coupling or FGW-OT coupling, and the displacement conditional path $\boldsymbol{x}_t = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1$ for conditional flow matching training. That is, for two probability densities μ_0, μ_1, we first calculate the GW-OT coupling or FGW-OT coupling $\gamma(\boldsymbol{x}_0, \boldsymbol{x}_1)$. Next, we parameterize a velocity network $\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x}, t)$, and train it by minimizing the conditional flow matching loss

$$\mathcal{L}_{\text{CFM}}(\boldsymbol{\theta}) = \mathbb{E}_{t\sim\mathcal{U}[0,1],(\boldsymbol{x}_0,\boldsymbol{x}_1)\sim\gamma(\boldsymbol{x}_0,\boldsymbol{x}_1),\boldsymbol{x}=(1-t)\boldsymbol{x}_0+t\boldsymbol{x}_1} \|\boldsymbol{v}_{\boldsymbol{\theta}}(\boldsymbol{x}, t) - (\boldsymbol{x}_1 - \boldsymbol{x}_0)\|_2^2 . \tag{82}$$

It can be viewed as replacing the OT coupling in OT-CFM by GW/FGW-OT coupling. Intuitively, it is a kind of linear interpolation between μ_0, μ_1 induced by the corresponding coupling, therefore a natural choice for dynamic extension. For FGW-OT, we fused OT and GW-OT with weight $7:3$.

B.8 STATIC QOT SOLVER

The static QOT problem is

$$\text{QOT}_S(\mu_0, \mu_1) = \inf_{\gamma\in\Pi(\mu_0,\mu_1)} \int_{\mathcal{X}^4} \mathcal{A}(\boldsymbol{z}, \boldsymbol{z}')\gamma(\boldsymbol{z})\gamma(\boldsymbol{z}')\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'. \tag{83}$$

In practice, μ_0, μ_1 are represented by point cloud data $\{X_i\}_{i=1:M}, \{Y_j\}_{j=1:N}$. The conditional variables $\boldsymbol{z} = (X_i, Y_j), \boldsymbol{z}' = (X_k, Y_l)$ are two end point pairs. The static cost in (83) can be written as

$$Q(\Gamma) = \sum A_{ij,kl}\Gamma_{ij}\Gamma_{kl} \tag{84}$$

where $\Gamma \in \{\Gamma \in \mathbb{R}^{M\times N}_{\geq 0} | \Gamma\mathbf{1} = \frac{1}{M}\mathbf{1}, \mathbf{1}^{\mathrm{T}}\Gamma = \frac{1}{N}\mathbf{1}^{\mathrm{T}}\}$ is the empirical coupling, $A_{ij,kl} = A(X_i, Y_j, X_k, Y_l)$ is the path action. The static QOT problem is then formulated as

$$\min_{\Gamma} \sum A_{ij,kl}\Gamma_{ij}\Gamma_{kl} \tag{85}$$

which is a quadratic programming w.r.t Γ.

B.8.1 FRANK-WOLFE ALGORITHM

The Frank-Wolfe algorithm (Kerdoncuff et al., 2021) is an iterative first-order optimization algorithm for convex constraint optimization. Though QOT problem may not be convex, it is standard and widely used in solving Gromov-Wasserstein type static QOT problem (Flamary et al., 2021). Consider a convex set D and an optimization problem

$$\min_{\boldsymbol{x}\in D} f(\boldsymbol{x}) \tag{86}$$

The Frank-Wolfe algorithm solves it by iteratively doing

$$\begin{aligned} &\text{Initialization} \quad \boldsymbol{x}_0 \in D, \quad k = 0 \\ &\text{Step1.} \quad \boldsymbol{s}_k = \arg\min_{\boldsymbol{s}\in D} \boldsymbol{s}^{\mathrm{T}}\nabla f(\boldsymbol{x}_k) \\ &\text{Step2.} \quad \boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \frac{2}{k+2}(\boldsymbol{s}_k - \boldsymbol{x}_k), \quad k = k+1 \end{aligned} \tag{87}$$

In our case, with symmetric conditions such as $\mathcal{L}$ is symmetric to $(\boldsymbol{x}, \dot{\boldsymbol{x}}), (\boldsymbol{y}, \dot{\boldsymbol{y}})$, it is easy to check $A_{ij,kl} = A_{kl,ij}$. Therefore, $\nabla Q(\Gamma) = 2\sum A_{ij,kl}\Gamma_{kl}$. The Frank-Wolfe algorithm (87) can be realized as

$$\begin{aligned} &\text{Initialization} \quad \Gamma^0, \quad n = 0 \\ &\text{Step1.} \quad G^n_{ij} = 2\sum A_{ij,kl}\Gamma^n_{kl} \\ &\text{Step2.} \quad \Pi^n = \arg\min_{\Pi} \langle \Pi, G^n \rangle \\ &\text{Step3.} \quad \Gamma^{n+1} = \Gamma^n + \frac{2}{n+2}(\Pi^n - \Gamma^n), \quad n = n+1 \end{aligned} \tag{88}$$

In step 1, we estimate the gradient G. In step 2, we solve a standard optimal transport subproblem. In step 3, we update Γ. The OT subproblem is relaxed and solved by Sinkhorn.

B.8.2 Gradient estimation

The full computation of step 1 is $\mathcal{O}(M^2N^2)$ which is expensive. To reduce the computational cost, we estimate the gradient via Monte Carlo.

$$\hat{G}_{ij} = \frac{2}{R}\sum_{h=1}^{R} A_{ij,k_h l_h}, \quad (k_h, l_h) \overset{i.i.d.}{\sim} \Gamma \tag{89}$$

The computational complexity is $\mathcal{O}(RMN)$ where R is the number of Monte Carlo samples. In the main text, we use $R = 64$.

B.8.3 Sparse KNN

If $\mathcal{O}(RMN)$ is still too expensive, we can further construct a bidirectional KNN graph between $\{X_i\}_{i=1:M}, \{Y_j\}_{j=1:N}$ and restrict Γ on the graph. Intuitively, X_i is more likely to transport to some Y_j near to itself rather than far away from itself. Therefore, we construct two KNN graphs on gene expression space. Each X_i is linked to its K nearest Y_j, and each Y_j is also linked to its K nearest X_i. We take the union of the edges, denoted E_0. That is, $(i,j) \in E_0$ means X_i is one of the K nearest neighbors of Y_j, or Y_j is one of the K nearest neighbors of X_i. It is easy to check $|E_0| \le K(M+N)$. We use bidirectional KNN rather than one way KNN so that all points have neighbors. We then want to restrict the coupling Γ on E_0, that is

$$\Gamma \in \mathbb{R}^{M\times N}_{\ge 0}, \quad \Gamma\mathbf{1} = \frac{1}{M}\mathbf{1}, \qquad \mathbf{1}^{\mathrm{T}}\Gamma = \frac{1}{N}\mathbf{1}^{\mathrm{T}}, \quad \Gamma_{ij} = 0, (i,j) \notin E_0. \tag{90}$$

Although the bidirectional construction guarantees that every row and column has at least one admissible edge, this alone does not necessarily guarantee the existence of a coupling with the prescribed marginals. We therefore augment E_0 by a small number of additional edges to guarantee feasibility. Specifically, initialize residual masses

$$r_i = \frac{1}{M}, \qquad s_j = \frac{1}{N},$$

and an auxiliary coupling $\bar{\Gamma} = 0$. We first scan all edges $(i,j) \in E_0$ in dictionary order. For each edge, we assign

$$\delta_{ij} = \min\{r_i, s_j\}, \qquad \bar{\Gamma}_{ij} \leftarrow \bar{\Gamma}_{ij} + \delta_{ij},$$

and update $r_i \leftarrow r_i - \delta_{ij}$, $s_j \leftarrow s_j - \delta_{ij}$. After all KNN edges have been scanned, if residual masses remain, we repeatedly select a pair i, j with $r_i > 0$ and $s_j > 0$, add the edge (i,j), and assign

$$\delta = \min\{r_i, s_j\}.$$

Each added edge exhausts at least one residual row or column. Therefore, at most $M + N - 1$ additional edges are required. Denoting these repair edges by E_{rep}, we finally set

$$E = E_0 \cup E_{\mathrm{rep}}.$$

By construction, $\bar{\Gamma}$ satisfies

$$\bar{\Gamma}\mathbf{1} = \frac{1}{M}\mathbf{1}, \qquad \mathbf{1}^{\mathrm{T}}\bar{\Gamma} = \frac{1}{N}\mathbf{1}^{\mathrm{T}}, \qquad \mathrm{supp}(\bar{\Gamma}) \subseteq E,$$

which explicitly guarantees that the sparse coupling constraint is feasible. Moreover,

$$|E| \le K(M+N) + M + N - 1,$$

and the feasibility repair requires only $\mathcal{O}(K(M+N))$ additional computation. We then restrict the coupling Γ on the augmented sparse support E, that is

$$\Gamma \in \mathbb{R}_{\ge 0}^{M\times N}, \quad \Gamma\mathbf{1} = \frac{1}{M}\mathbf{1}, \quad \mathbf{1}^{\mathrm{T}}\Gamma = \frac{1}{N}\mathbf{1}^{\mathrm{T}}, \quad \Gamma_{ij} = 0,\ (i,j) \notin E. \tag{91}$$

Mathematically, solving the OT subproblem under constraint (91) is equivalent to solving a standard OT problem with a masked cost

$$\begin{cases} \tilde{G}_{ij} = G_{ij}, & (i,j) \in E \\ \tilde{G}_{ij} = +\infty, & (i,j) \notin E \end{cases} \tag{92}$$

In Sinkhorn iteration, infinite cost leads to zero element in Gibbs kernel

$$\begin{cases} K_{ij} = \exp(-\tilde{G}_{ij}/\tau) = \exp(-G_{ij}/\tau), & (i,j) \in E \\ K_{ij} = \exp(-\tilde{G}_{ij}/\tau) = 0, & (i,j) \notin E \end{cases} \tag{93}$$

where τ is the coefficient of entropy relaxation. Further, a zero element in Gibbs kernel leads to zero element in updated coupling

$$\Gamma_{ij}^{\text{new}} = u_i K_{ij} v_j = 0, \quad (i,j) \notin E \tag{94}$$

where u, v are dual variables. The analysis above shows that we can only estimate G_{ij} and update Γ_{ij} for $(i,j) \in E$. For $(i,j) \notin E$, K_{ij}, Γ_{ij} are automatically zero. The complexity of gradient estimation is therefore further reduced to $\mathcal{O}(RK(M+N))$. In the main text, we use $\tau = 10^{-4}, K = 64$. We conducted an ablation study for K in C.1.

C Additional results

C.1 Ablation studies

C.1.1 Loss weights

To demonstrate the robustness of TP-DATE, we conducted an ablation study for hyperparameters $(\eta_{\text{spatial}}, \lambda_{\text{spatial}})$ B.3 on ARTISTA. As shown in Table 4, TP-DATE remains robust under different hyperparameter settings. We did not run different random seeds for this ablation experiment. All results are reported in one run.

C.1.2 Sparse KNN

We also conducted an ablation study for K in sparse KNN construction on the Tumor dataset. We tested $K = 64, 128, 256, 512, 1024$, on the Tumor dataset. The corresponding model performance are approximately the same, demonstrating the robustness of sparse KNN technique (Table 5).

We also recorded the training time and memory usage under different values of K. As shown in Figs.5, on a log–log scale, the slopes of the linear regressions for both training time and GPU memory usage with respect to K are close to 1, indicating that, as expected, the computational cost of TP-DATE scales approximately linearly with K. CPU memory usage does not exhibit the same linear dependence, mainly because CPU memory is substantially affected by data loading, preprocessing, and other fixed overheads, which dominate the algorithmic memory cost when K is relatively small.

C.1.3 Conditional path

In the Mouse Brain and ARTISTA experiments, we used relatively small weights for the interaction term. To isolate the contribution of the interaction-induced conditional path from that of the static QOT coupling, we trained an additional baseline using exactly the same QOT coupling as TP-DATE but replacing the travelling-pair conditional path with standard displacement interpolation. All other training settings and random seeds were kept unchanged. As shown in Table 6, TP-DATE achieves

Table 4: TP-DATE hyperparameter sensitivity on ARTISTA.

η_{spatial}	λ_{spatial}	Spatial W_2 (↓)	Expression W_2 (↓)	SC-eMSE (↓)	Balanced fused W_2 (↓)
10^{-2}	10^{-2}	0.10021	5.39525	0.89497	8.34454
10^{-2}	10^{-1}	0.10339	5.40739	0.90109	8.48391
10^{-2}	1	0.09922	5.41446	0.90192	8.35143
10^{-2}	10	0.10366	5.42960	0.90873	8.49660
10^{-2}	10^{2}	0.11070	5.35272	0.89337	8.75079
10^{-1}	10^{-2}	0.10039	5.38527	0.89438	8.34612
10^{-1}	10^{-1}	0.10099	5.36024	0.88781	8.35834
10^{-1}	1	0.10213	5.35141	0.88634	8.38445
10^{-1}	10	0.10323	5.40274	0.90401	8.48153
10^{-1}	10^{2}	0.10297	5.40772	0.91005	8.49803
1	10^{-2}	0.09734	5.40196	0.89723	8.24340
1	10^{-1}	0.09875	5.42559	0.90137	8.29199
1	1	0.09985	5.50784	0.92271	8.41832
1	10	0.09786	5.46790	0.91053	8.30708
1	10^{2}	0.10537	5.24299	0.86934	8.50025
10	10^{-2}	0.09331	5.36099	0.87991	8.06390
10	10^{-1}	0.09105	5.34367	0.87604	7.96595
10	1	0.09381	5.37037	0.87955	8.09346
10	10	0.09014	5.36759	0.87766	7.94178
10	10^{2}	0.09154	5.43260	0.89303	8.03001
10^{2}	10^{-2}	0.08922	5.23544	0.83938	7.79900
10^{2}	10^{-1}	0.08859	5.33771	0.86599	7.86204
10^{2}	1	0.09080	5.38808	0.87992	8.00044
10^{2}	10	0.08824	5.28488	0.85379	7.79917
10^{2}	10^{2}	0.08751	5.31516	0.86013	7.77876

Table 5: K ablation study on Tumor dataset. The dagger marks the value used in the main experiments.

k	Spatial W_2 (↓)	Expression W_2 (↓)	SC-eMSE (↓)	Balanced fused W_2 (↓)
$64^{\dagger}$	0.069590 ± 0.001515	4.819554 ± 0.138743	1.045954 ± 0.028812	7.374801 ± 0.087657
128	0.070472 ± 0.002885	4.802107 ± 0.196712	1.037978 ± 0.044555	7.399844 ± 0.127601
256	0.069701 ± 0.001579	4.766193 ± 0.215132	1.029943 ± 0.052658	7.348703 ± 0.145340
512	0.071247 ± 0.002133	4.816275 ± 0.175593	1.037625 ± 0.039205	7.395463 ± 0.142730
1024	0.070168 ± 0.003771	4.830717 ± 0.111594	1.042600 ± 0.017763	7.395537 ± 0.042100

lower mean errors across all evaluation metrics on both datasets, with particularly clear improvements in expression reconstruction, SC-eMSE, and fused W_2. These results indicate that the performance gain cannot be attributed solely to the static QOT coupling. Even with a relatively small interaction weight, the interaction induced dynamic conditional path provides an additional and consistent benefit. We do not additionally consider the zero-interaction coupling as a separate ablation, since removing the interaction term reduces the induced static QOT to the corresponding weighted standard OT formulation, whose flow-matching counterpart is already represented by OT-CFM.

Table 6: Conditional path ablation on spatial transcriptomics hold one out experiment. We report mean and standard deviation over 5 random seeds.

Dataset	Method	Spatial W_2 (↓)	Expression W_2 (↓)	SC-eMSE (↓)	Balanced fused W_2 (↓)
Mouse brain	TP-DATE (ours)	$\mathbf{0.08678 \pm 0.00291}$	$\mathbf{6.40989 \pm 0.04644}$	$\mathbf{1.47693 \pm 0.02312}$	$\mathbf{9.59301 \pm 0.09255}$
	QOT + straight path	0.09615 ± 0.00500	6.45808 ± 0.05616	1.52589 ± 0.03583	10.05324 ± 0.22572
ARTISTA	TP-DATE (ours)	$\mathbf{0.09196 \pm 0.00500}$	$\mathbf{5.39310 \pm 0.28347}$	$\mathbf{0.87824 \pm 0.07587}$	$\mathbf{8.01158 \pm 0.27289}$
	QOT + straight path	0.09210 ± 0.00345	5.94608 ± 0.39138	1.02462 ± 0.10252	8.42871 ± 0.30357

C.2 SCALABILITY

We trained TP-DATE and simulation-based baselines (stVCR (Peng et al., 2026a), CytoBridge (Zhang et al., 2025b)) on different cell numbers. We downsampled the subQ-3, subQ-5 slices of the Tumor dataset (Zhang et al., 2026a) to 10,000, 30,000, 50,000, 70,000, and 90,000 cells, respectively, and recorded the training time and memory usage of the three methods on datasets of the corresponding

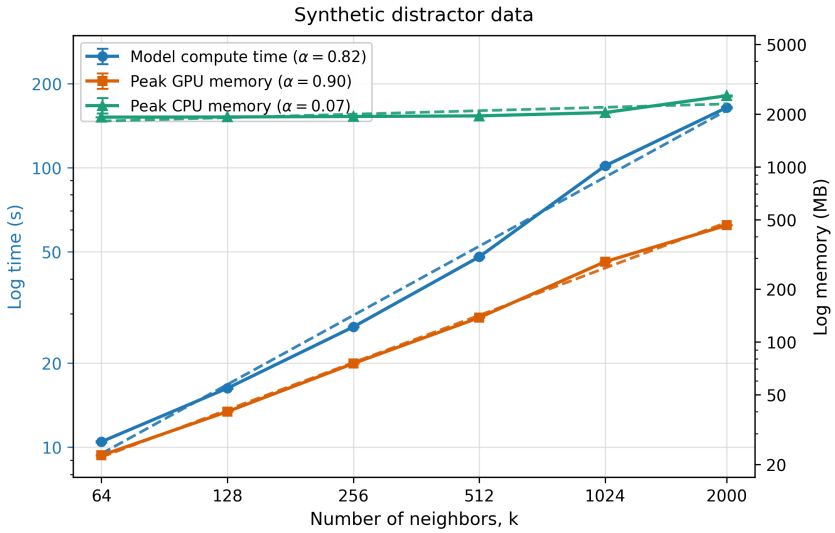


Figure 5: Training time and memory usage for different K on Tumor.

sizes. As shown in Figure 6, CytoBridge ran out of memory on 90000, stVCR and TP-DATE both require moderate memory. As shown in Figure 7, TP-DATE exhibits computational cost that scales linearly with dataset size and is more efficient than the two simulation-based baselines. This highlights the computational efficiency of simulation-free approaches such as flow matching.

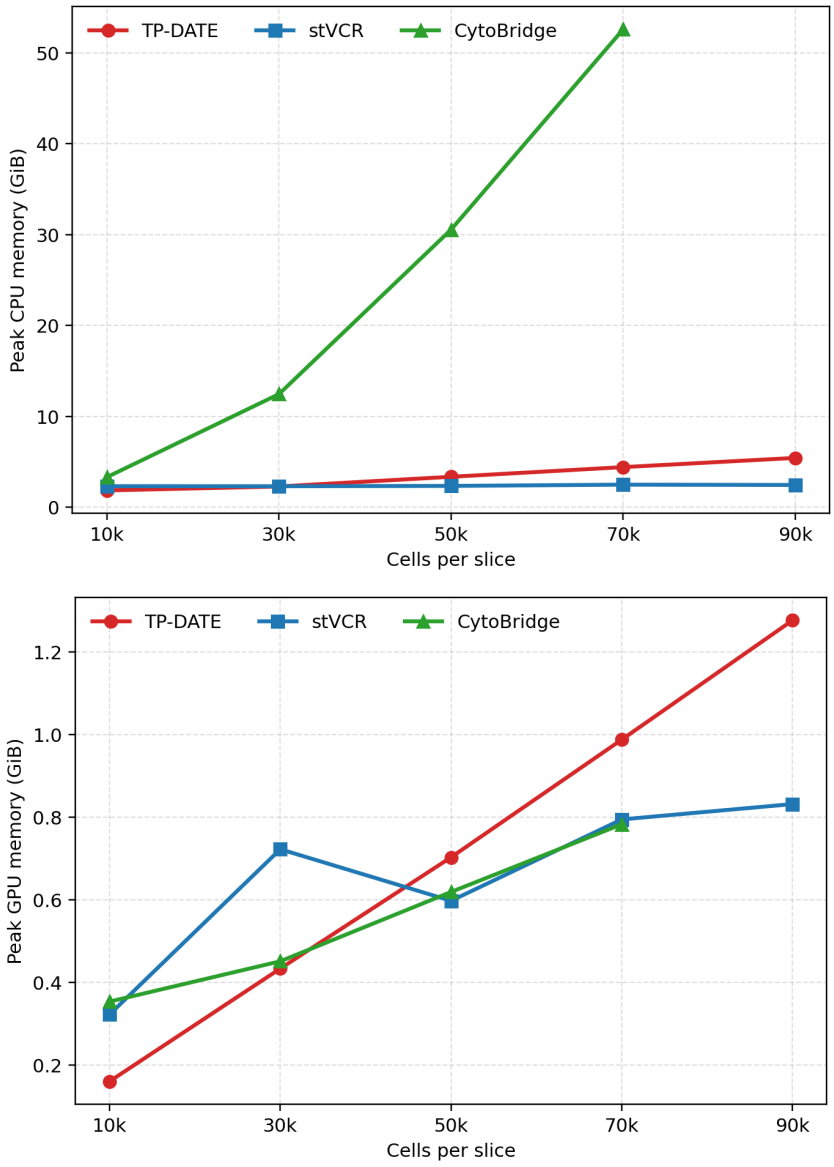


Figure 6: Training memory on different size data

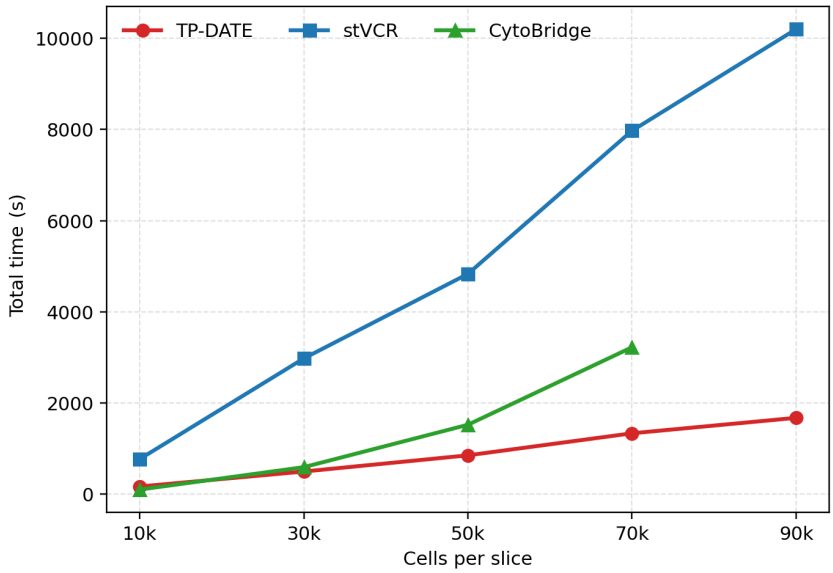


Figure 7: Training time on different size data

D RELATIONS TO OTHER WORKS

D.1 GW-OT AND FGW-OT

Consider the Lagrangian

$$\mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, \dot{\boldsymbol{x}}, \dot{\boldsymbol{y}}) = \frac{1}{2}\|\dot{\boldsymbol{x}}\|^2 + \frac{1}{2}\|\dot{\boldsymbol{y}}\|^2 + \lambda|\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{x}_t - \boldsymbol{y}_t\||^2. \tag{95}$$

D.1.1 STANDARD STATIC OT

When $\lambda = 0$, the static cost is

$$\mathcal{A}(\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{y}_0, \boldsymbol{y}_1) = \inf_{\boldsymbol{x}_t, \boldsymbol{y}_t} \int_0^1 \frac{1}{2}\|\dot{\boldsymbol{x}}_t\|^2 + \frac{1}{2}\|\dot{\boldsymbol{y}}_t\|^2 \mathrm{d}t = \frac{1}{2}\Big(\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|^2 + \|\boldsymbol{y}_1 - \boldsymbol{y}_0\|^2\Big). \tag{96}$$

The corresponding static form is

$$\begin{aligned}
\mathrm{QOT}_S(\mu_0, \mu_1) &= \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\mathcal{X}^4} \mathcal{A}(\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{y}_0, \boldsymbol{y}_1)\gamma(\boldsymbol{x}_0, \boldsymbol{x}_1)\gamma(\boldsymbol{y}_0, \boldsymbol{y}_1)\mathrm{d}\boldsymbol{x}_0\mathrm{d}\boldsymbol{x}_1\mathrm{d}\boldsymbol{y}_0\mathrm{d}\boldsymbol{y}_1 \\
&= \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \frac{1}{2}\Big(\int_{\mathcal{X}^2} \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|^2\gamma(\boldsymbol{x}_0, \boldsymbol{x}_1)\mathrm{d}\boldsymbol{x}_0\mathrm{d}\boldsymbol{x}_1 + \int_{\mathcal{X}^2} \|\boldsymbol{y}_1 - \boldsymbol{y}_0\|^2\gamma(\boldsymbol{y}_0, \boldsymbol{y}_1)\mathrm{d}\boldsymbol{y}_0\mathrm{d}\boldsymbol{y}_1\Big) \\
&= \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\mathcal{X}^2} \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|^2\gamma(\boldsymbol{x}_0, \boldsymbol{x}_1)\mathrm{d}\boldsymbol{x}_0\mathrm{d}\boldsymbol{x}_1 \\
&= \mathrm{OT}(\mu_0, \mu_1).
\end{aligned} \tag{97}$$

It is equivalent to the standard static OT.

D.1.2 STANDARD GW-OT

When there are no kinetic terms, which is $\mathcal{L} = |\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{x}_t - \boldsymbol{y}_t\||^2$, let $r_t = \|\boldsymbol{x}_t - \boldsymbol{y}_t\|$, the static cost is

$$\mathcal{A}(\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{y}_0, \boldsymbol{y}_1) = \inf_{r_t} \int_0^1 |\dot{r}_t|^2 \mathrm{d}t \quad \text{s.t.} \quad r_0 = \|\boldsymbol{x}_0 - \boldsymbol{y}_0\|, \quad r_1 = \|\boldsymbol{x}_1 - \boldsymbol{y}_1\|. \tag{98}$$

Using the Euler-Lagrange equation, it is easy to show that when $d \geq 2$, the optimal r_t is attained by $r_t = (1-t)r_0 + tr_1$. Therefore, the path action is

$$\mathcal{A}(\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{y}_0, \boldsymbol{y}_1) = (r_1 - r_0)^2 = (\|\boldsymbol{x}_1 - \boldsymbol{y}_1\| - \|\boldsymbol{x}_0 - \boldsymbol{y}_0\|)^2. \tag{99}$$

The corresponding static form is

$$\begin{aligned}
\mathrm{QOT}_S(\mu_0, \mu_1) &= \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\mathcal{X}^4} \mathcal{A}(\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{y}_0, \boldsymbol{y}_1)\gamma(\boldsymbol{x}_0, \boldsymbol{x}_1)\gamma(\boldsymbol{y}_0, \boldsymbol{y}_1)\mathrm{d}\boldsymbol{x}_0\mathrm{d}\boldsymbol{x}_1\mathrm{d}\boldsymbol{y}_0\mathrm{d}\boldsymbol{y}_1 \\
&= \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\mathcal{X}^4} (\|\boldsymbol{x}_1 - \boldsymbol{y}_1\| - \|\boldsymbol{x}_0 - \boldsymbol{y}_0\|)^2\gamma(\boldsymbol{x}_0, \boldsymbol{x}_1)\gamma(\boldsymbol{y}_0, \boldsymbol{y}_1)\mathrm{d}\boldsymbol{x}_0\mathrm{d}\boldsymbol{x}_1\mathrm{d}\boldsymbol{y}_0\mathrm{d}\boldsymbol{y}_1 \\
&= \mathrm{GWOT}(\mu_0, \mu_1)
\end{aligned} \tag{100}$$

which is equivalent to standard GW-OT (Mémoli, 2011; Klein et al., 2025). We point out that different to static OT which has a dynamic form, GW-OT does not naturally have an unique dynamic form. Because the static cost (98) is defined by only minimizing w.r.t r_t, which can not uniquely determine $\boldsymbol{q}_t$. Intuitively, the travelling pair path action only penalizes the variation of the length of $\boldsymbol{q}_t$, but doesn't penalize the rotation.

D.1.3 DYNAMIC ANALOG TO FGW-OT

Previous studies have combined the standard static OT and GW-OT together to obtain a fused GW-OT (FGW-OT) (Vayer et al., 2020; Klein et al., 2025), which simultaneously minimizes the kinetic energy and the preserves the local structure. Its cost is defined as a weighted sum of static OT cost and static GW-OT cost.

$$C_{\mathrm{FGW}}(\mu_0, \mu_1) = \alpha C_{\mathrm{GW}}(\mu_0, \mu_1) + (1-\alpha)C_{\mathrm{OT}}(\mu_0, \mu_1) \tag{101}$$

As discussed above, the Lagrangian (95) is also a weighted sum of OT and GW-OT up to a scalar, but in a dynamic sense. Hence, intuitively, the corresponding dynamic form can be viewed as a dynamic analog to FGW-OT.

As discussed in A.3, the Lagrangian (95) can be written as $\|\dot{\boldsymbol{c}}_t\|^2 + \frac{1}{4} r_t^2 \dot{\theta}_t^2 + (\frac{1}{4} + \lambda)\dot{r}_t^2$ under the C-Q decomposition and the polar coordinates on the plane spanned by $\boldsymbol{q}_0, \boldsymbol{q}_1$. The first term is the kinetic energy of the barycenter, which is an analog to the standard kinetic term $\frac{1}{2}\|\dot{\boldsymbol{x}}\|^2 + \frac{1}{2}\|\dot{\boldsymbol{y}}\|^2$. The third term exactly leads to the static path action of GW-OT (98). The second term further penalizes the variation of θ_t, i.e. the rotation of $\boldsymbol{q}_t$. Therefore, the Lagrangian (95) actually leads to a more strict preservation of local structure than FGW-OT.

We also derived the analytic form of the path action in A.3.

$$\mathcal{A}(\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{y}_0, \boldsymbol{y}_1) = \|\boldsymbol{c}_1 - \boldsymbol{c}_0\|^2 + (\frac{1}{4} + \lambda)(r_0^2 + r_1^2 - 2r_0 r_1 \cos k\alpha) \tag{102}$$

When $\lambda = 0$,

$$\begin{aligned} \mathcal{A}(\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{y}_0, \boldsymbol{y}_1) &= \|\boldsymbol{c}_1 - \boldsymbol{c}_0\|^2 + \frac{1}{4}(r_0^2 + r_1^2 - 2r_0 r_1 \cos\alpha) \\ &= \|\boldsymbol{c}_1 - \boldsymbol{c}_0\|^2 + \frac{1}{4}\|\boldsymbol{q}_1 - \boldsymbol{q}_0\|^2 = \frac{1}{2}\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|^2 + \frac{1}{2}\|\boldsymbol{y}_1 - \boldsymbol{y}_0\|^2. \end{aligned} \tag{103}$$

It reduces to OT. When $\lambda \to \infty$,

$$\begin{aligned} \lambda^{-1}\mathcal{A}(\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{y}_0, \boldsymbol{y}_1) &= \mathcal{O}(\lambda^{-1}) + (r_0^2 + r_1^2 - 2r_0 r_1 \cos\Big(\sqrt{\frac{1}{1+4\lambda}}\alpha\Big)) \\ &= \mathcal{O}(\lambda^{-1}) + (r_0^2 + r_1^2 - 2r_0 r_1(1 + \mathcal{O}(\lambda^{-1}))) \\ &= \mathcal{O}(\lambda^{-1}) + (r_0^2 + r_1^2 - 2r_0 r_1) \\ &= (r_1 - r_0)^2 + \mathcal{O}(\lambda^{-1}) \to (\|\boldsymbol{x}_1 - \boldsymbol{y}_1\| - \|\boldsymbol{x}_0 - \boldsymbol{y}_0\|)^2 \end{aligned} \tag{104}$$

It reduces to GW-OT. Therefore, for general $\lambda \in (0, +\infty)$, it can be viewed as a dynamic fusion of OT and GW-OT rather than FGW-OT, which is just a static fusion.

D.1.4 THE 1D CASE

We need $d \geq 2$ (d is the dimension of the space) to have (98) be equivalent to the static GW-OT cost. To be self-contained, we briefly discuss the 1D case ($d = 1$) here. When $d = 1$, the Lagrangian $\mathcal{L} = (\frac{\mathrm{d}}{\mathrm{d}t}|x_t - y_t|)^2 = \Big(\frac{\mathrm{d}}{\mathrm{d}t}(x_t - y_t)\Big)^2$. Let $q_1 = x_1 - y_1, q_0 = x_0 - y_0$, the path action is

$$\mathcal{A}(x_0, x_1, y_0, y_1) = \inf_{q_t} \int_0^1 |\dot{q}_t|^2 \mathrm{d}t \quad \text{s.t.} \quad q_0 = x_0 - y_0, \quad q_1 = x_1 - y_1. \tag{105}$$

The optimal path is $q_t = (1-t)q_0 + tq_1$, instead of $r_t = (1-t)r_0 + tr_1$. The path action is therefore

$$\mathcal{A}(x_0, x_1, y_0, y_1) = (q_1 - q_0)^2 = (x_1 - y_1 - x_0 + y_0)^2 \tag{106}$$

instead of $(|x_1 - y_1| - |x_0 - y_0|)^2$. Thus, when $d = 1$, our path action does not reduce to the static cost of GW-OT. In fact, the two cost are equal if and only if $(x_1 - y_1)(x_0 - y_0) \geq 0$. Geometrically, in a high dimensional space ($d \geq 2$), one can continuously deform the configuration and reverse the relative positions of two points while keeping the magnitude of their relative displacement equal to the linear interpolation between its endpoint values. But it is impossible in one dimension. Reversing the relative positions of two points necessarily forces their relative displacement to vanish at some intermediate time, so its magnitude cannot follow the linear interpolation between the initial and final values.

D.2 IGW-OT

Developing dynamic formulations of GW-like OT has long attracted considerable attention. A recent successful attempt is the inner product GW-OT (IGW-OT), proposed by (Zhang et al., 2026b), which replaces the cost based on discrepancies between pairwise distances in GW-OT with one based on discrepancies between inner products.

$$\text{IGW-OT}_S^2(\mu_0,\mu_1)=\inf_{\gamma\in\Pi(\mu_0,\mu_1)}\int_{\mathcal{X}^4}(\langle \boldsymbol{x}_1,\boldsymbol{y}_1\rangle-\langle \boldsymbol{x}_0,\boldsymbol{y}_0\rangle)^2\gamma(\boldsymbol{x}_0,\boldsymbol{x}_1)\gamma(\boldsymbol{y}_0,\boldsymbol{y}_1)\mathrm{d}\boldsymbol{x}_0\mathrm{d}\boldsymbol{x}_1\mathrm{d}\boldsymbol{y}_0\mathrm{d}\boldsymbol{y}_1. \tag{107}$$

They developed a dynamic formulation for the above static QOT problem, denoted as IGW-OT. Intuitively, IGW-OT likewise aims to preserve spatial structure as much as possible throughout the transport process. Different to TP-DATE, they focused on the gradient flow and Riemannian structure of IGW-OT.

We point out that, with an appropriate choice of Lagrangian within the TP-DATE framework, IGW-OT can also be recovered as a special case of the dynamic QOT formulation introduced in our work. Take

$$\mathcal{L}(t,\boldsymbol{x}_t,\boldsymbol{y}_t,\dot{\boldsymbol{x}}_t,\dot{\boldsymbol{y}}_t)=|\frac{\mathrm{d}}{\mathrm{d}t}\langle \boldsymbol{x}_t,\boldsymbol{y}_t\rangle|^2=|\langle\dot{\boldsymbol{x}}_t,\boldsymbol{y}_t\rangle+\langle \boldsymbol{x}_t,\dot{\boldsymbol{y}}_t\rangle|^2 \tag{108}$$

The path action is

$$\mathcal{A}(\boldsymbol{x}_0,\boldsymbol{x}_1,\boldsymbol{y}_0,\boldsymbol{y}_1)=\inf_{r_t}\int_0^1|\dot{r}_t|^2\mathrm{d}t\quad\text{s.t.}\quad r_0=\langle \boldsymbol{x}_0,\boldsymbol{y}_0\rangle,\quad r_1=\langle \boldsymbol{x}_1,\boldsymbol{y}_1\rangle. \tag{109}$$

Using the Euler-Lagrange equation, it is easy to show that when $d\geq 2$, the optimal r_t is attained by $r_t=(1-t)r_0+tr_1$. Therefore, the path action is

$$\mathcal{A}(\boldsymbol{x}_0,\boldsymbol{x}_1,\boldsymbol{y}_0,\boldsymbol{y}_1)=(r_1-r_0)^2=(\langle \boldsymbol{x}_1,\boldsymbol{y}_1\rangle-\langle \boldsymbol{x}_0,\boldsymbol{y}_0\rangle)^2. \tag{110}$$

The corresponding static form is

$$\begin{aligned}\text{QOT}_S(\mu_0,\mu_1)&=\inf_{\gamma\in\Pi(\mu_0,\mu_1)}\int_{\mathcal{X}^4}\mathcal{A}(\boldsymbol{x}_0,\boldsymbol{x}_1,\boldsymbol{y}_0,\boldsymbol{y}_1)\gamma(\boldsymbol{x}_0,\boldsymbol{x}_1)\gamma(\boldsymbol{y}_0,\boldsymbol{y}_1)\mathrm{d}\boldsymbol{x}_0\mathrm{d}\boldsymbol{x}_1\mathrm{d}\boldsymbol{y}_0\mathrm{d}\boldsymbol{y}_1\\&=\inf_{\gamma\in\Pi(\mu_0,\mu_1)}\int_{\mathcal{X}^4}(\langle \boldsymbol{x}_1,\boldsymbol{y}_1\rangle-\langle \boldsymbol{x}_0,\boldsymbol{y}_0\rangle)^2\gamma(\boldsymbol{x}_0,\boldsymbol{x}_1)\gamma(\boldsymbol{y}_0,\boldsymbol{y}_1)\mathrm{d}\boldsymbol{x}_0\mathrm{d}\boldsymbol{x}_1\mathrm{d}\boldsymbol{y}_0\mathrm{d}\boldsymbol{y}_1\\&=\text{IGW-OT}_S^2(\mu_0,\mu_1)\end{aligned} \tag{111}$$

The convexity of Lagrangian is also satisfied. To show that, let

$$\boldsymbol{v}_t=\begin{pmatrix}\dot{\boldsymbol{x}}_t\\ \dot{\boldsymbol{y}}_t\end{pmatrix}\qquad \boldsymbol{z}_t=\begin{pmatrix}\boldsymbol{y}_t\\ \boldsymbol{x}_t\end{pmatrix} \tag{112}$$

we have

$$|\langle\dot{\boldsymbol{x}}_t,\boldsymbol{y}_t\rangle+\langle \boldsymbol{x}_t,\dot{\boldsymbol{y}}_t\rangle|^2=|\boldsymbol{z}_t\cdot\boldsymbol{v}_t|^2=\boldsymbol{v}_t^{\mathrm{T}}(\boldsymbol{z}_t\boldsymbol{z}_t^{\mathrm{T}})\boldsymbol{v}_t. \tag{113}$$

Therefore, the Lagrangian is convex w.r.t $\boldsymbol{v}_t$, hence $(\dot{\boldsymbol{x}}_t,\dot{\boldsymbol{y}}_t)$. Thus, the static IGW objective is recovered as a special case of our path action construction, while the induced dynamic flow need not coincide with their intrinsic IGW flow (for mathematical details, please refer to their paper). From this perspective, our TP-DATE framework can be viewed as a more general dynamic QOT framework.

D.2.1 THE 1D CASE

In the IGW-OT case, we also need $d\geq 2$ to make $r_t=(1-t)r_0+tr_1$ attainable. We give an explicit construction here. Let $\boldsymbol{p}_t=\frac{\boldsymbol{x}_t+\boldsymbol{y}_t}{\sqrt{2}}$, $\boldsymbol{q}_t=\frac{\boldsymbol{x}_t-\boldsymbol{y}_t}{\sqrt{2}}$, we have $\langle \boldsymbol{x}_t,\boldsymbol{y}_t\rangle=\|\boldsymbol{p}_t\|^2-\|\boldsymbol{q}_t\|^2$. When $d\geq 2$, it is easy to construct $\boldsymbol{p}_t,\boldsymbol{q}_t$ such that $\|\boldsymbol{p}_t\|^2=(1-t)\|\boldsymbol{p}_0\|^2+t\|\boldsymbol{p}_1\|^2$, $\|\boldsymbol{q}_t\|^2=(1-t)\|\boldsymbol{q}_0\|^2+t\|\boldsymbol{q}_1\|^2$, hence $\langle \boldsymbol{x}_t,\boldsymbol{y}_t\rangle=(1-t)\langle \boldsymbol{x}_0,\boldsymbol{y}_0\rangle+t\langle \boldsymbol{x}_1,\boldsymbol{y}_1\rangle$.

When $d=1$, consider $x_0=1,y_0=-1,x_1=-1,y_1=1$. By direct calculation, we have $x_0y_0=x_1y_1=-1$. But, it is impossible to construct a travelling pair (x_t,y_t) such that $x_ty_t\equiv -1$, since there must exist some t such that $x_t=0$, hence $x_ty_t=0\neq -1$. The reason is similar to the GW-OT case. A reflection can be continuously realized by a rigid body rotation in high dimensional space $d\geq 2$, but can never be realized by a continuous rigid transformation in 1D spaces.

When $d = 1$, consider $\mathcal{L} = \left(\frac{\mathrm{d}}{\mathrm{d}t}(x_t y_t)\right)^2$, by direct calculation, one can easily check that the corresponding static cost is

$$\mathcal{A}(x_0, x_1, y_0, y_1) = \begin{cases} (|x_0 y_0| + |x_1 y_1|)^2, & x_0 x_1 < 0, y_0 y_1 < 0 \\ (x_1 y_1 - x_0 y_0)^2, & \text{otherwise} \end{cases} \tag{114}$$

The infimum is attained when $x_0 y_0 x_1 y_1 \neq 0$ or $x_0 y_0 = x_1 y_1 = 0$, but may not be attainable when $x_0 y_0 = 0, x_1 y_1 \neq 0$ or $x_0 y_0 \neq 0, x_1 y_1 = 0$.

D.3 OT-CFM

(Tong et al., 2024a) has proposed OT-CFM, a flow matching framework for solving standard dynamic OT problem. It uses the optimal OT coupling under squared Euclidean distance cost, and the displacement conditional path for flow matching training. The learned flow can be proved to solve the standard dynamic OT problem with kinetic energy cost.

$$\begin{aligned} &\mathrm{OT}(\mu_0, \mu_1) = \inf_{\rho, \boldsymbol{u}} \int_0^1 \int_{\mathcal{X}} \frac{1}{2} \|\boldsymbol{u}(\boldsymbol{x}, t)\|_2^2 \rho_t(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \mathrm{d}t \\ &\text{s.t.} \quad \partial_t \rho + \nabla_{\boldsymbol{x}} \cdot (\rho \boldsymbol{u}) = 0, \ \rho_0 = \mu_0, \ \rho_1 = \mu_1 \end{aligned} \tag{115}$$

Inspired by this pioneering work, we proposed a new travelling pair flow matching framework which can solve the dynamic QOT problem. When choosing the Lagrangian in our dynamic QOT framework as

$$\mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, \dot{\boldsymbol{x}}, \dot{\boldsymbol{y}}) = \frac{1}{2} \|\dot{\boldsymbol{x}}\|^2 + \frac{1}{2} \|\dot{\boldsymbol{y}}\|^2 \tag{116}$$

the corresponding dynamic QOT reduces to (115). Therefore, TP-DATE reduces to OT-CFM under this setting.

D.4 Static structured transport in spatial transcriptomics

PASTE (Zeira et al., 2022) and PASTE2 (Liu et al., 2023) formulate spatial transcriptomic slice alignment using FGW and partial-FGW, respectively, and primarily target spatial alignment, integration, and 3D reconstruction. MOSCOT (Klein et al., 2025) provides a unified OT framework for mapping cells across time and space, including GW- and FGW-type formulations for spatial and spatiotemporal correspondence. SOCS (Bryan et al., 2026) further targets time-series spatial transcriptomics by estimating an ancestor–descendant transport matrix between two observed time points while preserving gene-expression, geometric, and contiguous-structure information. Despite their different application settings, these methods operate at the level of couplings between observed datasets and do not model intermediate time measure paths or dynamical vector fields. In this sense, they perform static endpoint alignment. In contrast, TP-DATE goes beyond endpoint alignment by reconstructing continuous-time dynamics, and thus constitutes a dynamic trajectory inference framework.

D.5 GENOT

(Klein et al., 2024) also combines flow matching with GW-OT, but with a fundamentally different purpose. GENOT is a neural OT solver that uses flow matching to learn couplings associated with static entropic OT, GW-OT, or FGW-OT. In contrast, TP-DATE does not use flow matching to solve the QOT coupling. It uses flow matching to find the dynamic QOT probability flow. The flow in GENOT serves as a parameterization of a static transport plan, while the flow in TP-DATE represents the reconstructed dynamics between snapshots. The two approaches therefore address different problems.

D.6 NicheFlow

NicheFlow (Sakalyan et al., 2025) advances spatial trajectory inference by modeling local cellular microenvironments as structured point clouds, jointly capturing changes in spatial coordinates and gene expression states. It combines OT-based source–target matching with Variational Flow Matching

(VFM) to generate future microenvironments conditioned on observed source niches. However, its flow evolves from noise to the target state conditioned on the source, so the internal flow time represents a generative process rather than the continuous time biological dynamics between two tissue states. In contrast, TP-DATE directly models source-to-target continuous time dynamics and imposes structure preservation at the dynamical level through dynamic QOT.

D.7 CONTEXTFLOW

ContextFlow (Rathod et al., 2026) is a recently proposed context-aware generative modeling approach for transcriptomic snapshot interpolation. It incorporates spatial information and cell–cell communication signals into a static OT cost, thereby obtaining couplings that account for these contextual factors. These couplings are then used in conditional flow matching to reconstruct the dynamics. From this perspective, ContextFlow can be viewed as introducing a class of context-aware static OT formulations. In contrast, TP-DATE incorporates spatial structure directly at the dynamical level rather than indirectly through a static coupling. Viewed this way, TP-DATE extends the context-aware principle to dynamic settings through the framework of dynamic QOT.

D.8 STVCR

stVCR (Peng et al., 2026a) models spatiotemporal dynamics of single cells with Wasserstein-Fisher-Rao dynamics (WFR). WFR can be viewed as an unbalanced extension of standard OT. It extends OT from transporting probability densities to transporting unbalanced measures.

$$\begin{aligned} &\mathrm{WFR}_\delta^2(\mu_0,\mu_1) = \inf_{\rho,\boldsymbol{u}} \int_0^1 \int_{\mathcal{X}} \frac{1}{2}[\|\boldsymbol{u}(\boldsymbol{x},t)\|_2^2 + \delta^2 g_t^2(\boldsymbol{x})]\rho_t(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t \\ &\text{s.t.} \qquad \partial_t\rho + \nabla_{\boldsymbol{x}}\cdot(\rho\boldsymbol{u}) = g\rho,\ \rho_0 = \mu_0,\ \rho_1 = \mu_1 \end{aligned} \tag{117}$$

It models the cell proliferation and apoptosis by a growth rate term $g_t(\boldsymbol{x})$. When applied on scRNA-seq data without spatial information, this WFR problem can be solved by unbalanced flow matching (Peng et al., 2026b). On spatial transcriptomics data, to preserve the spatial structure, an optional spatial structure preserving loss is added. To account for possible slight misalignment between the coordinate systems of spatial transcriptomics slices, stVCR further introduces a rigid body transformation invariant OT formulation. Before each OT training step, it first solves an optimal rigid body transformation to align the coordinate systems. Both techniques make it unsolvable for flow matching based methods. Therefore, stVCR used NeuralODE method to solve the corresponding OT problem.

In contrast, TP-DATE naturally preserves the spatial structure and addresses slight misalignment by adding interaction term Φ, and is simulation-free. Although TP-DATE is more efficient than stVCR in modeling spatial structure preservation, it does not currently account for cell proliferation and apoptosis as stVCR does. This limitation points to an important direction for future work: extending the dynamic QOT framework to the unbalanced setting.

D.9 CYTOBRIDGE

CytoBridge (Zhang et al., 2025b) is a NeuralODE framework for learning interactions and mean-field interactive dynamics from snapshots. They focuses on the unbalanced mean-field Schrödinger bridge problem (UMFSB).

$$\begin{aligned} &\mathrm{UMFSB}(\mu_0,\mu_1) = \inf_{\rho,\boldsymbol{u}} \int_0^1 \int_{\mathcal{X}} \frac{1}{2}[\|\boldsymbol{u}(\boldsymbol{x},t)\|_2^2 + g_t^2(\boldsymbol{x})]\rho_t(\boldsymbol{x})\mathrm{d}\boldsymbol{x}\mathrm{d}t \\ &\text{s.t.}\ \ \partial_t\rho(\boldsymbol{x}) + \nabla_{\boldsymbol{x}}\cdot[\rho(\boldsymbol{x})(\boldsymbol{u}_t(\boldsymbol{x}) - \int k(\boldsymbol{x},\boldsymbol{y})\nabla_{\boldsymbol{x}}\Phi(\boldsymbol{x}-\boldsymbol{y})\rho_t(\boldsymbol{y})\mathrm{d}\boldsymbol{y})] = g\rho \\ &\qquad\qquad \rho_0 = \mu_0,\ \rho_1 = \mu_1 \end{aligned} \tag{118}$$

where Φ is some interaction potential for modeling cell-cell interactions, and k is some gate function to control the intensity of the interaction. The key distinction between CytoBridge and TP-DATE is where the interaction term is introduced. CytoBridge incorporates the interaction term directly into the dynamical equation as part of the constraint, whereas TP-DATE retains the standard continuity

equation constraint on the two-particle space and instead places the interaction term in the action functional.

One advantage of CytoBridge is that, because the interaction term is part of the dynamics to be learned, the interaction itself can be inferred from data through a NeuralODE. This is feasible for NeuralODE based methods. In principle, essentially any interaction term can be parameterized by a neural network and learned by minimizing the loss. For flow matching based methods, however, this is generally much more challenging, since the conditional paths determined by the interaction term must typically be computed in advance. Whether techniques from inverse problems can be used to overcome this limitation and enable flow matching methods to learn such interactions from the data is therefore a very interesting direction for future research.

Another interesting question is whether CytoBridge can be made simulation-free. From TrajectoryNet (Tong et al., 2020) to OT-CFM (Tong et al., 2024a), Diffusion Schrödinger bridge (De Bortoli et al., 2021; Shi et al., 2023) to SF^{2}M (Tong et al., 2024b), stVCR (Peng et al., 2026a) to WFR-FM Peng et al. (2026b), previous works have already built simulation-free flow matching based algorithms for OT, Schrödinger bridge, and WFR. There is a trend of converting simulation-based or iteration-based Neural OT solvers to simulation-free solvers by flow matching. However, the success of these algorithms rely on an important mathematical property of the underlying OT problems: the dynamical constraint is linear, or equivalently, local, with respect to the measure flow ρ_t. This property is precisely what allows the marginalization theorem to hold in these settings. Once a CytoBridge style interaction term is introduced directly into the dynamical constraint, this linearity or locality is lost. One can readily verify that, for the CytoBridge constraint equation, the usual marginalization argument no longer applies. This limitation is also reflected in the independence of conditional paths in standard flow matching. Once the initial and terminal states are fixed, each conditional path is determined independently of the others, with no coupling between different paths. As a result, standard conditional path constructions are not naturally equipped to represent interactions between particles. From this perspective, extending CytoBridge to a fully simulation-free flow matching formulation appears to be fundamentally challenging.

TP-DATE's travelling-pair flow matching may provide a possible starting point for addressing this challenge. By considering the problem on the two-particle space, we simultaneously enable interactions between conditional paths while preserving the linearity of the dynamical constraint. Moreover, through velocity decomposition, the resulting single-particle dynamics can be interpreted as an approximation to mean-field dynamics, in which interactions at the population level emerge from underlying pairwise interactions. Although TP-DATE was not originally designed for this purpose, a promising direction for future work is to build on travelling-pair flow matching and investigate whether formulating the problem on the two-particle space can enable efficient flow matching based modeling of nonlocal dynamics.

E VELOCITY DECOMPOSITION

In general cases, one may want to only have a velocity decomposition at conditional velocity level $\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}') = \boldsymbol{b}_t(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{z}') + \boldsymbol{f}_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')$. In this case, we want to regress the marginal velocities via flow matching.

$$\begin{aligned}\mathcal{L}_b(\boldsymbol{\theta}) &= \mathbb{E}_{t\sim\mathcal{U}[0,1],(\boldsymbol{z},\boldsymbol{z}')\sim q(\boldsymbol{z},\boldsymbol{z}'),(\boldsymbol{x},\boldsymbol{y})\sim\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')} \|\boldsymbol{b}_{\boldsymbol{\theta}}(\boldsymbol{x},t) - \boldsymbol{b}_t(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{z}')\|_2^2 \\ \mathcal{L}_f(\boldsymbol{\phi}) &= \mathbb{E}_{t\sim\mathcal{U}[0,1],(\boldsymbol{z},\boldsymbol{z}')\sim q(\boldsymbol{z},\boldsymbol{z}'),(\boldsymbol{x},\boldsymbol{y})\sim\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')} \|\boldsymbol{f}_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{y},t) - \boldsymbol{f}_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')\|_2^2\end{aligned} \tag{119}$$

Similarly to theorem 5.6, we still have

$$\begin{aligned}\mathcal{L}_b(\boldsymbol{\theta}) &= \mathbb{E}_{t\sim\mathcal{U}[0,1],(\boldsymbol{x},\boldsymbol{y})\sim\pi_t(\boldsymbol{x},\boldsymbol{y})} \|\boldsymbol{b}_{\boldsymbol{\theta}}(\boldsymbol{x},t) - \boldsymbol{b}_t(\boldsymbol{x})\|_2^2 + C_1 \\ &= \mathbb{E}_{t\sim\mathcal{U}[0,1],\boldsymbol{x}\sim\rho_t(\boldsymbol{x})} \|\boldsymbol{b}_{\boldsymbol{\theta}}(\boldsymbol{x},t) - \boldsymbol{b}_t(\boldsymbol{x})\|_2^2 + C_1 \\ \mathcal{L}_f(\boldsymbol{\phi}) &= \mathbb{E}_{t\sim\mathcal{U}[0,1],(\boldsymbol{x},\boldsymbol{y})\sim\pi_t(\boldsymbol{x},\boldsymbol{y})} \|\boldsymbol{f}_{\boldsymbol{\phi}}(\boldsymbol{x},\boldsymbol{y},t) - \boldsymbol{f}_t(\boldsymbol{x},\boldsymbol{y})\|_2^2 + C_2\end{aligned} \tag{120}$$

where C_1, C_2 are constants independent of $\boldsymbol{\theta}, \boldsymbol{\phi}$. Therefore, the minimizer is $\boldsymbol{b}_t(\boldsymbol{x})$ and $\boldsymbol{f}_t(\boldsymbol{x},\boldsymbol{y})$. The problem here is that the conditional velocity decomposition may not lead to a marginal velocity decomposition, i.e. $\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{b}_t(\boldsymbol{x}) + \boldsymbol{f}_t(\boldsymbol{x},\boldsymbol{y})$ does not naturally hold. By marginalization

theorems, we have

$$\begin{cases} \boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y}) = \int \boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')\dfrac{\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')q(\boldsymbol{z},\boldsymbol{z}')}{\pi_t(\boldsymbol{x},\boldsymbol{y})}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}' \\ \boldsymbol{b}_t(\boldsymbol{x}) = \int \Big(\int \boldsymbol{b}_t(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{z}')\dfrac{\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')q(\boldsymbol{z},\boldsymbol{z}')}{\pi_t(\boldsymbol{x},\boldsymbol{y})}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'\Big)\dfrac{\pi_t(\boldsymbol{x},\boldsymbol{y})}{\rho_t(\boldsymbol{x})}\mathrm{d}\boldsymbol{y} \\ \boldsymbol{f}_t(\boldsymbol{x},\boldsymbol{y}) = \int \boldsymbol{f}_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')\dfrac{\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')q(\boldsymbol{z},\boldsymbol{z}')}{\pi_t(\boldsymbol{x},\boldsymbol{y})}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}' \end{cases} \tag{121}$$

The conditional velocity decomposition yields

$$\begin{aligned} &\mathbb{E}[\boldsymbol{b}_t(\boldsymbol{X}|\boldsymbol{z},\boldsymbol{z}')|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Y}=\boldsymbol{y}] \\ =&\int \boldsymbol{b}_t(\boldsymbol{x}|\boldsymbol{z},\boldsymbol{z}')\frac{\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')q(\boldsymbol{z},\boldsymbol{z}')}{\pi_t(\boldsymbol{x},\boldsymbol{y})}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}' \\ =&\int [\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}') - \boldsymbol{f}_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')]\frac{\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')q(\boldsymbol{z},\boldsymbol{z}')}{\pi_t(\boldsymbol{x},\boldsymbol{y})}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}' \\ =&\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y}) - \boldsymbol{f}_t(\boldsymbol{x},\boldsymbol{y}) \end{aligned} \tag{122}$$

By definition,

$$\boldsymbol{b}_t(\boldsymbol{x}) = \mathbb{E}[\boldsymbol{b}_t(\boldsymbol{X}|\boldsymbol{z},\boldsymbol{z}')|\boldsymbol{X}=\boldsymbol{x}] \tag{123}$$

Therefore, the necessary and sufficient condition of marginal velocity decomposition is

$$\mathbb{E}[\boldsymbol{b}_t(\boldsymbol{X}|\boldsymbol{z},\boldsymbol{z}')|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Y}=\boldsymbol{y}] = \mathbb{E}[\boldsymbol{b}_t(\boldsymbol{X}|\boldsymbol{z},\boldsymbol{z}')|\boldsymbol{X}=\boldsymbol{x}] \tag{124}$$

Intuitively, the condition requires that, once the current state of the particle is given, the state of the interacting particle provides no additional information about the conditional mean of the self-driven velocity. It is a conditional mean-independence condition, which is weaker than conditional independence. The simplest example is what we introduced in the main text $\boldsymbol{b}_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}') = \boldsymbol{b}_t(\boldsymbol{x})$. It is easy to check that it satisfies the condition. Other kind of decompositions have to be designed carefully to satisfy this condition, in order to be learned by flow matching.

F THE BB-FORM

For convenience, we restate the three forms below. The static form is

$$\mathrm{QOT}_S(\mu_0,\mu_1) = \inf_{\gamma\in\Pi(\mu_0,\mu_1)}\int_{\mathcal{X}^4}\mathcal{A}(\boldsymbol{z},\boldsymbol{z}')\gamma(\boldsymbol{z})\gamma(\boldsymbol{z}')\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'. \tag{125}$$

The dynamic form is

$$\begin{aligned} &\mathrm{QOT}_D(\mu_0,\mu_1) = \\ &\inf\int_0^1\int\int_{\mathcal{X}^2}\mathcal{L}(t,\boldsymbol{x},\boldsymbol{y},u_t^1(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}'),u_t^2(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}'))\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')\gamma(\boldsymbol{z})\gamma(\boldsymbol{z}')\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'\mathrm{d}t. \end{aligned} \tag{126}$$

The BB-form is

$$\begin{aligned} &\mathrm{QOT}_{BB}(\mu_0,\mu_1) = \inf_{\pi,\boldsymbol{u}^1,\boldsymbol{u}^2}\int_0^1\int_{\mathcal{X}^2}\mathcal{L}(t,\boldsymbol{x},\boldsymbol{y},u_t^1(\boldsymbol{x},\boldsymbol{y}),u_t^2(\boldsymbol{x},\boldsymbol{y}))\pi_t(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}t \\ &\partial_t\pi_t(\boldsymbol{x},\boldsymbol{y}) + \nabla_{\boldsymbol{x}}\cdot(\pi_t(\boldsymbol{x},\boldsymbol{y})u_t^1(\boldsymbol{x},\boldsymbol{y})) + \nabla_{\boldsymbol{y}}\cdot(\pi_t(\boldsymbol{x},\boldsymbol{y})u_t^2(\boldsymbol{x},\boldsymbol{y})) = 0. \end{aligned} \tag{127}$$

F.1 DIFFICULTY OF THE EQUIVALENCE BETWEEN STATIC QOT AND BB-FORM

Inspired by the pioneering work of (Benamou & Brenier, 2000), one may expect to have an equivalence between the static form and the BB-form also for QOT. We explain why it can not. Let $\boldsymbol{z} = (\boldsymbol{x},\boldsymbol{y})$, and $\boldsymbol{u}_t(\boldsymbol{z}) = \begin{pmatrix} u_t^1(\boldsymbol{x},\boldsymbol{y}) \\ u_t^2(\boldsymbol{x},\boldsymbol{y}) \end{pmatrix}$, the BB-form can be reformulated as

$$\begin{aligned} &\mathrm{QOT}_{BB}(\mu_0,\mu_1) = \inf_{\pi,\boldsymbol{u}}\int_0^1\int_{\mathcal{X}^2}\mathcal{L}(t,\boldsymbol{z},u_t(\boldsymbol{z}))\pi_t(\boldsymbol{z})\mathrm{d}\boldsymbol{z}\mathrm{d}t \\ &\text{s.t.} \qquad \partial_t\pi_t(\boldsymbol{z}) + \nabla_{\boldsymbol{z}}\cdot(\pi_t(\boldsymbol{z})u_t(\boldsymbol{z})) = 0. \end{aligned} \tag{128}$$

which is just the standard OT on the two-particle space $\mathcal{X}^2$. When choosing $\mathcal{L}(t, \boldsymbol{z}, u_t(\boldsymbol{z})) = \frac{1}{2}\|\dot{\boldsymbol{z}}\|^2$, (Benamou & Brenier, 2000) proved that it is equivalent to the static OT on the two-particle space

$$\mathrm{OT}(\mu_0, \mu_1) = \inf_{q \in \Pi(\mu_0 \otimes \mu_0, \mu_1 \otimes \mu_1)} \int_{\mathcal{X}^4} \mathcal{A}(\boldsymbol{z}, \boldsymbol{z}') q(\boldsymbol{z}, \boldsymbol{z}') \mathrm{d}\boldsymbol{z} \mathrm{d}\boldsymbol{z}'. \tag{129}$$

This static form is slightly different to the static QOT. In static QOT (125), we restrict the coupling q to a smaller set

$$\Pi^2(\mu_0, \mu_1) \triangleq \{q = \gamma \otimes \gamma | \gamma \in \Pi(\mu_0, \mu_1)\} \subset \Pi(\mu_0 \otimes \mu_0, \mu_1 \otimes \mu_1). \tag{130}$$

From this perspective, the static QOT on $\mathcal{X}$ can be viewed as kind of special OT on $\mathcal{X}^2$ with more strict coupling constraint. Therefore, it is naturally to have only $\mathrm{QOT}_S(\mu_0, \mu_1) \geq \mathrm{QOT}_{BB}(\mu_0, \mu_1)$, and it is essentially hard to improve it.

F.2 Difficulty of the equivalence between static QOT and the constrained BB-form

One may naturally ask whether the BB-form can also be equipped with additional constraints so that it becomes compatible with the extra coupling constraint imposed in static QOT. The answer is also no.

By superposition principle (Ambrosio et al., 2008; Lisini, 2007), for any admissible probability path $\pi_t(\boldsymbol{x}, \boldsymbol{y})$ satisfying the continuity equation $\partial_t \pi_t(\boldsymbol{x}, \boldsymbol{y}) + \nabla_{\boldsymbol{x}} \cdot (\pi_t(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{u}_t^1(\boldsymbol{x}, \boldsymbol{y})) + \nabla_{\boldsymbol{y}} \cdot (\pi_t(\boldsymbol{x}, \boldsymbol{y}) \boldsymbol{u}_t^2(\boldsymbol{x}, \boldsymbol{y})) = 0$, the superposition principle states that there exists a probability measure η over the path space $C([0,1], \mathcal{X}^2)$ such that for a path $(\boldsymbol{x}_t, \boldsymbol{y}_t)$, $\dot{\boldsymbol{x}}_t = \boldsymbol{u}_t^1(\boldsymbol{x}_t, \boldsymbol{y}_t), \dot{\boldsymbol{y}}_t = \boldsymbol{u}_t^2(\boldsymbol{x}_t, \boldsymbol{y}_t)$ holds $\eta - a.e.$, and the projection of η at time t is π_t. That means, for path Γ and evaluation mapping $e_t : e_t(\Gamma) = \Gamma_t$, we have $(e_t)_{\#}\eta = \pi_t$. Intuitively, it says that a probability path generated by a continuity equation can be viewed as a superposition of several Dirac-Dirac paths which in this paper we also called the conditional paths.

To be compatible to the $\gamma \otimes \gamma$ structure of static QOT, the condition

$$\exists \gamma \in \Pi(\mu_0, \mu_1), \quad \text{s.t.} \quad \pi_{0,1} \triangleq (e_0, e_1)_{\#}\eta = \gamma \otimes \gamma \tag{131}$$

naturally follows. We call it the QOT compatibility condition. With this condition, we define the QOT compatible BB-form

$$\begin{aligned} &\mathrm{QOT}_C(\mu_0, \mu_1) = \inf_{\pi, \boldsymbol{u}^1, \boldsymbol{u}^2} \int_0^1 \int_{\mathcal{X}^2} \mathcal{L}(t, \boldsymbol{x}, \boldsymbol{y}, u_t^1(\boldsymbol{x}, \boldsymbol{y}), u_t^2(\boldsymbol{x}, \boldsymbol{y})) \pi_t(\boldsymbol{x}, \boldsymbol{y}) \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{y} \mathrm{d}t \\ &\text{s.t.} \quad \exists \gamma \in \Pi(\mu_0, \mu_1), \quad \pi_{0,1} = \gamma \otimes \gamma \\ &\partial_t \pi_t(\boldsymbol{x}, \boldsymbol{y}) + \nabla_{\boldsymbol{x}} \cdot (\pi_t(\boldsymbol{x}, \boldsymbol{y}) u_t^1(\boldsymbol{x}, \boldsymbol{y})) + \nabla_{\boldsymbol{y}} \cdot (\pi_t(\boldsymbol{x}, \boldsymbol{y}) u_t^2(\boldsymbol{x}, \boldsymbol{y})) = 0. \end{aligned} \tag{132}$$

With this condition, one can prove the inverse inequality $\mathrm{QOT}_S(\mu_0, \mu_1) \leq \mathrm{QOT}_C(\mu_0, \mu_1)$.

Proof. For any conditional path $(\boldsymbol{x}_{[0,1]}, \boldsymbol{y}_{[0,1]})$, let $\boldsymbol{z} = (\boldsymbol{x}_0, \boldsymbol{x}_1), \boldsymbol{z}' = (\boldsymbol{y}_0, \boldsymbol{y}_1)$, we have

$$\int_0^1 \mathcal{L}(t, \boldsymbol{x}_t, \boldsymbol{y}_t, \dot{\boldsymbol{x}}_t, \dot{\boldsymbol{y}}_t) \mathrm{d}t \geq \mathcal{A}(\boldsymbol{z}, \boldsymbol{z}') \tag{133}$$

by definition. Now consider the projection of η on the two time points $\{0,1\}$. Since $(e_0,e_1)_{\#}\eta = \pi_{0,1}$, there exists a coupling $\gamma \in \Pi(\mu_0,\mu_1)$ such that $(e_0,e_1)_{\#}\eta = \gamma\otimes\gamma$. By (133), we have

$$
\begin{aligned}
&\int \mathcal{L}(t,\boldsymbol{x},\boldsymbol{y},\boldsymbol{u}_t^1(\boldsymbol{x},\boldsymbol{y}),\boldsymbol{u}_t^2(\boldsymbol{x},\boldsymbol{y}))\pi_t(\boldsymbol{x},\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}\mathrm{d}t\\
&=\int \mathcal{L}(t,\boldsymbol{x}_t,\boldsymbol{y}_t,\boldsymbol{u}_t^1(\boldsymbol{x}_t,\boldsymbol{y}_t),\boldsymbol{u}_t^2(\boldsymbol{x}_t,\boldsymbol{y}_t))\mathrm{d}\eta(\boldsymbol{x}_{[0,1]},\boldsymbol{y}_{[0,1]})\mathrm{d}t\\
&=\int \mathcal{L}(t,\boldsymbol{x}_t,\boldsymbol{y}_t,\dot{\boldsymbol{x}}_t,\dot{\boldsymbol{y}}_t)\mathrm{d}\eta(\boldsymbol{x}_{[0,1]},\boldsymbol{y}_{[0,1]})\mathrm{d}t\\
&=\int \Big(\int_0^1 \mathcal{L}(t,\boldsymbol{x}_t,\boldsymbol{y}_t,\dot{\boldsymbol{x}}_t,\dot{\boldsymbol{y}}_t)\mathrm{d}t\Big)\mathrm{d}\eta(\boldsymbol{x}_{[0,1]},\boldsymbol{y}_{[0,1]})\\
&=\int \Big(\int_0^1 \mathcal{L}(t,\boldsymbol{x}_t,\boldsymbol{y}_t,\dot{\boldsymbol{x}}_t,\dot{\boldsymbol{y}}_t)\mathrm{d}t\Big)\mathrm{d}(e_0,e_1)_{\#}\eta(\boldsymbol{x}_{[0,1]},\boldsymbol{y}_{[0,1]})\\
&=\int \Big(\int_0^1 \mathcal{L}(t,\boldsymbol{x}_t,\boldsymbol{y}_t,\dot{\boldsymbol{x}}_t,\dot{\boldsymbol{y}}_t)\mathrm{d}t\Big)\gamma(\boldsymbol{z})\gamma(\boldsymbol{z}')\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'\\
&\geq \int \mathcal{A}(\boldsymbol{z},\boldsymbol{z}')\gamma(\boldsymbol{z})\gamma(\boldsymbol{z}')\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}'\\
&\geq \mathrm{QOT}_S(\mu_0,\mu_1)
\end{aligned}
\tag{134}
$$

By taking infimum w.r.t $\pi_t(\boldsymbol{x},\boldsymbol{y})$, we have $\mathrm{QOT}_C(\mu_0,\mu_1)\geq \mathrm{QOT}_S(\mu_0,\mu_1)$. □

Although we can prove the inverse inequality, unfortunately, the original side cannot be guaranteed in general under this condition. That is, we no longer have $\mathrm{QOT}_C(\mu_0,\mu_1)\leq \mathrm{QOT}_S(\mu_0,\mu_1)$. The reason is that, we proved the original inequality by Jensen's inequality in A.1. To use Jensen's inequality, we marginalize all the travelling pair conditional velocities into a marginal velocity pair

$$
\boldsymbol{u}_t^i(\boldsymbol{x},\boldsymbol{y}) = \int \boldsymbol{u}_t^i(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')\frac{\pi_t(\boldsymbol{x},\boldsymbol{y}|\boldsymbol{z},\boldsymbol{z}')q(\boldsymbol{z},\boldsymbol{z}')}{\pi_t(\boldsymbol{x},\boldsymbol{y})}\mathrm{d}\boldsymbol{z}\mathrm{d}\boldsymbol{z}' \tag{135}
$$

where $q=\gamma\otimes\gamma$. The point here is that although q has the tensor structure, the probability flow $\pi_t(\boldsymbol{x},\boldsymbol{y})$ generated by the marginal velocity pair does not naturally satisfy the QOT compatible condition. Therefore, we can only prove $\mathrm{QOT}_S(\mu_0,\mu_1)\geq \mathrm{QOT}_{BB}(\mu_0,\mu_1)$, but can never prove $\mathrm{QOT}_S(\mu_0,\mu_1)\geq \mathrm{QOT}_C(\mu_0,\mu_1)$. The optimal result we can obtain is therefore $\mathrm{QOT}_C(\mu_0,\mu_1)\geq \mathrm{QOT}_S(\mu_0,\mu_1)\geq \mathrm{QOT}_{BB}(\mu_0,\mu_1)$.

F.3 The intuitive relation between the three forms

In the main text, we define the static QOT, the dynamic QOT and the BB-form. The intuition of the BB-form is to minimize the population level action induced by the Lagrangian during the transport, which is what, in practice, one may actually want to minimize. Since $\mathrm{QOT}_D(\mu_0,\mu_1) = \mathrm{QOT}_S(\mu_0,\mu_1)\geq \mathrm{QOT}_{BB}(\mu_0,\mu_1)$, minimizing the static QOT action is equivalent to minimizing a tractable upper bound of the BB-form. Intuitively, it also implicitly tries to minimize the BB-form action. But, the static form contains no dynamic information. To model the continuous time dynamics, we have to extend it to the dynamic form, which is equivalent to the static form. In the main text, we introduce flow matching to solve the dynamic QOT flow by obtaining a Markovian projection. By Jensen's inequality, the BB-form action induced by this marginal flow is less than the corresponding static action, but greater than the infimum BB-form action by definition. Therefore, the learned flow can be viewed as a finer upper bound approximation of the true BB-form action.