ESFM: RECONSTRUCTING UNBALANCED DYNAMICS VIA REDUCTION AND EXTENDED-SPACE FLOW MATCHING

Anonymous authors
Paper under double-blind review

ABSTRACT

Biological population dynamics are often unbalanced, involving simultaneous changes in cellular states and population abundance. Existing unbalanced flow-matching methods typically rely on cone geometry-based loss formulations. While mathematically consistent, these formulations often induce biologically unfavorable state evolutions. We propose **Extended Space Flow Matching (ESFM)**, which lifts unbalanced dynamics into conservative transport in an extended space of state and log mass. This jointly models state transitions and abundance changes as a single conservative process, enabling prescribed conditional paths to be integrated with compatible endpoint couplings and simulation-free flow matching objectives. For linear interpolation of state and log mass, ESFM admits an analytical reduction to a tractable optimal transport formulation. On synthetic and real single-cell snapshot datasets, ESFM accurately predicts both state and abundance dynamics. Extended space formulations thus offer a flexible, biologically preferred framework for unbalanced flow matching.

1 INTRODUCTION

Single-cell RNA sequencing (scRNA-seq) provides discrete snapshots of a continuously evolving biological system. Because sequencing is destructive, individual cells cannot be followed over time, while experimental cost limits data collection to a few time points (Zheng et al., 2017; Bunne et al., 2024; Zhang et al., 2025b). Single-cell trajectory inference therefore seeks to recover continuous dynamics from population snapshots (Schiebinger et al., 2019; Tong et al., 2020).

Continuous-time models provide several approaches to this problem. Neural ordinary differential equations (ODEs) parameterize time-dependent velocity fields but typically require repeated numerical integration during training (Chen et al., 2018; Tong et al., 2020; Huguet et al., 2022). Dynamic optimal transport (OT) provides a complementary geometric view, with the Benamou–Brenier formulation linking static endpoint couplings to least-action continuous flows (Benamou & Brenier, 2000). Flow matching (FM) instead learns continuous flows by regressing conditional vector fields without simulating the ODE during training (Lipman et al., 2023; Liu et al., 2023; Albergo & Vanden-Eijnden, 2023; Klein et al., 2024; Rohbeck et al., 2025). In general FM constructions, the endpoint coupling and conditional path may be specified separately. OT conditional flow matching (OT-CFM) links them through a common transport geometry by combining OT endpoint couplings with displacement interpolation (Tong et al., 2024a; Pooladian et al., 2023). This coupling–path consistency provides a principled geometric basis for flow matching. However, standard FM and OT-CFM operate on normalized probability measures and do not directly model changes in total population abundance.

Biological populations can grow or decline through proliferation, death, and differential survival (Sha et al., 2024; Zhang et al., 2025a; Sun et al., 2025; Peng et al., 2026c). Unbalanced OT relaxes mass conservation through source terms or relaxed marginal constraints (Liero et al., 2016; Chizat et al., 2018a;b), and recent unbalanced flow matching methods incorporate mass variation through unbalanced couplings, explicit growth dynamics, or unbalanced transport geometries (Eyring et al., 2024; Wang et al., 2025; Peng et al., 2026b; Ying et al., 2026b). A prominent class of unbalanced flow-matching methods derives endpoint couplings and conditional dynamics from a prescribed unbalanced transport geometry, particularly WFR formulation (Liero et al., 2016; Chizat 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.

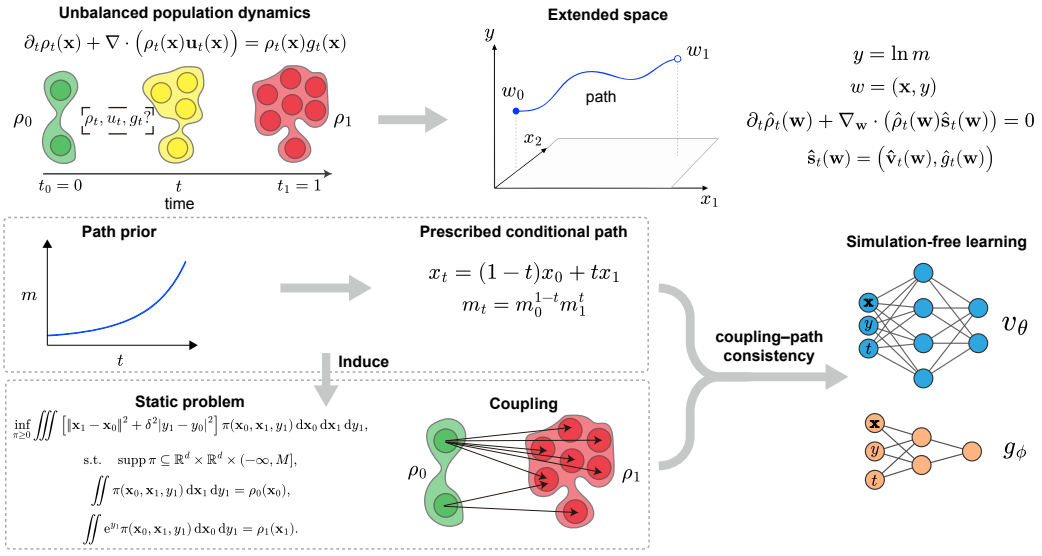


Figure 1: **Overview of Extended Space Flow Matching (ESFM).** ESFM lifts unbalanced dynamics to conservative transport in the extended space $\boldsymbol{w} = (\boldsymbol{x}, \ln m)$, starts from a prescribed conditional path, and derives a compatible endpoint coupling and flow matching objective.

2018a;b; Peng et al., 2026b). These approaches follow a forward design principle: a prescribed transport geometry or growth penalty in expression state space determines both the endpoint coupling and the associated conditional dynamics. While this construction ensures internal consistency, it does not necessarily yield biologically favorable state evolutions, as the resulting dynamics are driven solely by a prescribed, simplistic loss function.

We address unbalanced dynamics through a reduction approach in an extended space and propose the **Extended Space Flow Matching (ESFM)** framework for unbalanced single-cell dynamics. ESFM introduces a log mass coordinate alongside the cell state, lifting nonconservative dynamics into conservative transport in an extended space. In this representation, state transitions and relative growth are modeled jointly, while physical population mass is recovered by projection.

Within this formulation, ESFM adopts a path-first construction for the specific instance studied here, where linear interpolation of cell state and log mass is paired with a compatible action, endpoint coupling, and simulation-free flow matching objective. This design allows application specific dynamical assumptions to be incorporated through path specification and admits an analytical reduction to a static optimal transport problem. More general path and action designs can also be incorporated into the extended space. Figure 1 summarizes the overall construction.

Our main contributions are as follows:

- We introduce an extended space formulation for unbalanced flow matching that lifts unbalanced dynamics into conservative transport by introducing a log mass coordinate alongside the state space. This formulation enables state transport and relative growth to be modeled jointly, while conditional paths and endpoint couplings are derived consistently from a common extended-space transport formulation.
- We establish theoretical connections between the extended-space dynamic formulation and its static transport counterpart. For linear state motion and linear interpolation in log mass, corresponding to constant relative growth, we analytically eliminate the latent log mass coordinate and derive a reduced optimal transport problem with an explicit abundance change cost.
- We develop an entropically regularized Sinkhorn-like solver and a simulation-free flow matching objective for jointly learning single-cell state transport and population growth. Experiments demonstrate accurate prediction of state and abundance dynamics on synthetic and real single-cell snapshot datasets.

2 Related Work

Flow matching. Flow matching learns continuous dynamics by regressing vector fields associated with tractable conditional probability paths, avoiding ODE simulation during training (Lipman et al., 2023; Liu et al., 2023; Albergo & Vanden-Eijnden, 2023). A key design choice is the conditional path: many methods use simple linear interpolation, while others adapt or learn interpolants from data geometry, conditioning variables, or multiple observed snapshots (Liu et al., 2023; Kapusniak et al., 2024; Rohbeck et al., 2025; Lee et al., 2025; Kviman et al., 2026). Flow matching has also been applied directly to temporal single-cell snapshots using OT-based couplings and coupling refinement (Pu et al., 2026). Another key choice is the endpoint coupling, which determines how source and target samples are paired and may be independent, prescribed, or data dependent (Albergo et al., 2024). In general flow matching constructions, the coupling and conditional path can be specified independently and need not arise from a common geometric or variational principle.

Optimal transport and coupling–path consistency. Optimal transport links endpoint coupling and conditional paths through a common geometry. In quadratic OT, the static Kantorovich and dynamic Benamou–Brenier formulations induce the same squared Euclidean endpoint cost and constant-speed linear interpolation (Kantorovich, 1942; Benamou & Brenier, 2000). OT-CFM exploits this correspondence by sampling endpoint pairs from the OT coupling and regressing velocities along the associated displacement paths (Tong et al., 2024a). Related methods extend this construction through minibatch couplings, entropic regularization, Schrödinger bridges, or geometry-aware paths (Pooladian et al., 2023; Klein et al., 2024; Tong et al., 2024b; Kapusniak et al., 2024). Lagrangian Flow Matching further derives least-action paths and static transport costs from a prescribed Lagrangian (Du et al., 2026). These methods start from a prescribed geometry or action. ESFM instead starts from a prescribed conditional path together with a compatible action and derives the corresponding endpoint coupling.

Unbalanced transport and single-cell dynamics. Changes in cell abundance due to proliferation and death make unbalanced transport particularly relevant for single-cell snapshot dynamics. Unbalanced OT relaxes mass conservation, while WFR jointly models state transport and population growth (Benamou, 2003; Chizat et al., 2018b; Liero et al., 2016; Chizat et al., 2018a; Kondratyev et al., 2016). These ideas have motivated continuous single-cell models including TIGON, DeepRUOT, Var-RUOT, stVCR, and MIOFlow 2.0 (Sha et al., 2024; Zhang et al., 2025a; Sun et al., 2025; Peng et al., 2026c; Sun et al., 2026), as well as simulation-free approaches based on relaxed couplings, explicit growth fields, branching dynamics, or general growth penalties (Eyring et al., 2024; Wang et al., 2025; Maddu et al., 2026; Ying et al., 2026a;b). WFR-FM and its extensions derive conditional paths and couplings from WFR geometry, while SUDO generalizes this geometry-first view to convex growth penalties (Peng et al., 2026b;a; Wang et al., 2026; Ying et al., 2026b). ESFM instead adopts a path-first design, deriving a compatible action, endpoint coupling, and flow matching objective from a prescribed conditional path.

3 Mathematical Background

Conditional flow matching. Flow matching learns a time-dependent vector field that transports a source distribution to a target distribution. Conditional flow matching (CFM) avoids requiring the marginal probability path and its velocity field in closed form by introducing tractable conditional probability paths (Lipman et al., 2023; Tong et al., 2024a). For a condition z, typically an endpoint pair $z = (\boldsymbol{x}_0, \boldsymbol{x}_1)$, CFM minimizes

$$\mathcal{L}_{\mathrm{CFM}}(\theta) = \mathbb{E}_{t,z,\boldsymbol{x}\sim\rho_t(\cdot|z)}\left[\left\|\boldsymbol{u}_\theta(\boldsymbol{x},t) - \boldsymbol{u}_t(\boldsymbol{x} \mid z)\right\|^2\right]. \tag{3.1}$$

Thus, an endpoint-conditioned CFM construction involves two key components: the distribution of endpoint pairs and the conditional path associated with each pair. Additional background on marginal flow matching is provided in Appendix A.1.

OT and coupling–path consistency. Optimal transport provides a principled way to link these two components through a common transport geometry. For quadratic OT, the static Kantorovich

problem determines an optimal endpoint coupling, while the corresponding particle-level geodesic is the constant-speed interpolation

$$\boldsymbol{x}_t = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1, \qquad \boldsymbol{u}_t(\boldsymbol{x}_t \mid \boldsymbol{x}_0, \boldsymbol{x}_1) = \boldsymbol{x}_1 - \boldsymbol{x}_0.$$

The endpoint coupling and conditional path therefore arise from the same transport geometry. OT-CFM exploits this correspondence by sampling endpoint pairs from the OT coupling and regressing the velocity field along the associated displacement interpolation (Benamou & Brenier, 2000; Tong et al., 2024a). We refer to this shared geometric origin as *coupling–path consistency*.

Unbalanced dynamics and WFR geometry. To describe changes in population abundance, the state velocity is augmented by a growth field. Along a particle trajectory carrying physical mass $m_t > 0$, the growth rate is

$$g_t = \frac{\mathrm{d} \ln m_t}{\mathrm{d}t} = \frac{\dot{m}_t}{m_t}. \tag{3.2}$$

At the population level, a finite mass density ρ_t evolves according to the unbalanced continuity equation

$$\partial_t \rho_t(\boldsymbol{x}) + \nabla \cdot (\rho_t(\boldsymbol{x}) \boldsymbol{u}_t(\boldsymbol{x})) = \rho_t(\boldsymbol{x}) g_t(\boldsymbol{x}). \tag{3.3}$$

A representative geometry for such dynamics is Wasserstein–Fisher–Rao (WFR), whose dynamic formulation jointly penalizes state transport and growth (Liero et al., 2016; Chizat et al., 2018b):

$$\begin{aligned} \inf_{\rho, \boldsymbol{u}, g} \quad & \int_0^1 \int_{\mathbb{R}^d} \left(\|\boldsymbol{u}_t(\boldsymbol{x})\|^2 + \delta^2 \, |g_t(\boldsymbol{x})|^2 \right) \rho_t(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}t, \\ \text{s.t.} \quad & \partial_t \rho_t + \nabla \cdot (\rho_t \boldsymbol{u}_t) = \rho_t g_t, \qquad \rho_{t=0} = \rho_0, \quad \rho_{t=1} = \rho_1. \end{aligned}$$

As in balanced OT, the WFR geometry jointly determines endpoint pairing and particle-level interpolation, providing the geometric basis of WFR-FM (Peng et al., 2026b).

4 ESFM Framework

In this section, we formulate unbalanced population dynamics as conservative transport in the joint space of cell state and log mass. Using linear state and log mass paths with a compatible quadratic action, we derive the corresponding dynamic and static transport problems, a computable coupling, and a flow matching objective while preserving coupling–path consistency.

4.1 Extended Space

ESFM defines the dynamics in the extended space $\mathcal{W} = \mathbb{R}^d \times \mathbb{R}$, where each particle is described by its cell state $\boldsymbol{x}$ and log mass $y = \ln m$. The joint state $\boldsymbol{w} = (\boldsymbol{x}, y)$ evolves according to

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} \boldsymbol{x}_t \\ y_t \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{u}}_t(\boldsymbol{x}_t, y_t) \\ \hat{g}_t(\boldsymbol{x}_t, y_t) \end{pmatrix}.$$

Here $\hat{\boldsymbol{u}}_t$ and $\hat{g}_t$ denote the state velocity and relative growth field in the extended space. Let $\hat{\rho}_t$ denote the particle number density. Its weighted projection recovers the physical mass density:

$$\rho_t(\boldsymbol{x}) = \int_{\mathbb{R}} \mathrm{e}^y \hat{\rho}_t(\boldsymbol{x}, y) \, \mathrm{d}y.$$

We assume $\int_{\mathbb{R}^d} \rho_0(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = 1$. We assume that each source particle has unit mass at $t = 0$. The boundary conditions are

$$\hat{\rho}_0(\boldsymbol{x}, y) = \rho_0(\boldsymbol{x}) \delta_0(y), \qquad \int_{\mathbb{R}} \mathrm{e}^y \hat{\rho}_1(\boldsymbol{x}, y) \, \mathrm{d}y = \rho_1(\boldsymbol{x}). \tag{4.1}$$

Thus $\hat{\rho}_0$ and $\hat{\rho}_1$ are probability measures, whereas ρ_1 need not have unit mass.

Particle-number conservation is expressed by the extended continuity equation:

$$\partial_t \hat{\rho}_t + \nabla_{\boldsymbol{x}} \cdot (\hat{\rho}_t \hat{\boldsymbol{u}}_t) + \partial_y (\hat{\rho}_t \hat{g}_t) = 0. \tag{4.2}$$

Theorem 4.1. *Assume that the extended density and growth field satisfy*

$$\lim_{y\to\pm\infty} \mathrm{e}^{y}\hat{\rho}_t(\boldsymbol{x}, y)\hat{g}_t(\boldsymbol{x}, y) = 0.$$

After mass projection onto the physical state space, Equation (4.2) *implies Equation* (3.3) *with the corresponding projected velocity and growth fields.*

The projected fields and proof are given in Appendix A.2.

4.2 Prescribed Path and Compatible Action

For extended endpoints $\boldsymbol{w}_0 = (\boldsymbol{x}_0, y_0)$ and $\boldsymbol{w}_1 = (\boldsymbol{x}_1, y_1)$, we first prescribe linear interpolation in both state and log mass:

$$\boldsymbol{x}_t = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1, \qquad y_t = (1-t)y_0 + ty_1, \qquad m_t = m_0^{1-t}m_1^t. \tag{4.3}$$

This path corresponds to constant state velocity and constant relative growth between the prescribed endpoints.

We then construct a compatible microscopic action

$$\mathcal{A}_\delta(\boldsymbol{w}) = \int_0^1 \left(\|\dot{\boldsymbol{x}}_t\|^2 + \delta^2 \left|\dot{y}_t\right|^2\right) \mathrm{d}t, \tag{4.4}$$

where $\delta > 0$ controls the cost of changing log mass relative to state transport. Its minimum defines the endpoint cost

$$c(\boldsymbol{w}_0, \boldsymbol{w}_1) = \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|^2 + \delta^2 \left|y_1 - y_0\right|^2. \tag{4.5}$$

The optimality argument is given in Appendix A.3.

More general formulation and examples with different prescribed paths are discussed in Appendix D; however, such cases generally do not admit analytically reduced formulations.

4.3 Extended Space Dynamic and Static OT

Integrating the single-particle action over the population gives the dynamic transport problem, subject to Equations (4.1) and (4.2):

$$\mathcal{B}_\delta(\rho_0, \rho_1) = \inf_{\hat{\rho}, \hat{\boldsymbol{u}}, \hat{g}} \int_0^1 \int_{\mathcal{W}} \left(\|\hat{\boldsymbol{u}}_t(\boldsymbol{w})\|^2 + \delta^2 \left|\hat{g}_t(\boldsymbol{w})\right|^2\right) \hat{\rho}_t(\boldsymbol{w}) \,\mathrm{d}\boldsymbol{w} \,\mathrm{d}t. \tag{4.6}$$

The corresponding static formulation uses the endpoint cost in Equation (4.5):

$$\mathcal{K}_\delta(\rho_0, \rho_1) = \inf_{\substack{\hat{\rho}_1 \in \mathcal{P}(\mathcal{W}) \\ \hat{\pi} \in \Pi(\hat{\rho}_0, \hat{\rho}_1)}} \int_{\mathcal{W}^2} c(\boldsymbol{w}_0, \boldsymbol{w}_1) \,\mathrm{d}\hat{\pi}(\boldsymbol{w}_0, \boldsymbol{w}_1). \tag{4.7}$$

Here $\mathcal{P}(\mathcal{W})$ denotes the set of probability measures on $\mathcal{W}$, and $\Pi(\hat{\rho}_0, \hat{\rho}_1)$ denotes probability couplings with these marginals, subject to Equation (4.1). Unlike standard OT, ESFM optimizes $\hat{\rho}_1$ jointly with the coupling, while fixing only its weighted projection.

Theorem 4.2. *The extended dynamic and static problems have equal optimal values:* $\mathcal{B}_\delta(\rho_0, \rho_1) = \mathcal{K}_\delta(\rho_0, \rho_1)$.

The proof is provided in Appendix A.3.

4.4 Reduced ESFM Static OT

The source condition fixes $y_0 = 0$, so the extended plan can be written as $\pi(\boldsymbol{x}_0, \boldsymbol{x}_1, y_1)$. To bound individual particle growth, we impose $y_1 \le M$ with $M > 1$. The choice of M is described in Appendix B.1. Since $y_1 = \ln m_1$, negative values of y_1 remain allowed and correspond to mass decay. Define the two semi-couplings

$$\gamma_0(\boldsymbol{x}_0, \boldsymbol{x}_1) = \int \pi(\boldsymbol{x}_0, \boldsymbol{x}_1, y_1) \,\mathrm{d}y_1, \qquad \gamma_1(\boldsymbol{x}_0, \boldsymbol{x}_1) = \int \mathrm{e}^{y_1} \pi(\boldsymbol{x}_0, \boldsymbol{x}_1, y_1) \,\mathrm{d}y_1.$$

For each endpoint pair, γ_0 and γ_1 represent the initial and final mass, respectively. Their ratio $R(\boldsymbol{x}_0, \boldsymbol{x}_1) = \mathrm{d}\gamma_1 / \mathrm{d}\gamma_0(\boldsymbol{x}_0, \boldsymbol{x}_1)$ is the local growth ratio, with $0 < R(\boldsymbol{x}_0, \boldsymbol{x}_1) \le \mathrm{e}^M$.

Theorem 4.3. *For $M > 1$, the extended static problem in Equation* (4.7) *with the additional constraint $y_1 \le M$ is equivalent to*

$$\begin{aligned} \inf_{\gamma_0,\gamma_1\ge 0} \quad & \iint \left[\|\boldsymbol{x}_1-\boldsymbol{x}_0\|^2 + \delta^2 \Phi_M(R(\boldsymbol{x}_0,\boldsymbol{x}_1))\right] \mathrm{d}\gamma_0, \\ \textit{s.t.} \quad & \int \gamma_0(\boldsymbol{x}_0,\boldsymbol{x}_1)\,\mathrm{d}\boldsymbol{x}_1 = \rho_0(\boldsymbol{x}_0), \qquad \int \gamma_1(\boldsymbol{x}_0,\boldsymbol{x}_1)\,\mathrm{d}\boldsymbol{x}_0 = \rho_1(\boldsymbol{x}_1), \\ & 0 < R(\boldsymbol{x}_0,\boldsymbol{x}_1) \le \mathrm{e}^M. \end{aligned} \tag{4.8}$$

Here $\Phi_M(\cdot)$ is given by

$$\Phi_M(r) = \begin{cases} (\ln r)^2, & 0 < r \le \mathrm{e}^{y_\star}, \\ \dfrac{2y_\star}{\mathrm{e}^{y_\star}} r + y_\star^2 - 2y_\star, & \mathrm{e}^{y_\star} < r \le \mathrm{e}^M. \end{cases}$$

where $y_\star \in (0,1)$ is the threshold determined by M through the equaion $M^2 - y_\star^2 + 2y_\star - 2y_\star \mathrm{e}^{M-y_\star} = 0$.

The reduction and local optimizer are proved in Appendix A.4.

Proposition 4.1. *The entropic regularization of Equation* (4.8) *admits a Sinkhorn-like algorithm with explicit alternating row and column updates.*

The update formulas and proof are given in Appendix A.5. For training, sample state pairs from normalized γ_0, set $y_0 = 0$, and draw y_1 from the local optimizer in Equation (A.25).

Proposition 4.2. *Splitting a source particle into particles at the same state, while preserving its total weight, does not change the optimal value of Equation* (A.15).

The proof is given in Appendix A.6.

4.5 Conditional Paths and ESFM Training

We apply the balanced CFM construction of Lipman et al. (2023) in extended space, using separate networks $\boldsymbol{v}_\theta$ and g_ϕ to form $\boldsymbol{s}_{\theta,\phi} = (\boldsymbol{v}_\theta, g_\phi)$.

Marginal objective. For an extended density path $\hat{\rho}_t$ with velocity components $(\boldsymbol{u}_t, g_t)$, the ideal regression objective is

$$\mathcal{L}_{\mathrm{ESFM}}^{\mathrm{marg}}(\theta,\phi) = \int_0^1 \int_{\mathcal{W}} \Big[\, \|\boldsymbol{v}_\theta(\boldsymbol{w},t) - \boldsymbol{u}_t(\boldsymbol{w})\|^2 + \lambda\, |g_\phi(\boldsymbol{w},t) - g_t(\boldsymbol{w})|^2 \,\Big] \hat{\rho}_t(\boldsymbol{w})\,\mathrm{d}\boldsymbol{w}\,\mathrm{d}t. \tag{4.9}$$

Since the marginal fields are unknown, we instead regress tractable conditional fields.

Conditional path. Let $\boldsymbol{z} = (\boldsymbol{w}_0, \boldsymbol{w}_1) \in \mathcal{Z} = \mathcal{W}^2$ have the endpoint distribution q obtained in Section 4.4. We use the ESFM geodesic $\boldsymbol{\mu}_t(\boldsymbol{z}) = (1-t)\boldsymbol{w}_0 + t\boldsymbol{w}_1$ as the mean of a Gaussian path with fixed covariance:

$$\hat{\rho}_t(\boldsymbol{w} \mid \boldsymbol{z}) = \mathcal{N}\big(\boldsymbol{w} \mid \boldsymbol{\mu}_t(\boldsymbol{z}), \sigma^2 I_{d+1}\big)\,.$$

As $\sigma \to 0$, the Gaussian path recovers the deterministic ESFM path $\boldsymbol{w}_t = \boldsymbol{\mu}_t(\boldsymbol{z})$. Its conditional velocity is constant:

$$\boldsymbol{s}_t(\boldsymbol{w} \mid \boldsymbol{z}) = \big(\boldsymbol{u}_t(\boldsymbol{w} \mid \boldsymbol{z}), g_t(\boldsymbol{w} \mid \boldsymbol{z})\big) = \boldsymbol{w}_1 - \boldsymbol{w}_0.$$

Mixing the conditional densities over q gives $\hat{\rho}_t$; averaging their velocities conditional on the current state gives $(\boldsymbol{u}_t, g_t)$. The explicit formulas are provided in Appendix A.7.

Conditional objective. The computable training loss is

$$\begin{aligned} \mathcal{L}_{\mathrm{ESFM}}^{\mathrm{cond}}(\theta,\phi) = \int_0^1 \int_{\mathcal{Z}} \int_{\mathcal{W}} \Big[& \|\boldsymbol{v}_\theta(\boldsymbol{w},t) - \boldsymbol{u}_t(\boldsymbol{w} \mid \boldsymbol{z})\|^2 \\ & + \lambda\, |g_\phi(\boldsymbol{w},t) - g_t(\boldsymbol{w} \mid \boldsymbol{z})|^2 \Big] \hat{\rho}_t(\boldsymbol{w} \mid \boldsymbol{z}) q(\boldsymbol{z})\,\mathrm{d}\boldsymbol{w}\,\mathrm{d}\boldsymbol{z}\,\mathrm{d}t. \end{aligned} \tag{4.10}$$

Here $\lambda > 0$ balances the state and log mass regression errors. The following result justifies this conditional objective.

Theorem 4.4. *Assume that $\hat{\rho}_t(\boldsymbol{w}) > 0$ for all $\boldsymbol{w} \in \mathcal{W}$ and $t \in [0,1]$, and that $q(\boldsymbol{z})$ is independent of $\boldsymbol{w}$ and t. Then*

$$\nabla_{\theta,\phi}\mathcal{L}^{\mathrm{cond}}_{\mathrm{ESFM}}(\theta,\phi) = \nabla_{\theta,\phi}\mathcal{L}^{\mathrm{marg}}_{\mathrm{ESFM}}(\theta,\phi).$$

The proof is provided in Appendix A.7.

Algorithm 1 ESFM conditional flow matching training

Input: endpoint law $q(\boldsymbol{z})$ with $\boldsymbol{z} = (\boldsymbol{w}_0, \boldsymbol{w}_1)$; λ, σ; batch size B, training steps K, and learning rates η_v, η_g.
for $k = 1, \dots, K$ **do**
  Sample $\{\boldsymbol{z}_b = (\boldsymbol{w}_{0,b}, \boldsymbol{w}_{1,b})\}_{b=1}^{B}$ independently from $q(\boldsymbol{z})$.
  for $b = 1, \dots, B$ **do**
    Sample $t_b \sim \mathcal{U}[0,1]$ and $\boldsymbol{\epsilon}_b \sim \mathcal{N}(0, I_{d+1})$.
    $\boldsymbol{w}_{t,b} \leftarrow (1-t_b)\boldsymbol{w}_{0,b} + t_b\boldsymbol{w}_{1,b} + \sigma\boldsymbol{\epsilon}_b$.
    $\boldsymbol{s}_b^\star \leftarrow \boldsymbol{w}_{1,b} - \boldsymbol{w}_{0,b} = (\boldsymbol{u}_b^\star, g_b^\star)$.
  end for
  $\widehat{\mathcal{L}}_{\boldsymbol{u}} \leftarrow B^{-1}\sum_{b=1}^{B} \|\boldsymbol{v}_\theta(\boldsymbol{w}_{t,b}, t_b) - \boldsymbol{u}_b^\star\|^2$.
  $\widehat{\mathcal{L}}_g \leftarrow B^{-1}\sum_{b=1}^{B} |g_\phi(\boldsymbol{w}_{t,b}, t_b) - g_b^\star|^2$.
  $\theta \leftarrow \theta - \eta_v \nabla_\theta \widehat{\mathcal{L}}_{\boldsymbol{u}}$.
  $\phi \leftarrow \phi - \eta_g \lambda \nabla_\phi \widehat{\mathcal{L}}_g$.
end for
Output: the trained vector field in extended space $\boldsymbol{s}_{\theta,\phi} = (\boldsymbol{v}_\theta, g_\phi)$.

Detailed comparisons with VGFM, WFR-FM, and USB are provided in Appendix C. The inference procedure is given in Appendix F.

5 Numerical Results

We evaluate ESFM through four questions: **Q1**: Can a prescribed path consistent with known growth prior improve recovery of state-dependent growth dynamics? **Q2**: Does the numerical ESFM construction recover the expected path–coupling consistency? **Q3**: Does the prescribed path design preserve accurate state and mass prediction? **Q4**: Does ESFM generalize to unseen snapshots in real single-cell data? Implementation settings and metric definitions are provided in Appendices B.1 and B.2.

Recovery of state-dependent growth dynamics (Q1). We first test whether a prescribed growth path consistent with known qualitative growth behavior improves recovery of the underlying growth field. We use a controlled variant of the Simulation Gene benchmark with a known state-dependent growth field, containing a proliferative branch with growth near $g_{\max}$ and an approximately quiescent branch with near-zero growth. ESFM and WFR-FM are trained on the same population snapshots and evaluated against the ground-truth growth field.

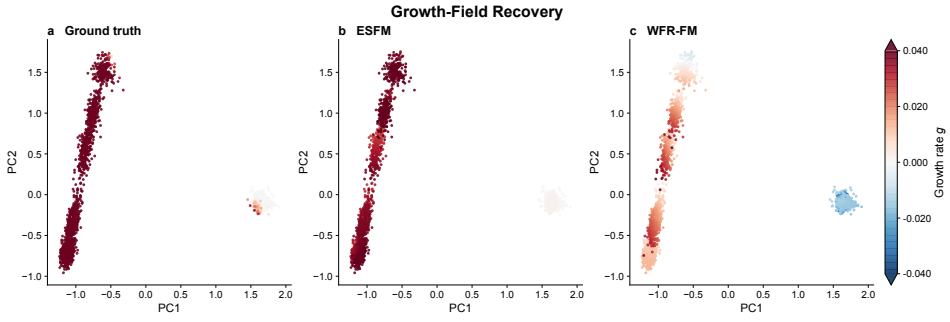


Figure 2: Growth field recovery on the saturating growth Simulation Gene benchmark. (a) Ground-truth growth rate. (b) ESFM prediction. (c) WFR-FM prediction.

As shown in Figure 2, ESFM recovers both the near-zero growth of the quiescent population and the saturated growth of the proliferative branch. Using the coefficient of determination R^2, ESFM achieves $R^2 = 0.922$, whereas WFR-FM underestimates the growth magnitude and yields $R^2 = -0.931$. Thus, in this controlled setting, ESFM more faithfully recovers the underlying state-dependent growth dynamics.

Numerical verification of path–coupling consistency (Q2). We compare the static coupling cost A_{stat} with the dynamic action A_{dyn} of ESFM's learned dynamics. The static value is obtained by solving the entropy-regularized static OT problem and evaluating the coupling's cost. The dynamic value is computed by numerical integration along ODE trajectories generated by the learned velocity and growth fields. Across the evaluated datasets, Table 1 shows close agreement between the two values, indicating that the learned dynamics approximately recover the corresponding static coupling costs. These results provide numerical support for ESFM's path–coupling consistency. Computational details are provided in Appendix B.6.

Table 1: Static coupling cost and learned dynamic action.

Dataset	A_{stat}	A_{dyn}	Gap (%)
Gene	2.405	2.340	2.677
Dyngen	29.628	30.187	1.885
Gaussian	13.343	13.159	1.374
EMT	1.304	1.280	1.819
EB	886.117	884.563	0.175
CITE	3251.004	3228.955	0.678
Mouse	310.677	304.636	1.944

Predictive accuracy under prescribed path design (Q3). We next test whether imposing the prescribed path compromises predictive accuracy. ESFM is evaluated on Simulation Gene, Dyngen, and a 1000D Gaussian mixture using the 1-Wasserstein distance $\mathcal{W}_1$ for distributional accuracy and relative mass error (RME) for mass prediction. Additional results are provided in Appendix B.3.

As shown in Table 2, ESFM achieves the best $\mathcal{W}_1$ and RME on Gene and Dyngen, and the lowest RME on Gaussian while remaining competitive in distributional accuracy. These results show that the prescribed path design preserves accurate endpoint prediction while providing reliable mass estimation.

Table 2: Synthetic benchmark performance measured by mean $\mathcal{W}_1$ and RME. RME is reported only for unbalanced methods.

Method	Gene (2D)		Dyngen (5D)		Gaussian (1000D)	
	$\mathcal{W}_1$ (↓)	RME (↓)	$\mathcal{W}_1$ (↓)	RME (↓)	$\mathcal{W}_1$ (↓)	RME (↓)
MMFM (Rohbeck et al., 2025)	0.298	—	1.371	—	2.833	—
Metric FM (Kapusniak et al., 2024)	0.311	—	1.767	—	3.794	—
SF2M (Tong et al., 2024b)	$0.224_{\pm 0.007}$	—	$1.277_{\pm 0.017}$	—	$3.543_{\pm 0.002}$	—
MIOFlow (Huguet et al., 2022)	0.148	—	0.965	—	2.858	—
DeepRUOT (Zhang et al., 2025a)	$0.043_{\pm 0.002}$	$0.017_{\pm 0.001}$	$0.623_{\pm 0.032}$	$0.065_{\pm 0.011}$	$3.785_{\pm 0.009}$	$0.303_{\pm 0.070}$
Var-RUOT (Sun et al., 2025)	$0.079_{\pm 0.003}$	$0.008_{\pm 0.002}$	$0.522_{\pm 0.008}$	$0.177_{\pm 0.007}$	$2.813_{\pm 0.004}$	$0.041_{\pm 0.006}$
UOT-FM (Eyring et al., 2024)	0.093	0.010	1.204	0.097	2.771	0.033
VGFM (Wang et al., 2025)	0.046	0.006	0.598	0.037	3.010	0.037
WFR-FM (Peng et al., 2026b)	<u>0.019</u>	<u>0.001</u>	0.135	<u>0.005</u>	<u>2.233</u>	0.044
USB (Ying et al., 2026a)	$\underline{0.019}_{\pm 0.000}$	$0.002_{\pm 0.000}$	$\underline{0.131}_{\pm 0.001}$	$0.007_{\pm 0.000}$	$\mathbf{2.136}_{\pm 0.002}$	$\underline{0.004}_{\pm 0.004}$
ESFM (ours)	**0.017**	**0.0005**	**0.111**	**0.003**	2.428	**0.0001**

Figure 3 visualizes the learned dynamics on Gene and Dyngen. The predicted endpoint distributions closely match the observed targets, while the propagated source trajectories illustrate the learned state evolution.

Generalization to held-out single-cell snapshots (Q4). Finally, we evaluate ESFM on unseen snapshots using hold-one-out interpolation on EMT, embryoid body (EB), CITE-seq, and mouse hematopoiesis. We report the mean 1-Wasserstein distance $\mathcal{W}_1$ over held-out time points. Additional results are provided in Appendices B.4 and B.5.

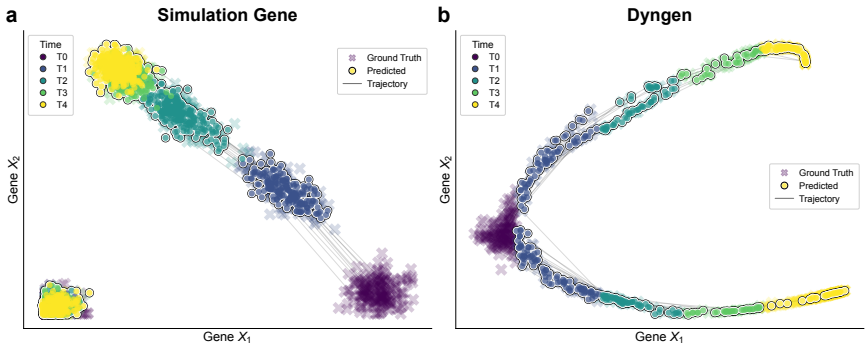


Figure 3: Qualitative trajectory reconstruction by ESFM on synthetic benchmarks. Crosses denote observed target samples, circles denote ESFM predictions, and gray curves denote propagated source trajectories.

Table 3: Hold-one-out interpolation performance on real single-cell datasets.

Method	EMT (10D)	EB (50D)	CITE (50D)	Mouse (50D)
MMFM	0.323	11.213	38.521	8.263
Metric FM	0.314	10.726	37.342	7.753
SF2M	0.308$\pm$0.001	10.986$\pm$0.006	38.333$\pm$0.002	8.646$\pm$0.004
MIOFlow	0.325	10.960	39.574	7.779
TIGON	0.360	11.080	38.159	6.868
DeepRUOT	0.323$\pm$0.002	**10.075**$\pm$0.004	37.892$\pm$0.002	6.847$\pm$0.003
Var-RUOT	0.320$\pm$0.003	11.035$\pm$0.017	38.393$\pm$0.029	8.672$\pm$0.040
UOT-FM	0.322	11.344	38.649	9.332
VGFM	0.301	10.370	37.386	8.496
WFR-FM	0.298	10.157	37.221	**6.586**
USB	0.298$\pm$0.0001	10.177$\pm$0.0001	**37.087**$\pm$0.0001	6.988$\pm$0.0001
ESFM (ours)	**0.295**	10.134	37.147	6.714

As shown in Table 3, ESFM achieves the lowest reported mean on EMT and the second-lowest on EB, CITE-seq, and Mouse. These results show that ESFM maintains competitive distributional prediction and generalizes well to unseen snapshots across diverse real single-cell systems.

6 Conclusion, Limitations, and Discussion

We introduced ESFM, an extended space flow matching framework for unbalanced dynamics from population snapshots. ESFM lifts nonconservative dynamics to conservative transport in a joint space of state and log mass, where state transport and relative growth are modeled jointly. For the linear state and log mass paths studied here, we use a compatible quadratic action and derive the corresponding static transport problem. The terminal log mass coordinate can be eliminated analytically, yielding a tractable reduced coupling problem and a simulation-free flow matching objective. Experiments demonstrate improved recovery of state-dependent growth dynamics while maintaining competitive predictive performance.

Several limitations remain. The prescribed path is an inductive prior rather than a uniquely identifiable biological trajectory, and solving the static transport problem can be costly at scale. Minibatch and entropic approximations improve efficiency but require further study. Future work could explore more biologically informed path and action designs within the extended space framework.

AI use statement

In this work, we used generative AI tools only for parameter suggestions and language editing. We did not use generative AI tools for research ideation, mathematical derivations, literature search, experimental design, code implementation or debugging, data analysis, result interpretation, or manuscript drafting. All AI-assisted suggestions and edits were carefully reviewed and verified by the authors. The authors made all methodological and scientific decisions and take full responsibility for the final content of this work, including all text, claims, and results.

Reproducibility statement

Mathematical assumptions and derivations are provided in Appendix A, including the dynamic–static correspondence, local mass reduction, and entropic solver updates. Network architectures, experimental settings, evaluation metrics, dataset descriptions, and additional results are reported in Appendix B. The inference procedure and pseudocode are given in Appendix F. These materials specify the mathematical construction and numerical procedures used to evaluate ESFM.

References

Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=li7qeBbCR1t.

Michael Samuel Albergo, Mark Goldstein, Nicholas Matthew Boffi, Rajesh Ranganath, and Eric Vanden-Eijnden. Stochastic interpolants with data-dependent couplings. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 921–937, 2024. URL https://proceedings.mlr.press/v235/albergo24a.html.

Aymeric Baradat and Hugo Lavenant. *Regularized Unbalanced Optimal Transport as Entropy Minimization with Respect to Branching Brownian Motion*, volume 458 of *Astérisque*. Société Mathématique de France, 2025. ISBN 978-2-37905-211-8. doi: 10.24033/ast.1247.

Jean-David Benamou. Numerical resolution of an unbalanced mass transport problem. *ESAIM: Mathematical Modelling and Numerical Analysis*, 37(5):851–868, 2003. doi: 10.1051/m2an:2003058.

Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000. doi: 10.1007/s002110050002.

Charlotte Bunne, Geoffrey Schiebinger, Andreas Krause, Aviv Regev, and Marco Cuturi. Optimal transport for single-cell and spatial omics. *Nature Reviews Methods Primers*, 4(1):58, 2024. doi: 10.1038/s43586-024-00334-2.

Robrecht Cannoodt, Wouter Saelens, Louise Deconinck, and Yvan Saeys. Spearheading future omics analyses using dyngen, a multi-modal simulator of single cells. *Nature Communications*, 12(1):3942, 2021. doi: 10.1038/s41467-021-24152-2.

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 31, pp. 6571–6583, 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/hash/69386f6bb1dfed68692a24c8686939b9-Abstract.html.

Lénaïc 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. doi: 10.1007/s10208-016-9331-y.

Lénaïc 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. doi: 10.1016/j.jfa.2018.03.008.

David P. Cook and Barbara C. Vanderhyden. Context specificity of the EMT transcriptional response. *Nature Communications*, 11(1):2142, 2020. doi: 10.1038/s41467-020-16066-2.

Shukai Du, Junzhe Zhang, and Yiming Li. Lagrangian flow matching: A least-action framework for principled path design. *arXiv preprint arXiv:2605.15419*, 2026.

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. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/022a39052abf9ca467e268923057dfc0-Abstract-Conference.html.

Guillaume Huguet, Daniel Sumner Magruder, Alexander Tong, Oluwadamilola Fasina, Manik Kuchroo, Guy Wolf, and Smita Krishnaswamy. Manifold interpolating optimal-transport flows for trajectory inference. In *Advances in Neural Information Processing Systems*, volume 35, pp. 29705–29718, 2022. doi: 10.52202/068431-2154.

Leonid V. Kantorovich. On the translocation of masses. *Doklady Akademii Nauk SSSR*, 37:199–201, 1942.

Kacper Kapusniak, Peter Potaptchik, Teodora Reu, Leo Zhang, Alexander Tong, Michael Bronstein, Joey Bose, and Francesco Di Giovanni. Metric flow matching for smooth interpolations on the data manifold. In *Advances in Neural Information Processing Systems*, volume 37, pp. 135011–135042, 2024. doi: 10.52202/079017-4291.

Dominik Klein, Théo Uscidda, Fabian Theis, and Marco Cuturi. GENOT: Entropic (Gromov) Wasserstein flow matching with applications to single-cell genomics. In *Advances in Neural Information Processing Systems*, volume 37, pp. 103897–103944, 2024. doi: 10.52202/079017-3301.

Dominik Klein, Jonas Simon Fleck, Daniil Bobrovskiy, Lea Zimmermann, Sören Becker, Alessandro Palma, Leander Dony, Alejandro Tejada-Lapuerta, Guillaume Huguet, Hsiu-Chuan Lin, Nadezhda Azbukina, Fátima Sanchís-Calleja, Théo Uscidda, Artur Szalata, Manuel Gander, Aviv Regev, Barbara Treutlein, J. Gray Camp, and Fabian J. Theis. CellFlow enables generative single-cell phenotype modeling with flow matching. *bioRxiv*, 2025. doi: 10.1101/2025.04.11.648220. URL https://www.biorxiv.org/content/10.1101/2025.04.11.648220.

Stanislav Kondratyev, Léonard Monsaingeon, and Dmitry Vorotnikov. A new optimal transport distance on the space of finite Radon measures. *Advances in Differential Equations*, 21(11/12): 1117–1164, 2016. doi: 10.57262/ade/1476369298.

Oskar Kviman, Kirill Tamogashev, Nicola Branchini, Víctor Elvira, Jens Lagergren, and Nikolay Malkin. Multi-marginal flow matching with adversarially learnt interpolants. In *The Fourteenth International Conference on Learning Representations*, 2026.

Christopher Lance, Malte D. Luecken, Daniel B. Burkhardt, Robrecht Cannoodt, Pia Rautenstrauch, Anna Laddach, Aidyn Ubingazhibov, Zhi-Jie Cao, Kaiwen Deng, Sumeer Khan, Qiao Liu, Nikol Russkikh, Gleb Ryazantsev, Uwe Ohler, NeurIPS 2021 Multimodal Data Integration Competition Participants, Angela Oliveira Pisco, Jonathan Bloom, Smita Krishnaswamy, and Fabian J. Theis. Multimodal single-cell data integration challenge: Results and lessons learned. In *Proceedings of the NeurIPS 2021 Competitions and Demonstrations Track*, volume 176 of *Proceedings of Machine Learning Research*, pp. 162–176, 2022. URL https://proceedings.mlr.press/v176/lance22a.html.

Justin Lee, Behnaz Moradijamei, and Heman Shakeri. Multi-marginal stochastic flow matching for high-dimensional snapshot data at irregular time points. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 33476–33498, 2025. URL https://proceedings.mlr.press/v267/lee25x.html.

Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal transport in competition with reaction: The Hellinger–Kantorovich distance and geodesic curves. *SIAM Journal on Mathematical Analysis*, 48(4):2869–2911, 2016. doi: 10.1137/15M1041420.

Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures. *Inventiones Mathematicae*, 211 (3):969–1117, 2018. doi: 10.1007/s00222-017-0759-8.

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.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=XVjTT1nw5z.

Suryanarayana Maddu, Victor Chardès, and Michael J. Shelley. Learning biophysical models of gene regulation with probability flow matching. *arXiv preprint arXiv:2604.25062*, 2026. URL https://arxiv.org/abs/2604.25062.

Kevin R. Moon, David van Dijk, Zheng Wang, Scott Gigante, Daniel B. Burkhardt, William S. Chen, Kristina Yim, Antonia van den Elzen, Matthew J. Hirn, Ronald R. Coifman, Natalia B. Ivanova, Guy Wolf, and Smita Krishnaswamy. Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology*, 37(12):1482–1492, 2019. doi: 10.1038/s41587-019-0336-3.

Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in PyTorch. In *NIPS 2017 Workshop on Autodiff*, 2017. URL https://openreview.net/forum?id=BJJsrmfCZ.

Qiangwei Peng, Lezhi Chen, and Peijie Zhou. Multiscale supervised unbalanced optimal transport flow matching. *arXiv preprint arXiv:2605.16529*, 2026a. URL https://arxiv.org/abs/2605.16529.

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.

Qiangwei Peng, Peijie Zhou, and Tiejun Li. stVCR: Spatiotemporal dynamics of single cells. *Nature Methods*, 23(3):542–553, 2026c. doi: 10.1038/s41592-026-03010-3.

Aram-Alexandre Pooladian, Heli Ben-Hamu, Carles Domingo-Enrich, Brandon Amos, Yaron Lipman, and Ricky T. Q. Chen. Multisample flow matching: Straightening flows with minibatch couplings. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 28100–28127, 2023. URL https://proceedings.mlr.press/v202/pooladian23a.html.

Siyu Pu, Qingqing Long, Xiaohan Huang, Haotian Chen, Jiajia Wang, Meng Xiao, Xiao Luo, Hengshu Zhu, Yuanchun Zhou, and Xuezhi Wang. From snapshots to trajectories: Learning single-cell gene expression dynamics via conditional flow matching. *arXiv preprint arXiv:2605.22340*, 2026.

Martin Rohbeck, Edward De Brouwer, Charlotte Bunne, Jan-Christian Huetter, Anne Biton, Kelvin Y. Chen, Aviv Regev, and Romain Lopez. Modeling complex system dynamics with flow matching across time and conditions. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/206b764b00195b8c896c5af8bde80726-Abstract-Conference.html.

Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, Lia Lee, Jenny Chen, Justin Brumbaugh, Philippe Rigollet, Konrad Hochedlinger, Rudolf Jaenisch, Aviv Regev, and Eric S. Lander. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943.e22, 2019. doi: 10.1016/j.cell.2019.01.006.

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. doi: 10.1038/s42256-023-00763-w.

Xingzhi Sun, João Felipe Rocha, Brett Phelan, Dhananjay Bhaskar, Guillaume Huguet, Yanlei Zhang, Daniel Sumner Magruder, Alexander Tong, Ke Xu, Oluwadamilola Fasina, Mark Gerstein, Guy Wolf, Natalia Ivanova, Christine L. Chaffer, and Smita Krishnaswamy. MIOFlow 2.0: A unified framework for inferring cellular stochastic dynamics from single cell and spatial transcriptomics data. *arXiv preprint arXiv:2603.22564*, 2026.

Yuhao Sun, Zhenyi Zhang, Zihan Wang, Tiejun Li, and Peijie Zhou. Variational regularized unbalanced optimal transport: Single network, least action. In *Advances in Neural Information Processing Systems*, volume 38, pp. 19150–19192, 2025. doi: 10.52202/085713-0572.

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, 2020. URL https://proceedings.mlr.press/v119/tong20a.html.

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*, 2024a. URL https://openreview.net/forum?id=CD9Snc73AW.

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, 2024b. URL https://proceedings.mlr.press/v238/tong24a.html.

Dongyi Wang, Yuanwei Jiang, Zhenyi Zhang, Xiang Gu, Peijie Zhou, and Jian Sun. Joint velocity-growth flow matching for single-cell dynamics modeling. In *Advances in Neural Information Processing Systems*, volume 38, pp. 178180–178216, 2025. doi: 10.52202/085713-5363.

Xinyu Wang, Ruoyu Wang, Qiangwei Peng, Peijie Zhou, and Tiejun Li. WFR-MFM: One-step inference for dynamic unbalanced optimal transport. In *Forty-Third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=yVGbbB3ERW.

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

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*, 2026a. URL https://openreview.net/forum?id=IABjbJlinz.

Junda Ying, Yuxuan Wang, Bowen Yang, Peijie Zhou, and Lei Zhang. Simulation-free unbalanced dynamic optimal transport with general growth penalty. *arXiv preprint arXiv:2609.04710*, 2026b.

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://proceedings.iclr.cc/paper_files/paper/2025/hash/32b8a612105de5c22db337b774ce7b61-Abstract-Conference.html.

Zhenyi Zhang, Yuhao Sun, Qiangwei Peng, Tiejun Li, and Peijie Zhou. Integrating dynamical systems modeling with spatiotemporal scRNA-Seq data analysis. *Entropy*, 27(5):453, 2025b. doi: 10.3390/e27050453.

Grace X. Y. Zheng, Jessica M. Terry, Phillip Belgrader, Paul Ryvkin, Zachary W. Bent, Ryan Wilson, Solongo B. Ziraldo, Tobias D. Wheeler, Geoff P. McDermott, Junjie Zhu, et al. Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8(1):14049, 2017. doi: 10.1038/ncomms14049.

A Mathematical Derivations and Proofs

A.1 Additional Background on Flow Matching and Dynamic OT

This section provides additional background, including marginal flow matching and the dynamic–static correspondence underlying OT-based conditional flow matching.

Flow matching. Consider normalized endpoint distributions $\rho_0, \rho_1 \in \mathcal{P}(\mathbb{R}^d)$. Flow matching learns a time-dependent vector field whose ODE transports samples from ρ_0 to ρ_1:

$$\frac{\mathrm{d}\boldsymbol{x}_t}{\mathrm{d}t} = \boldsymbol{u}_t(\boldsymbol{x}_t). \tag{A.1}$$

Given a probability path $\{\rho_t\}_{t\in[0,1]}$ and its velocity field $\boldsymbol{u}_t$, the marginal flow-matching objective is

$$\mathcal{L}_{\mathrm{FM}}(\theta) = \mathbb{E}_{t,\boldsymbol{x}\sim\rho_t}\left[\|\boldsymbol{u}_\theta(\boldsymbol{x},t) - \boldsymbol{u}_t(\boldsymbol{x})\|^2\right].$$

The learned vector field is used for inference by integrating Equation (A.1). In practice, the marginal path ρ_t and marginal velocity $\boldsymbol{u}_t$ are generally unavailable in closed form, which motivates the conditional formulation used in Equation (3.1).

Dynamic optimal transport. In the balanced setting, a density ρ_t and velocity field $\boldsymbol{u}_t$ satisfy the continuity equation

$$\partial_t \rho_t + \nabla \cdot (\rho_t \boldsymbol{u}_t) = 0.$$

The Benamou–Brenier formulation selects the interpolation of minimum kinetic energy (Benamou & Brenier, 2000):

$$\begin{aligned} W_2^2(\rho_0,\rho_1) = \inf_{\rho,\boldsymbol{u}} \quad & \int_0^1 \int_{\mathbb{R}^d} \|\boldsymbol{u}_t(\boldsymbol{x})\|^2 \rho_t(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}t, \\ \text{s.t.} \quad & \partial_t \rho_t + \nabla\cdot(\rho_t \boldsymbol{u}_t) = 0, \\ & \rho_{t=0} = \rho_0, \qquad \rho_{t=1} = \rho_1. \end{aligned}$$

For a pair of Dirac measures at $\boldsymbol{x}_0$ and $\boldsymbol{x}_1$, the minimizing particle path is

$$\boldsymbol{x}_t = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1, \qquad \boldsymbol{u}_t = \boldsymbol{x}_1 - \boldsymbol{x}_0. \tag{A.2}$$

Thus, the dynamic OT geometry determines the conditional path and its velocity at the particle level.

Static OT and endpoint coupling. The Benamou–Brenier dynamic problem is equivalent to the quadratic Kantorovich problem

$$\pi^\star \in \arg\min_{\pi\in\Pi(\rho_0,\rho_1)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|^2 \,\mathrm{d}\pi(\boldsymbol{x}_0,\boldsymbol{x}_1).$$

The static formulation determines which source and target states are paired, whereas the corresponding dynamic formulation determines how each pair moves between its endpoints. Aggregating the Dirac-to-Dirac geodesics in Equation (A.2) over the optimal coupling $\pi^\star$ recovers the balanced displacement interpolation. This dynamic–static correspondence is the geometric basis of OT-CFM (Tong et al., 2024a).

Extension to unbalanced dynamics. For mass-changing systems, the conservative continuity equation is replaced by

$$\partial_t \rho_t + \nabla \cdot (\rho_t \boldsymbol{u}_t) = \rho_t g_t,$$

where g_t is the growth rate defined in Equation (3.2). Dynamic unbalanced OT formulations, including WFR, augment the transport action with a growth penalty and admit corresponding static formulations based on relaxed couplings or semi-couplings (Chizat et al., 2018a; Liero et al., 2018).

The same construction principle can therefore be used in unbalanced flow matching: a dynamic transport geometry determines the single-particle state and mass paths, while its static formulation determines the endpoint pairing. Examples include relaxed-coupling approaches such as UOT-FM and CellFlow (Eyring et al., 2024; Klein et al., 2025), semi-relaxed growth estimation as in VGFM (Wang et al., 2025), and WFR-geodesic constructions such as WFR-FM and MUST-FM (Peng et al., 2026b;a). Related methods model stochastic branching or mechanistic gene-regulatory dynamics, including USB and PFM (Ying et al., 2026a; Maddu et al., 2026).

A.2 Proof of Theorem 4.1

Let $y = \ln m$ and let $\hat{\rho}_t(\boldsymbol{x}, y)$ be a particle number density. The physical density is

$$\rho_t(\boldsymbol{x}) = \int_{\mathbb{R}} \mathrm{e}^y \hat{\rho}_t(\boldsymbol{x}, y)\, \mathrm{d}y.$$

If the extended velocity is $(\hat{\boldsymbol{u}}_t, \hat{g}_t)$, then the physical velocity and growth rate are weighted averages:

$$\boldsymbol{u}_t(\boldsymbol{x}) = \frac{\int_{\mathbb{R}} \hat{\boldsymbol{u}}_t(\boldsymbol{x}, y)\hat{\rho}_t(\boldsymbol{x}, y)\mathrm{e}^y\, \mathrm{d}y}{\rho_t(\boldsymbol{x})}, \quad g_t(\boldsymbol{x}) = \frac{\int_{\mathbb{R}} \hat{g}_t(\boldsymbol{x}, y)\hat{\rho}_t(\boldsymbol{x}, y)\mathrm{e}^y\, \mathrm{d}y}{\rho_t(\boldsymbol{x})}.$$

The conservative extended equation is

$$\partial_t \hat{\rho}_t + \nabla_{\boldsymbol{x}} \cdot (\hat{\rho}_t \hat{\boldsymbol{u}}_t) + \partial_y(\hat{\rho}_t \hat{g}_t) = 0. \tag{A.3}$$

We now derive the physical nonconservative continuity equation in Equation (3.3). Multiplying Equation (A.3) by e^y and integrating over y gives

$$\int_{\mathbb{R}} \mathrm{e}^y \partial_t \hat{\rho}_t\, \mathrm{d}y + \int_{\mathbb{R}} \mathrm{e}^y \nabla_{\boldsymbol{x}} \cdot (\hat{\rho}_t \hat{\boldsymbol{u}}_t)\, \mathrm{d}y + \int_{\mathbb{R}} \mathrm{e}^y \partial_y(\hat{\rho}_t \hat{g}_t)\, \mathrm{d}y = 0. \tag{A.4}$$

The first two terms are the projected time derivative and flux:

$$\int_{\mathbb{R}} \mathrm{e}^y \partial_t \hat{\rho}_t(\boldsymbol{x}, y)\, \mathrm{d}y = \partial_t \rho_t(\boldsymbol{x}), \tag{A.5}$$

$$\int_{\mathbb{R}} \mathrm{e}^y \nabla_{\boldsymbol{x}} \cdot (\hat{\rho}_t(\boldsymbol{x}, y)\hat{\boldsymbol{u}}_t(\boldsymbol{x}, y))\ \mathrm{d}y = \nabla_{\boldsymbol{x}} \cdot (\rho_t(\boldsymbol{x})\boldsymbol{u}_t(\boldsymbol{x}))\,, \tag{A.6}$$

where the second identity follows from the definition of the weighted physical velocity. For the y-transport term, integration by parts gives

$$\begin{aligned} \int_{\mathbb{R}} \mathrm{e}^y \partial_y\, (\hat{\rho}_t(\boldsymbol{x}, y)\hat{g}_t(\boldsymbol{x}, y))\ \mathrm{d}y &= [\mathrm{e}^y \hat{\rho}_t(\boldsymbol{x}, y)\hat{g}_t(\boldsymbol{x}, y)]_{-\infty}^{+\infty} - \int_{\mathbb{R}} \mathrm{e}^y \hat{\rho}_t(\boldsymbol{x}, y)\hat{g}_t(\boldsymbol{x}, y)\, \mathrm{d}y \\ &= -\rho_t(\boldsymbol{x}) g_t(\boldsymbol{x}), \end{aligned} \tag{A.7}$$

where the boundary term vanishes by the assumption in Theorem 4.1. Since $y = \ln m$, the two limits correspond to particle masses approaching zero and infinity. Substituting Equations (A.5) to (A.7) into Equation (A.4) yields

$$\partial_t \rho_t + \nabla_{\boldsymbol{x}} \cdot (\rho_t \boldsymbol{u}_t) = \rho_t g_t,$$

which is the physical unbalanced continuity equation in Equation (3.3).

A.3 Proof of Theorem 4.2

Following the dynamic–static equivalence arguments of Benamou & Brenier (2000) and Chizat et al. (2018b), we present the proof from a particle-trajectory perspective in the extended space

$$\mathcal{W} = \mathbb{R}^d \times \mathbb{R}, \qquad \|(\boldsymbol{a}, b)\|_\delta^2 := \|\boldsymbol{a}\|^2 + \delta^2 |b|^2 .$$

Single-particle optimality.

Indeed, for any absolutely continuous path connecting $\boldsymbol{w}_0$ and $\boldsymbol{w}_1$, the Cauchy–Schwarz inequality gives

$$\int_0^1 \|\dot{\boldsymbol{x}}_t\|^2\, \mathrm{d}t \geq \left\| \int_0^1 \dot{\boldsymbol{x}}_t\, \mathrm{d}t \right\|^2 = \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|^2 , \qquad \int_0^1 |\dot{y}_t|^2\, \mathrm{d}t \geq |y_1 - y_0|^2 ,$$

with equality when $\dot{\boldsymbol{x}}_t$ and $\dot{y}_t$ are constant in time. Thus the path in Equation (4.3) minimizes Equation (4.4), with minimum Equation (4.5).

Let $\mathcal{B}_\delta$ and $\mathcal{K}_\delta$ denote the values of Equations (4.6) and (4.7), respectively. We prove the two inequalities separately.

Dynamic cost is no smaller than static cost: $\mathcal{B}_\delta \geq \mathcal{K}_\delta$.

For a dynamic solution represented by a particle flow, let $\hat{\boldsymbol{s}}_t = (\hat{\boldsymbol{u}}_t, \hat{g}_t)$. Let $\boldsymbol{W}_t(\boldsymbol{w}_0)$ denote the extended state reached at time t by a particle that starts from $\boldsymbol{w}_0$ and follows the extended velocity field $\hat{\boldsymbol{s}}_t$; equivalently, $\boldsymbol{W}_t(\boldsymbol{w}_0)$ satisfies

$$\dot{\boldsymbol{W}}_t(\boldsymbol{w}_0) = \hat{\boldsymbol{s}}_t\big(\boldsymbol{W}_t(\boldsymbol{w}_0)\big), \qquad \boldsymbol{W}_0(\boldsymbol{w}_0) = \boldsymbol{w}_0.$$

The extended continuity equation implies

$$\hat{\rho}_t = (\boldsymbol{W}_t)_{\#}\hat{\rho}_0.$$

Hence

$$\hat{\pi} := (\mathrm{id}, \boldsymbol{W}_1)_{\#}\hat{\rho}_0$$

is a coupling between $\hat{\rho}_0$ and $\hat{\rho}_1$. Its terminal marginal is the same $\hat{\rho}_1$ as in the dynamic problem and therefore satisfies the required weighted projection onto ρ_1.

Using the pushforward identity and then Cauchy–Schwarz along each particle path gives

$$\begin{aligned}
&\int_0^1 \int_{\mathcal{W}} \|\hat{\boldsymbol{s}}_t(\boldsymbol{w})\|_\delta^2 \,\mathrm{d}\hat{\rho}_t(\boldsymbol{w})\,\mathrm{d}t \\
&= \int_{\mathcal{W}} \int_0^1 \left\|\dot{\boldsymbol{W}}_t(\boldsymbol{w}_0)\right\|_\delta^2 \mathrm{d}t\,\mathrm{d}\hat{\rho}_0(\boldsymbol{w}_0) \\
&\geq \int_{\mathcal{W}} \|\boldsymbol{W}_1(\boldsymbol{w}_0) - \boldsymbol{w}_0\|_\delta^2 \,\mathrm{d}\hat{\rho}_0(\boldsymbol{w}_0) \\
&= \iint_{\mathcal{W}^2} \|\boldsymbol{w}_1 - \boldsymbol{w}_0\|_\delta^2 \,\mathrm{d}\hat{\pi}(\boldsymbol{w}_0, \boldsymbol{w}_1) \geq \mathcal{K}_\delta(\rho_0, \rho_1).
\end{aligned}$$

Since this holds for every feasible dynamic solution, taking the infimum gives

$$\mathcal{B}_\delta(\rho_0, \rho_1) \geq \mathcal{K}_\delta(\rho_0, \rho_1).$$

Static cost is no smaller than dynamic cost: $\mathcal{K}_\delta \geq \mathcal{B}_\delta$.

Conversely, let $(\hat{\pi}, \hat{\rho}_1)$ be any feasible static solution and define the straight conditional path

$$\boldsymbol{W}_t(\boldsymbol{w}_0, \boldsymbol{w}_1) := (1-t)\boldsymbol{w}_0 + t\boldsymbol{w}_1, \qquad \hat{\rho}_t := (\boldsymbol{W}_t)_{\#}\hat{\pi}.$$

The boundary conditions follow from the two marginals of $\hat{\pi}$. At an intermediate state $\boldsymbol{w}$, define the macroscopic extended velocity as the conditional average

$$\hat{\boldsymbol{s}}_t(\boldsymbol{w}) := \mathbb{E}_{\hat{\pi}}\left[\boldsymbol{w}_1 - \boldsymbol{w}_0 \mid \boldsymbol{W}_t = \boldsymbol{w}\right].$$

Applying the chain rule to a smooth test function along every straight path and then integrating over $\hat{\pi}$ shows that $(\hat{\rho}_t, \hat{\boldsymbol{s}}_t)$ satisfies the extended continuity equation in the weak sense.

Indeed, let $\varphi \in C_c^1([0,1] \times \mathcal{W})$. Along each path,

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi\big(t, \boldsymbol{W}_t(\boldsymbol{w}_0, \boldsymbol{w}_1)\big) = \partial_t\varphi(t, \boldsymbol{W}_t) + \nabla_{\boldsymbol{w}}\varphi(t, \boldsymbol{W}_t)\cdot(\boldsymbol{w}_1 - \boldsymbol{w}_0).$$

Integrating first over t and then over the endpoint pairs gives

$$\begin{aligned}
&\int_{\mathcal{W}} \varphi(1, \boldsymbol{w})\,\mathrm{d}\hat{\rho}_1(\boldsymbol{w}) - \int_{\mathcal{W}} \varphi(0, \boldsymbol{w})\,\mathrm{d}\hat{\rho}_0(\boldsymbol{w}) \\
&= \int_0^1 \iint_{\mathcal{W}^2} \Big[\partial_t\varphi(t, \boldsymbol{W}_t) + \nabla_{\boldsymbol{w}}\varphi(t, \boldsymbol{W}_t)\cdot(\boldsymbol{w}_1 - \boldsymbol{w}_0)\Big]\,\mathrm{d}\hat{\pi}\,\mathrm{d}t \\
&= \int_0^1 \int_{\mathcal{W}} \Big[\partial_t\varphi(t, \boldsymbol{w}) + \nabla_{\boldsymbol{w}}\varphi(t, \boldsymbol{w})\cdot\hat{\boldsymbol{s}}_t(\boldsymbol{w})\Big]\,\mathrm{d}\hat{\rho}_t(\boldsymbol{w})\,\mathrm{d}t.
\end{aligned}$$

The last equality uses $\hat{\rho}_t = (\boldsymbol{W}_t)_{\#}\hat{\pi}$ and the definition of $\hat{\boldsymbol{s}}_t$ as the conditional mean. This identity is precisely the weak form of

$$\partial_t\hat{\rho}_t + \mathrm{div}_{\boldsymbol{w}}\big(\hat{\rho}_t\hat{\boldsymbol{s}}_t\big) = 0,$$

which is the extended continuity equation in Equation (4.6).

We next compare the costs. For fixed t, the definition of the conditional mean and Jensen's inequality imply

$$\begin{aligned}\|\hat{\boldsymbol{s}}_t(\boldsymbol{w})\|_\delta^2 &= \|\mathbb{E}_{\hat{\pi}}\left[\boldsymbol{w}_1 - \boldsymbol{w}_0 \mid \boldsymbol{W}_t = \boldsymbol{w}\right]\|_\delta^2 \\ &\le \mathbb{E}_{\hat{\pi}}\left[\|\boldsymbol{w}_1 - \boldsymbol{w}_0\|_\delta^2 \mid \boldsymbol{W}_t = \boldsymbol{w}\right].\end{aligned}$$

Since $\hat{\rho}_t = (\boldsymbol{W}_t)_{\#}\hat{\pi}$, averaging this inequality over the intermediate states $\boldsymbol{w}$ recovers the expectation over all endpoint pairs:

$$\begin{aligned}\int_{\mathcal{W}} \|\hat{\boldsymbol{s}}_t(\boldsymbol{w})\|_\delta^2 \,\mathrm{d}\hat{\rho}_t(\boldsymbol{w}) &\le \int_{\mathcal{W}} \mathbb{E}_{\hat{\pi}}\left[\|\boldsymbol{w}_1 - \boldsymbol{w}_0\|_\delta^2 \mid \boldsymbol{W}_t = \boldsymbol{w}\right] \mathrm{d}\hat{\rho}_t(\boldsymbol{w}) \\ &= \mathbb{E}_{\hat{\pi}}\left[\|\boldsymbol{w}_1 - \boldsymbol{w}_0\|_\delta^2\right] \\ &= \iint_{\mathcal{W}^2} \|\boldsymbol{w}_1 - \boldsymbol{w}_0\|_\delta^2 \,\mathrm{d}\hat{\pi}.\end{aligned}$$

Integrating over $t \in [0,1]$ therefore yields

$$\int_0^1 \int_{\mathcal{W}} \|\hat{\boldsymbol{s}}_t(\boldsymbol{w})\|_\delta^2 \,\mathrm{d}\hat{\rho}_t(\boldsymbol{w})\,\mathrm{d}t \le \iint_{\mathcal{W}^2} \|\boldsymbol{w}_1 - \boldsymbol{w}_0\|_\delta^2 \,\mathrm{d}\hat{\pi}.$$

Thus every feasible static solution generates a feasible dynamic solution with no larger cost. Taking the infimum gives

$$\mathcal{B}_\delta(\rho_0, \rho_1) \le \mathcal{K}_\delta(\rho_0, \rho_1).$$

Combining the two inequalities gives

$$\mathcal{B}_\delta(\rho_0, \rho_1) = \mathcal{K}_\delta(\rho_0, \rho_1),$$

which proves the equivalence of Equations (4.6) and (4.7).

A.4 Proof of Theorem 4.3

This section proves Theorem 4.3 by first eliminating y_0 and then solving the local optimization over y_1.

Source reduction and the local problem.

Since $\hat{\rho}_0(\boldsymbol{x}_0, y_0) = \rho_0(\boldsymbol{x}_0)\delta_0(y_0)$, every feasible extended plan is concentrated on $y_0 = 0$. After imposing $y_1 \le M$, the static problem becomes

$$\begin{aligned}\inf_{\pi \ge 0} \quad & \iiint \left[\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|^2 + \delta^2 y_1^2\right] \pi(\boldsymbol{x}_0, \boldsymbol{x}_1, y_1)\,\mathrm{d}\boldsymbol{x}_0\,\mathrm{d}\boldsymbol{x}_1\,\mathrm{d}y_1, \\ \text{s.t.} \quad & \operatorname{supp}\pi \subseteq \mathbb{R}^d \times \mathbb{R}^d \times (-\infty, M], \\ & \iint \pi(\boldsymbol{x}_0, \boldsymbol{x}_1, y_1)\,\mathrm{d}\boldsymbol{x}_1\,\mathrm{d}y_1 = \rho_0(\boldsymbol{x}_0), \\ & \iint \mathrm{e}^{y_1}\pi(\boldsymbol{x}_0, \boldsymbol{x}_1, y_1)\,\mathrm{d}\boldsymbol{x}_0\,\mathrm{d}y_1 = \rho_1(\boldsymbol{x}_1).\end{aligned} \tag{A.8}$$

For a fixed endpoint state pair, write $A = \gamma_0(\boldsymbol{x}_0, \boldsymbol{x}_1)$, $B = \gamma_1(\boldsymbol{x}_0, \boldsymbol{x}_1)$, $r = B/A$, and $f(y_1) = \pi(\boldsymbol{x}_0, \boldsymbol{x}_1, y_1)$. We consider $A > 0$ and $B > 0$; pairs with $A = B = 0$ contribute nothing. The state cost is fixed, and the two moments of f must remain equal to A and B. Removing the common coefficient δ^2 gives the local problem:

$$\begin{aligned}\inf_{f \in \mathcal{M}_+((-\infty, M])} \quad & \int_{-\infty}^{M} y^2 f(y)\,\mathrm{d}y, \\ \text{s.t.} \quad & \int_{-\infty}^{M} f(y)\,\mathrm{d}y = A, \qquad \int_{-\infty}^{M} \mathrm{e}^y f(y)\,\mathrm{d}y = B.\end{aligned} \tag{A.9}$$

Let $r = B/A$. Feasibility requires $0 < r \le \mathrm{e}^M$.

Lagrangian and dual problem.

The Lagrangian associated with Equation (A.9) is

$$\mathcal{L} = \int_{-\infty}^{M} \left(y^2 - \lambda_1 - \lambda_2 \mathrm{e}^y\right) f(y)\,\mathrm{d}y + \lambda_1 A + \lambda_2 B.$$

Minimization over $f \in \mathcal{M}_+\big((-\infty, M]\big)$ requires

$$y^2 - \lambda_1 - \lambda_2 \mathrm{e}^y \geq 0, \qquad \text{for } y \leq M.$$

Hence

$$\lambda_1 \leq \inf_{y \leq M} \left(y^2 - \lambda_2 \mathrm{e}^y\right).$$

Define

$$g_{\lambda_2}(y) := y^2 - \lambda_2 \mathrm{e}^y.$$

For fixed λ_2, the dual objective is maximized by taking

$$\lambda_1 = \inf_{y \leq M} g_{\lambda_2}(y).$$

Therefore, the dual problem can be written as

$$\sup_{\lambda_2 \in \mathbb{R}} \left\{ A \inf_{y \leq M} g_{\lambda_2}(y) + \lambda_2 B \right\}. \tag{A.10}$$

Solving the dual problem and determining $y_\star$.

To solve the dual problem in Equation (A.10), we first examine the minimum of g_{λ_2} over $(-\infty, M]$. Its derivative is

$$g'_{\lambda_2}(y) = 2y - \lambda_2 \mathrm{e}^y.$$

An interior stationary point $y_\star(\lambda_2)$ satisfies

$$2y_\star(\lambda_2)\mathrm{e}^{-y_\star(\lambda_2)} = \lambda_2.$$

The extrema of g_{λ_2} have the following structure:

- If $\lambda_2 \leq 0$, then $g'_{\lambda_2}(y) = 0$ has a unique solution, and the minimizer satisfies $y_\star(\lambda_2) \leq 0$.
- If $0 < \lambda_2 < \frac{2}{\mathrm{e}}$, the function
  $$\phi(y) := 2y\mathrm{e}^{-y}$$
  reaches its maximum $\frac{2}{\mathrm{e}}$ at $y = 1$. In this case, $g'_{\lambda_2}(y) = 0$ has two roots, and its unique local minimum satisfies
  $$y_\star(\lambda_2) \in (0, 1).$$
  The global minimum may occur either at this interior point or at the boundary point M.
- If $\lambda_2 \geq \frac{2}{\mathrm{e}}$, then $g'_{\lambda_2}(y) \leq 0$ for all y. Thus, g_{λ_2} is decreasing and reaches its minimum at M.

Consequently,

$$\inf_{y \leq M} g_{\lambda_2}(y) = \min\left\{g_{\lambda_2}\big(y_\star(\lambda_2)\big), g_{\lambda_2}(M)\right\}.$$

Moreover,

$$g_{\lambda_2}\big(y_\star(\lambda_2)\big) - g_{\lambda_2}(M) = y_\star(\lambda_2)^2 - M^2 + \lambda_2\left(\mathrm{e}^M - \mathrm{e}^{y_\star(\lambda_2)}\right).$$

By the envelope theorem,

$$\frac{\mathrm{d}}{\mathrm{d}\lambda_2}\left[g_{\lambda_2}\big(y_\star(\lambda_2)\big) - g_{\lambda_2}(M)\right] = \mathrm{e}^M - \mathrm{e}^{y_\star(\lambda_2)} > 0.$$

Thus, as λ_2 increases, the global minimizer switches from the interior point $y_\star(\lambda_2)$ to the boundary point M. Let λ_2^* be the switching value and denote the corresponding interior minimizer by

$$y_\star := y_\star(\lambda_2^*).$$

At the switching point,

$$g_{\lambda_2^*}(y_\star) = g_{\lambda_2^*}(M). \tag{A.11}$$

Therefore,

$$y_{\min}(\lambda_2) = \begin{cases} y_\star(\lambda_2), & \lambda_2 < \lambda_2^*, \\ y_\star \text{ or } M, & \lambda_2 = \lambda_2^*, \\ M, & \lambda_2 > \lambda_2^*. \end{cases}$$

We now determine $y_\star$. Since it is an interior local minimizer,

$$\lambda_2^* = 2y_\star \mathrm{e}^{-y_\star}.$$

Substituting this identity into Equation (A.11) gives

$$y_\star^2 - \left(2y_\star \mathrm{e}^{-y_\star}\right) \mathrm{e}^{y_\star} = M^2 - \left(2y_\star \mathrm{e}^{-y_\star}\right) \mathrm{e}^{M}.$$

After simplification,

$$M^2 - y_\star^2 + 2y_\star - 2y_\star \mathrm{e}^{M-y_\star} = 0. \tag{A.12}$$

Define

$$h(y) := M^2 - y^2 + 2y - 2y\mathrm{e}^{M-y}.$$

Then

$$h'(y) = 2(y-1)\left(\mathrm{e}^{M-y} - 1\right).$$

For $M > 1$,

$$h(0) = M^2 > 0, \qquad h(1) = M^2 + 1 - 2\mathrm{e}^{M-1} < 0, \qquad h(M) = 0.$$

Moreover, h is strictly decreasing on $(0, 1)$ and strictly increasing on $(1, M)$. Hence, besides the trivial root $y = M$, the equation $h(y) = 0$ has a unique nontrivial solution $y_\star \in (0, 1)$, which is uniquely determined by M.

Maximizing the dual objective and recovering the optimizer.

Define the dual objective

$$D(\lambda_2) := A g_{\lambda_2}\big(y_{\min}(\lambda_2)\big) + \lambda_2 B.$$

By the envelope theorem,

$$D'(\lambda_2) = B - A\mathrm{e}^{y_{\min}(\lambda_2)}.$$

Therefore, if $D'(\lambda_2) = 0$, the desired minimizer satisfies

$$y_{\min}(\lambda_2) = \ln \frac{B}{A}.$$

The solution has two cases, depending on the position of $\ln(B/A)$ relative to $y_\star$.

Case 1: $\ln(B/A) \le y_\star$**.** In this case, $D'(\lambda_2) = 0$ can be satisfied before λ_2 reaches λ_2^*. Hence

$$y_{\min}(\lambda_2) = y_\star(\lambda_2) = \ln \frac{B}{A}.$$

The optimal plan is concentrated at the single point $\ln(B/A)$:

$$f^*(y) = A\delta\left(y - \ln \frac{B}{A}\right).$$

The corresponding minimum is

$$\inf_f \int_{-\infty}^{M} y^2 f(y)\,\mathrm{d}y = A\left(\ln \frac{B}{A}\right)^2.$$

Case 2: $y_\star \le \ln(B/A) \le M$**.** When $\lambda_2 < \lambda_2^*$,

$$y_{\min}(\lambda_2) = y_\star(\lambda_2) < y_\star \le \ln \frac{B}{A},$$

and therefore

$$D'(\lambda_2) = B - A\mathrm{e}^{y_{\min}(\lambda_2)} > 0.$$

When $\lambda_2 > \lambda_2^*$,

$$y_{\min}(\lambda_2) = M \geq \ln \frac{B}{A},$$

and therefore

$$D'(\lambda_2) = B - A\mathrm{e}^{M} \leq 0.$$

Thus, the dual objective reaches its maximum at λ_2^*. At this point,

$$g_{\lambda_2^*}(y_\star) = g_{\lambda_2^*}(M),$$

and the optimal plan is supported on $y_\star$ and M:

$$f^*(y) = w_1\delta(y - y_\star) + w_2\delta(y - M).$$

The weights are uniquely determined by

$$w_1 + w_2 = A,$$
$$w_1\mathrm{e}^{y_\star} + w_2\mathrm{e}^{M} = B,$$

which gives

$$w_1 = \frac{A\mathrm{e}^{M} - B}{\mathrm{e}^{M} - \mathrm{e}^{y_\star}}, \qquad w_2 = \frac{B - A\mathrm{e}^{y_\star}}{\mathrm{e}^{M} - \mathrm{e}^{y_\star}}.$$

The corresponding minimum is

$$\begin{aligned}\inf_f \int_{-\infty}^{M} y^2 f(y)\,\mathrm{d}y &= w_1 y_\star^2 + w_2 M^2 \\ &= \frac{A\mathrm{e}^{M} - B}{\mathrm{e}^{M} - \mathrm{e}^{y_\star}}\, y_\star^2 + \frac{B - A\mathrm{e}^{y_\star}}{\mathrm{e}^{M} - \mathrm{e}^{y_\star}}\, M^2.\end{aligned}$$

Writing $r = B/A$, an optimizer is

$$f^\star(y_1) = \begin{cases} A\delta(y_1 - \ln r)\,, & 0 < r \leq \mathrm{e}^{y_\star}, \\ A\left[\dfrac{\mathrm{e}^{M} - r}{\mathrm{e}^{M} - \mathrm{e}^{y_\star}}\delta(y_1 - y_\star) + \dfrac{r - \mathrm{e}^{y_\star}}{\mathrm{e}^{M} - \mathrm{e}^{y_\star}}\delta(y_1 - M)\right], & \mathrm{e}^{y_\star} < r \leq \mathrm{e}^{M}. \end{cases} \tag{A.13}$$

The minimum is denoted by $A\Phi_M(r)$, where

$$\Phi_M(r) = \begin{cases} (\ln r)^2, & 0 < r \leq \mathrm{e}^{y_\star}, \\ \dfrac{\mathrm{e}^{M} - r}{\mathrm{e}^{M} - \mathrm{e}^{y_\star}} y_\star^2 + \dfrac{r - \mathrm{e}^{y_\star}}{\mathrm{e}^{M} - \mathrm{e}^{y_\star}} M^2, & \mathrm{e}^{y_\star} < r \leq \mathrm{e}^{M}. \end{cases}$$

Using the critical equation,

$$\frac{M^2 - y_\star^2}{\mathrm{e}^{M} - \mathrm{e}^{y_\star}} = \frac{2y_\star}{\mathrm{e}^{y_\star}}, \qquad \frac{\mathrm{e}^{M} y_\star^2 - \mathrm{e}^{y_\star} M^2}{\mathrm{e}^{M} - \mathrm{e}^{y_\star}} = y_\star^2 - 2y_\star.$$

Therefore,

$$\Phi_M(r) = \begin{cases} (\ln r)^2, & 0 < r \leq \mathrm{e}^{y_\star}, \\ \dfrac{2y_\star}{\mathrm{e}^{y_\star}} r + y_\star^2 - 2y_\star, & \mathrm{e}^{y_\star} < r \leq \mathrm{e}^{M}. \end{cases} \tag{A.14}$$

This function is continuous and continuously differentiable at $r = \mathrm{e}^{y_\star}$.

Recovering the reduced problem.

For any feasible plan in Equation (A.8), the local bound gives

$$\int_{-\infty}^{M} y_1^2 \pi(\boldsymbol{x}_0, \boldsymbol{x}_1, y_1)\,\mathrm{d}y_1 \geq \gamma_0(\boldsymbol{x}_0, \boldsymbol{x}_1)\Phi_M(R(\boldsymbol{x}_0, \boldsymbol{x}_1)).$$

Integrating over endpoint state pairs bounds its cost from below by the reduced objective in Equation (4.8). Conversely, for any feasible (γ_0, γ_1) in that reduced problem, use Equation (A.13) at each pair with $A = \gamma_0$ and $r = R$. The resulting plan has marginals γ_0 and γ_1, satisfies the growth cap, and attains the local bound. Thus the two problems have the same infimum. This proves Theorem 4.3.

A.5 Proof of Proposition 4.1 and Update Rules

This section derives the Sinkhorn-like updates and proves Proposition 4.1.

Discrete reduced problem.

Let the empirical source and target measures be

$$\rho_0 = \sum_{i=1}^{n} a_i \delta_{\boldsymbol{x}_i^0}, \qquad \rho_1 = \sum_{j=1}^{m} b_j \delta_{\boldsymbol{x}_j^1},$$

Define the state space cost

$$C_{ij} = \left\| \boldsymbol{x}_i^0 - \boldsymbol{x}_j^1 \right\|^2.$$

We write Γ_{ij} for the discrete base coupling γ_0 and R_{ij} for the local growth ratio $\mathrm{d}\gamma_1 / \mathrm{d}\gamma_0$. Then Equation (4.8) becomes

$$\begin{aligned} \min_{\Gamma \geq 0,\ 0 < R \leq \mathrm{e}^M} \quad & \sum_{i=1}^{n} \sum_{j=1}^{m} \left[C_{ij} + \delta^2 \Phi_M(R_{ij}) \right] \Gamma_{ij}, \\ \text{s.t.} \quad & \sum_{j=1}^{m} \Gamma_{ij} = a_i, \qquad i = 1, \ldots, n, \\ & \sum_{i=1}^{n} R_{ij} \Gamma_{ij} = b_j, \qquad j = 1, \ldots, m. \end{aligned} \tag{A.15}$$

Entropic regularization and Lagrangian.

For computation we add an entropy term to the base coupling:

$$\min_{\Gamma \geq 0,\ 0 < R \leq \mathrm{e}^M} \quad \sum_{ij} \Big[C_{ij} \Gamma_{ij} + \delta^2 \Gamma_{ij} \Phi_M(R_{ij}) + \varepsilon \Gamma_{ij} (\ln \Gamma_{ij} - 1) \Big], \tag{A.16}$$

subject to the same constraints as in Equation (A.15). Introduce row dual variables u_i and column dual variables v_j. The Lagrangian of Equation (A.16) is

$$\begin{aligned} \mathcal{L}(\Gamma, R; u, v) = & \sum_{ij} \Gamma_{ij} \left[C_{ij} + \delta^2 \Phi_M(R_{ij}) + \varepsilon (\ln \Gamma_{ij} - 1) - u_i - v_j R_{ij} \right] \\ & + \sum_i u_i a_i + \sum_j v_j b_j. \end{aligned}$$

For fixed R, stationarity with respect to Γ_{ij} gives

$$\Gamma_{ij} = \exp\left(\frac{u_i + v_j R_{ij} - C_{ij} - \delta^2 \Phi_M(R_{ij})}{\varepsilon} \right). \tag{A.17}$$

Substituting Equation (A.17) into the Lagrangian gives the dual function

$$\inf_{\Gamma} \mathcal{L}(\Gamma, R; u, v) = \sum_i u_i a_i + \sum_j v_j b_j - \varepsilon \sum_{ij} \exp\left(\frac{u_i + v_j R_{ij} - C_{ij} - \delta^2 \Phi_M(R_{ij})}{\varepsilon} \right).$$

Thus the remaining local step is a scalar maximization over the growth ratio. We make this step explicit next.

The scalar R subproblem.

For a fixed target index j and dual variable v_j, define the scalar objective for the growth-ratio update as

$$h_j(R; v_j) := v_j R - \delta^2 \Phi_M(R), \qquad 0 < R \leq \mathrm{e}^M.$$

Define the column value function as the optimal value of this one-dimensional problem:

$$H_j(v_j) := \max_{0 < R \leq \mathrm{e}^M} h_j(R; v_j).$$

Using the piecewise expression of Φ_M in Equation (A.14), the objective h_j becomes

$$h_j(R; v_j) = \begin{cases} v_j R - \delta^2 (\ln R)^2, & 0 < R \le \mathrm{e}^{y_\star}, \\ \left(v_j - \delta^2 \dfrac{2y_\star}{\mathrm{e}^{y_\star}} \right) R + \delta^2 (2y_\star - y_\star^2), & \mathrm{e}^{y_\star} < R \le \mathrm{e}^M. \end{cases}$$

The slope of the second branch is

$$v_j - \delta^2 \frac{2y_\star}{\mathrm{e}^{y_\star}},$$

and its sign determines the following three regimes.

Interior branch. If $v_j < \delta^2 2y_\star / \mathrm{e}^{y_\star}$, the second branch is decreasing, so the maximizer lies on the first branch. Setting $p = \ln R$ and differentiating gives

$$v_j - 2\delta^2 p \mathrm{e}^{-p} = 0.$$

Thus the interior optimizer is

$$R^\star(v_j) = \exp\left[-W_0\left(-\frac{v_j}{2\delta^2} \right) \right],$$

where W_0 denotes the principal branch of the Lambert W function, defined by $W_0(z)\mathrm{e}^{W_0(z)} = z$. The principal branch W_0 is increasing on $[-1/\mathrm{e}, \infty)$. Since the argument $-v_j/(2\delta^2)$ decreases as v_j increases, the optimizer $R^\star(v_j) = \exp[-W_0(-v_j/(2\delta^2))]$ is increasing in v_j.

Transition branch. If $v_j = \delta^2 2y_\star / \mathrm{e}^{y_\star}$, the derivative on the first branch is

$$\frac{\partial}{\partial R} h_j(R; v_j) = 2\delta^2 \left(\frac{y_\star}{\mathrm{e}^{y_\star}} - \frac{\ln R}{R} \right).$$

For $0 < R \le \mathrm{e}^{y_\star}$ and $y_\star < 1$, we have $\ln R / R \le y_\star / \mathrm{e}^{y_\star}$, so the first branch is nondecreasing and reaches its maximum at $R = \mathrm{e}^{y_\star}$. The second branch is flat and has the same value as this boundary value. Every point in $[\mathrm{e}^{y_\star}, \mathrm{e}^M]$ is therefore optimal.

Capped branch. If $v_j > \delta^2 2y_\star / \mathrm{e}^{y_\star}$, the second branch is increasing, so the maximizer is attained at the growth cap $R = \mathrm{e}^M$.

Combining the three regimes gives the optimal set

$$\mathcal{R}^\star(v_j) = \operatorname*{arg\,max}_{0 < R \le \mathrm{e}^M} h_j(R; v_j) = \begin{cases} \left\{ \exp\left[-W_0\left(-\dfrac{v_j}{2\delta^2} \right) \right] \right\}, & v_j < \delta^2 \dfrac{2y_\star}{\mathrm{e}^{y_\star}}, \\ [\mathrm{e}^{y_\star}, \mathrm{e}^M], & v_j = \delta^2 \dfrac{2y_\star}{\mathrm{e}^{y_\star}}, \\ \{\mathrm{e}^M\}, & v_j > \delta^2 \dfrac{2y_\star}{\mathrm{e}^{y_\star}}. \end{cases}$$

The corresponding value function is

$$H_j(v_j) = \begin{cases} \delta^2 \left[2 \ln R^\star(v_j) - \left(\ln R^\star(v_j) \right)^2 \right], & v_j < \delta^2 \dfrac{2y_\star}{\mathrm{e}^{y_\star}}, \\ \delta^2 (2y_\star - y_\star^2), & v_j = \delta^2 \dfrac{2y_\star}{\mathrm{e}^{y_\star}}, \\ \left(v_j - \delta^2 \dfrac{2y_\star}{\mathrm{e}^{y_\star}} \right) \mathrm{e}^M + \delta^2 (2y_\star - y_\star^2), & v_j > \delta^2 \dfrac{2y_\star}{\mathrm{e}^{y_\star}}. \end{cases}$$

After this elimination, the dual problem can be written as

$$\max_{u,v} \quad \mathcal{D}(u, v) = \sum_i u_i a_i + \sum_j v_j b_j - \varepsilon \sum_{ij} \exp\left(\frac{u_i + H_j(v_j) - C_{ij}}{\varepsilon} \right). \tag{A.18}$$

In what follows, R_j denotes a selected element of $\mathcal{R}^\star(v_j)$. By the envelope theorem, $H_j'(v_j) = R_j$ when the maximizer is unique; on the transition branch, the same relation is understood as a

subgradient condition with $R_j \in [\mathrm{e}^{y_\star}, \mathrm{e}^{M}]$. Taking derivatives of $\mathcal{D}$ in Equation (A.18) with respect to u_i and v_j gives

$$\frac{\partial \mathcal{D}}{\partial u_i} = a_i - \sum_j \exp\left(\frac{u_i + H_j - C_{ij}}{\varepsilon}\right), \qquad \frac{\partial \mathcal{D}}{\partial v_j} = b_j - R_j \sum_i \exp\left(\frac{u_i + H_j - C_{ij}}{\varepsilon}\right),$$

where $H_j = H_j(v_j)$. Setting these conditions to zero yields

$$a_i = \sum_j \exp\left(\frac{u_i + H_j - C_{ij}}{\varepsilon}\right), \tag{A.19a}$$

$$b_j = R_j \sum_i \exp\left(\frac{u_i + H_j - C_{ij}}{\varepsilon}\right). \tag{A.19b}$$

Alternating updates in (u, H).

The explicit solver uses u_i and the column values H_j as update variables. Given H, the row condition Equation (A.19a) gives the closed-form update

$$u_i = \varepsilon \ln a_i - \varepsilon \log \sum_j \exp\left(\frac{H_j - C_{ij}}{\varepsilon}\right). \tag{A.20}$$

Given u, define

$$K_j := \frac{b_j}{\sum_i \exp((u_i - C_{ij})/\varepsilon)}. \tag{A.21}$$

The column condition Equation (A.19b) becomes

$$R_j \exp(H_j/\varepsilon) = K_j, \qquad \text{or} \qquad \ln R_j + \frac{H_j}{\varepsilon} = \ln K_j. \tag{A.22}$$

We solve this scalar relation by cases. The only nontrivial case is the interior branch, where the left-hand side is a strictly increasing function of $\ln R_j$ because $\ln R_j \le y_\star < 1$; the transition and capped branches give explicit values. Thus, once the branch is selected by K_j, $\ln R_j$ and H_j is unique. Define

$$T_1 := y_\star + \frac{\delta^2(2y_\star - y_\star^2)}{\varepsilon}, \qquad T_2 := M + \frac{\delta^2(2y_\star - y_\star^2)}{\varepsilon}. \tag{A.23}$$

Using the thresholds in Equation (A.23), the three regimes give the following updates.

Interior branch. If $\ln K_j \le T_1$, set $p_j = \ln R_j$. On this branch,

$$H_j = \delta^2(2p_j - p_j^2),$$

so Equation (A.22) gives

$$p_j + \frac{\delta^2}{\varepsilon}(2p_j - p_j^2) = \ln K_j.$$

Solving the quadratic and taking the root with $p_j \le y_\star$ yields

$$p_j = \frac{(\varepsilon + 2\delta^2) - \sqrt{(\varepsilon + 2\delta^2)^2 - 4\delta^2 \varepsilon \ln K_j}}{2\delta^2},$$

and then

$$R_j = \mathrm{e}^{p_j}, \qquad H_j = \delta^2(2p_j - p_j^2).$$

Transition branch. If $T_1 < \ln K_j \le T_2$, the column value is constant:

$$H_j = \delta^2(2y_\star - y_\star^2).$$

Then Equation (A.22) gives

$$R_j = K_j \exp\left(-\frac{\delta^2(2y_\star - y_\star^2)}{\varepsilon}\right).$$

The threshold condition is exactly the statement that R_j lies in $[\mathrm{e}^{y_\star}, \mathrm{e}^{M}]$.

Capped branch. If $\ln K_j > T_2$, the maximizer is $R_j = \mathrm{e}^M$. Equation (A.22) gives

$$R_j = \mathrm{e}^M, \qquad H_j = \varepsilon(\ln K_j - M).$$

Recovery and sampling quantities.

After convergence, the base coupling and the target-mass coupling are recovered as

$$\Gamma^{(0)}_{ij} = \exp\left(\frac{u_i + H_j - C_{ij}}{\varepsilon}\right), \qquad \Gamma^{(1)}_{ij} = R_j \Gamma^{(0)}_{ij}. \tag{A.24}$$

The recovered base coupling $\Gamma^{(0)}$ is used to sample $(\boldsymbol{x}_0, \boldsymbol{x}_1)$ and fix the initial log-mass as $y_0 = 0$. The terminal log-mass y_1 is then sampled according to the normalized version of the local optimizer in Equation (A.13), using the recovered growth ratio R_j:

$$f^*(y_1) = \begin{cases} \delta(y_1 - \ln R_j), & 0 < R_j \le \mathrm{e}^{y_\star}, \\ \dfrac{\mathrm{e}^M - R_j}{\mathrm{e}^M - \mathrm{e}^{y_\star}}\delta(y_1 - y_\star) + \dfrac{R_j - \mathrm{e}^{y_\star}}{\mathrm{e}^M - \mathrm{e}^{y_\star}}\delta(y_1 - M), & \mathrm{e}^{y_\star} < R_j \le \mathrm{e}^M. \end{cases} \tag{A.25}$$

Thus, if $R_j \le \mathrm{e}^{y_\star}$, we set $y_1 = \ln R_j$; otherwise we sample $y_1 = y_\star$ with probability $(\mathrm{e}^M - R_j)/(\mathrm{e}^M - \mathrm{e}^{y_\star})$ and $y_1 = M$ with probability $(R_j - \mathrm{e}^{y_\star})/(\mathrm{e}^M - \mathrm{e}^{y_\star})$.

The complete Sinkhorn-like solver is summarized below; it computes the recovered couplings and growth ratios used for endpoint and terminal log-mass sampling.

Algorithm 2 Entropic Solver for the ESFM Static Coupling

1: **Input:** source weights and states $\{a_i, \boldsymbol{x}_i^0\}_{i=1}^n$, target weights and states $\{b_j, \boldsymbol{x}_j^1\}_{j=1}^m$, parameters δ, M, ε.
2: Compute $C_{ij} = \left\|\boldsymbol{x}_i^0 - \boldsymbol{x}_j^1\right\|^2$ and solve for $y_\star \in (0,1)$ using Equation (A.12).
3: Check $\sum_j b_j \le \mathrm{e}^M \sum_i a_i$.
4: Initialize H_j.
5: **repeat**
6: Update every u_i using Equation (A.20).
7: Compute every K_j using Equation (A.21).
8: Update every H_j and R_j using the three branch formulas above.
9: **until** the marginal residuals or iterate changes are below tolerance
10: Recover $\Gamma^{(0)}$ and $\Gamma^{(1)}$ using Equation (A.24).
11: **Output:** $\Gamma^{(0)}$, $\Gamma^{(1)}$, and R.

A.6 Proof of Proposition 4.2

Consider the discrete problem

$$\begin{aligned} \min_{\Gamma \ge 0,\ 0 < R \le \mathrm{e}^M} \quad & \sum_{i,j} \left[C_{ij} + \delta^2 \Phi_M(R_{ij})\right] \Gamma_{ij}, \\ \text{s.t.} \quad & \sum_j \Gamma_{ij} = a_i, \qquad \sum_i R_{ij}\Gamma_{ij} = b_j. \end{aligned} \tag{A.26}$$

Let $a_i = a_{i_1} + a_{i_2}$ with $a_{i_1}, a_{i_2} > 0$.

Proof. Because the two split particles remain at the same location as the original particle,

$$C_{i_1 j} = C_{i_2 j} = C_{ij}.$$

First consider any feasible solution of the split problem, allowing $R_{i_1 j}$ and $R_{i_2 j}$ to be chosen independently. For each target particle j, define

$$\Gamma_{ij} = \Gamma_{i_1 j} + \Gamma_{i_2 j},$$

and, whenever $\Gamma_{ij} > 0$,

$$R_{ij} = \frac{R_{i_1 j}\Gamma_{i_1 j} + R_{i_2 j}\Gamma_{i_2 j}}{\Gamma_{i_1 j} + \Gamma_{i_2 j}}.$$

If $\Gamma_{ij} = 0$, choose any $R_{ij} \in (0, \mathrm{e}^M]$; the corresponding transport and growth contributions are zero. This aggregation preserves the source and target constraints as well as the bound $0 < R_{ij} \le \mathrm{e}^M$. The spatial contribution is unchanged because the split particles are at the same location. Since Φ_M is convex, Jensen's inequality gives

$$\Gamma_{i_1 j}\Phi_M(R_{i_1 j}) + \Gamma_{i_2 j}\Phi_M(R_{i_2 j}) \ge \Gamma_{ij}\Phi_M(R_{ij}).$$

Therefore aggregation cannot increase the objective. The optimal value of the original problem is consequently no larger than that of the split problem.

To attain equality, take any feasible solution (Γ, R) of the original problem and define

$$\alpha = \frac{a_{i_1}}{a_i}.$$

Construct a feasible split solution by setting

$$\begin{aligned} \Gamma_{i_1 j} &= \alpha\Gamma_{ij}, \quad & \Gamma_{i_2 j} &= (1-\alpha)\Gamma_{ij}, \\ R_{i_1 j} &= R_{ij}, & R_{i_2 j} &= R_{ij}, \end{aligned}$$

This choice gives equality in Jensen's inequality and preserves both marginal constraints and the objective value. Therefore, the original and split problems have the same optimal value. □

A.7 Proof of Theorem 4.4

The endpoint distribution $q(\boldsymbol{z})$ is defined on $\mathcal{Z} = \mathcal{W} \times \mathcal{W}$, with $\boldsymbol{z} = (\boldsymbol{w}_0, \boldsymbol{w}_1)$. Write the conditional field as $\boldsymbol{s}_t(\boldsymbol{w} \mid \boldsymbol{z}) = (\boldsymbol{u}_t(\boldsymbol{w} \mid \boldsymbol{z}), g_t(\boldsymbol{w} \mid \boldsymbol{z}))$. The marginal path and fields used in Equation (4.9) and Theorem 4.4 are

$$\hat{\rho}_t(\boldsymbol{w}) = \int_{\mathcal{Z}} \hat{\rho}_t(\boldsymbol{w} \mid \boldsymbol{z}) q(\boldsymbol{z}) \,\mathrm{d}\boldsymbol{z}, \tag{A.27}$$

$$\boldsymbol{u}_t(\boldsymbol{w}) = \frac{\int_{\mathcal{Z}} \boldsymbol{u}_t(\boldsymbol{w} \mid \boldsymbol{z}) \hat{\rho}_t(\boldsymbol{w} \mid \boldsymbol{z}) q(\boldsymbol{z}) \,\mathrm{d}\boldsymbol{z}}{\hat{\rho}_t(\boldsymbol{w})}, \tag{A.28}$$

$$g_t(\boldsymbol{w}) = \frac{\int_{\mathcal{Z}} g_t(\boldsymbol{w} \mid \boldsymbol{z}) \hat{\rho}_t(\boldsymbol{w} \mid \boldsymbol{z}) q(\boldsymbol{z}) \,\mathrm{d}\boldsymbol{z}}{\hat{\rho}_t(\boldsymbol{w})}. \tag{A.29}$$

Proof of Theorem 4.4. We prove the claim for the state velocity component; the log mass growth component is identical with scalar products. The conditional velocity part of Equation (4.10) is

$$\mathcal{L}_{\boldsymbol{v}}^{\mathrm{cond}}(\theta) = \int_0^1 \int_{\mathcal{W}} \int_{\mathcal{Z}} \left\| \boldsymbol{v}_\theta(\boldsymbol{w}, t) - \boldsymbol{u}_t(\boldsymbol{w} \mid \boldsymbol{z}) \right\|^2 \hat{\rho}_t(\boldsymbol{w} \mid \boldsymbol{z}) q(\boldsymbol{z}) \,\mathrm{d}\boldsymbol{z} \,\mathrm{d}\boldsymbol{w} \,\mathrm{d}t.$$

Expanding the square gives

$$\begin{aligned} \mathcal{L}_{\boldsymbol{v}}^{\mathrm{cond}}(\theta) = \int_0^1 \int_{\mathcal{W}} \Bigg[& \left\| \boldsymbol{v}_\theta(\boldsymbol{w}, t) \right\|^2 \int_{\mathcal{Z}} \hat{\rho}_t(\boldsymbol{w} \mid \boldsymbol{z}) q(\boldsymbol{z}) \,\mathrm{d}\boldsymbol{z} \\ & - 2 \left\langle \boldsymbol{v}_\theta(\boldsymbol{w}, t), \int_{\mathcal{Z}} \boldsymbol{u}_t(\boldsymbol{w} \mid \boldsymbol{z}) \hat{\rho}_t(\boldsymbol{w} \mid \boldsymbol{z}) q(\boldsymbol{z}) \,\mathrm{d}\boldsymbol{z} \right\rangle \\ & + \int_{\mathcal{Z}} \left\| \boldsymbol{u}_t(\boldsymbol{w} \mid \boldsymbol{z}) \right\|^2 \hat{\rho}_t(\boldsymbol{w} \mid \boldsymbol{z}) q(\boldsymbol{z}) \,\mathrm{d}\boldsymbol{z} \Bigg] \,\mathrm{d}\boldsymbol{w} \,\mathrm{d}t. \end{aligned}$$

Expanding the marginal velocity loss in Equation (4.9) gives

$$\mathcal{L}_{\boldsymbol{v}}^{\mathrm{marg}}(\theta) = \int_0^1 \int_{\mathcal{W}} \Big[\left\| \boldsymbol{v}_\theta(\boldsymbol{w}, t) \right\|^2 - 2 \left\langle \boldsymbol{v}_\theta(\boldsymbol{w}, t), \boldsymbol{u}_t(\boldsymbol{w}) \right\rangle + \left\| \boldsymbol{u}_t(\boldsymbol{w}) \right\|^2 \Big] \hat{\rho}_t(\boldsymbol{w}) \,\mathrm{d}\boldsymbol{w} \,\mathrm{d}t,$$

By Equation (A.27), the first square term in the conditional expansion equals the corresponding term in the marginal expansion. By Equation (A.28),

$$\int_{\mathcal{Z}} \boldsymbol{u}_t(\boldsymbol{w} \mid \boldsymbol{z})\hat{\rho}_t(\boldsymbol{w} \mid \boldsymbol{z})q(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = \boldsymbol{u}_t(\boldsymbol{w})\hat{\rho}_t(\boldsymbol{w}),$$

so the cross term also matches. The remaining target-field square terms do not depend on θ. Therefore,

$$\mathcal{L}_{\boldsymbol{v}}^{\mathrm{cond}}(\theta) = \mathcal{L}_{\boldsymbol{v}}^{\mathrm{marg}}(\theta) + C_{\boldsymbol{v}}.$$

The same expansion applies to the scalar growth component g_ϕ, using Equation (A.29) in place of Equation (A.28). Hence the conditional and marginal growth losses differ only by a ϕ-independent constant. Since λ is fixed during training, the full losses satisfy

$$\mathcal{L}_{\mathrm{ESFM}}^{\mathrm{cond}}(\theta, \phi) = \mathcal{L}_{\mathrm{ESFM}}^{\mathrm{marg}}(\theta, \phi) + C,$$

with C independent of (θ, ϕ), and taking gradients with respect to both parameter sets gives the result. □

B Dataset Details and Additional Numerical Results

B.1 Experimental Details

Experiments were run on an NVIDIA A100 GPU cluster with 128 CPU cores or a laptop with an NVIDIA RTX 5060 Laptop GPU. We use two separate multilayer perceptrons for velocity and growth, with five and three hidden layers, respectively. Both take $(t, \boldsymbol{x}, y)$ as input and use LeakyReLU activations. The prediction experiments use 256 hidden units per layer, while held-out experiments use 512. The models are implemented in PyTorch (Paszke et al., 2017) and trained with Adam.

Coupling parameter δ. The parameter $\delta > 0$ weights the mass change cost relative to the state transport cost when solving the static coupling problem in Equation (4.8):

$$\begin{aligned} \inf_{\gamma_0,\gamma_1 \geq 0} \quad & \iint \left[\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|^2 + \delta^2 \Phi_M(R(\boldsymbol{x}_0, \boldsymbol{x}_1)) \right] \mathrm{d}\gamma_0, \\ \text{s.t.} \quad & \int \gamma_0(\boldsymbol{x}_0, \boldsymbol{x}_1)\,\mathrm{d}\boldsymbol{x}_1 = \rho_0(\boldsymbol{x}_0), \qquad \int \gamma_1(\boldsymbol{x}_0, \boldsymbol{x}_1)\,\mathrm{d}\boldsymbol{x}_0 = \rho_1(\boldsymbol{x}_1), \\ & 0 < R(\boldsymbol{x}_0, \boldsymbol{x}_1) \leq \mathrm{e}^M. \end{aligned}$$

Increasing δ penalizes growth or decay more strongly, favoring state transport when matching the endpoint distributions. Decreasing δ makes mass variation less costly. This balance determines the optimal coupling, including the endpoint pairing and terminal log mass allocation, and hence the conditional samples used for training.

Training weight λ. The parameter $\lambda > 0$ balances the velocity and growth regression errors in the conditional ESFM loss from Equation (4.10):

$$\begin{aligned} \mathcal{L}_{\mathrm{ESFM}}^{\mathrm{cond}}(\theta, \phi) = \int_0^1 \int_{\mathcal{Z}} \int_{\mathcal{W}} \Big[& \|\boldsymbol{v}_\theta(\boldsymbol{w}, t) - \boldsymbol{u}_t(\boldsymbol{w} \mid \boldsymbol{z})\|^2 \\ & + \lambda \left| g_\phi(\boldsymbol{w}, t) - g_t(\boldsymbol{w} \mid \boldsymbol{z}) \right|^2 \Big] \hat{\rho}_t(\boldsymbol{w} \mid \boldsymbol{z}) q(\boldsymbol{z})\,\mathrm{d}\boldsymbol{w}\,\mathrm{d}\boldsymbol{z}\,\mathrm{d}t. \end{aligned}$$

A larger λ gives greater weight to the growth regression error relative to the velocity error. This parameter acts only during training and does not change the transport action, coupling, or conditional targets.

Choice of M. The parameter M sets the upper bound on the local growth ratio through $R \leq \mathrm{e}^M$ and determines the mass cost function Φ_M in Equation (4.8), thereby affecting the computation of the optimal coupling.

Increasing M enlarges the feasible set of the local mass problem, so $\Phi_M(r)$ is nonincreasing in M for any fixed feasible growth ratio r. By Equation (A.12), the threshold $y_\star$ tends to zero as $M \to \infty$. For any fixed $r > 1$, the second branch of Φ_M in Theorem 4.3 applies for sufficiently large M, giving

$$\Phi_M(r) = \frac{2y_\star}{\mathrm{e}^{y_\star}} r + y_\star^2 - 2y_\star \longrightarrow 0 \qquad \text{as } M \to \infty.$$

A finite growth bound therefore limits individual particle growth and keeps the mass cost strictly positive for any fixed feasible growth ratio $r > 1$.

For two adjacent snapshots containing n_0 and n_1 cells, the observed average growth ratio is n_1/n_0. We choose the maximum allowed growth ratio as a multiple of this average:

$$\mathrm{e}^M = \kappa \frac{n_1}{n_0}, \qquad \kappa > 1.$$

Thus, the maximum growth ratio of a single particle is κ times the population average n_1/n_0. To maintain the assumption $M > 1$ used in the derivation, our implementation sets

$$M = \max\left\{1.05,\, \log\left(\kappa \frac{n_1}{n_0}\right)\right\}.$$

We use $\kappa = 2$ as the default choice for most datasets, which allows individual particles to grow by up to twice the observed average population growth ratio. This is an empirical choice that provides room for heterogeneous growth while limiting extreme mass allocation. The values used for individual datasets are reported below.

B.2 Evaluation Metrics

We use two metrics to evaluate prediction results: the 1-Wasserstein distance ($\mathcal{W}_1$) to measure the difference between normalized predicted and observed distributions, and relative mass error (RME) to assess the accuracy of total mass prediction. They are defined as

$$\mathcal{W}_1(p, q) = \min_{\pi \in \Pi(p,q)} \int \|\boldsymbol{x} - \boldsymbol{x}'\|_2 \, \mathrm{d}\pi(\boldsymbol{x}, \boldsymbol{x}'),$$

$$\mathrm{RME}(t_k) = \frac{|\sum_i m_i(t_k) - n_k|}{n_k}.$$

Here p and q are the normalized predicted and observed empirical distributions, and $\Pi(p, q)$ is the set of couplings between them. The target cell count at time t_k is denoted by n_k, and $m_i(t_k)$ is the predicted mass of particle i at that time. Lower values of both metrics indicate better performance. We report RME only for methods that model mass variation.

For ESFM, each prediction starts from the previous observed snapshot with $m_i = 1$, corresponding to $y_i = 0$, consistent with the source condition of the interval coupling. We solve the learned ODE in the extended space to obtain the predicted cell states $\boldsymbol{x}_i(t_k)$ and log masses $y_i(t_k)$. The prediction in the original state space is the weighted particle distribution $\sum_i \mathrm{e}^{y_i(t_k)} \delta_{\boldsymbol{x}_i(t_k)}$: each particle contributes at its predicted cell state with mass $m_i(t_k) = \mathrm{e}^{y_i(t_k)}$. This is the weighted projection of the extended distribution. We compute $\mathcal{W}_1$ between its normalized form and the normalized observed distribution using POT, with distances measured in the cell state space. We compute RME by comparing its total mass $\sum_i m_i(t_k)$ with the observed cell count n_k. For the held-out experiments, we exclude one intermediate snapshot from training and predict it from the previous observed snapshot. We report the mean $\mathcal{W}_1$ over the held-out time points.

To compare ESFM with other methods, we run the available reference implementations with method and dataset specific configurations. For WFR-FM (Peng et al., 2026b), VGFM (Wang et al., 2025) and USB (Ying et al., 2026a), we use their reference configurations as the starting point and retain their respective network architectures and training procedures. For MMFM (Rohbeck et al., 2025), Metric FM (Kapusniak et al., 2024), SF2M (Tong et al., 2024b) and UOT-FM (Eyring et al., 2024), we use the corresponding released model and matching implementations, with learning rates, training budgets and sample sizes set separately for each method and dataset. Cosine annealing is used in WFR-FM, USB and MMFM training scripts. For MIOFlow (Huguet et al., 2022), TIGON (Sha et al.,

Table 4: Sensitivity analysis of δ, λ and κ on the Simulation Gene dataset.

Parameter	**Value**	$\mathcal{W}_1$ (↓)	RME (↓)
δ	1.0	0.027	0.008
δ	2.0	0.018	0.001
δ	3.0	0.024	0.013
κ	1.2	0.018	0.001
κ	1.5	0.018	0.001
κ	3.0	0.018	0.001
λ	1.0	0.018	0.001
λ	4.0	0.018	0.001
λ	8.0	0.017	0.001

2024), DeepRUOT (Zhang et al., 2025a) and Var-RUOT (Sun et al., 2025), we retain the method specific optimization procedures, including local and global training in MIOFlow, density estimation in TIGON and the staged training pipeline in DeepRUOT. For methods with stochastic inference, we report the mean and standard deviation over five inference runs using a fixed trained model.

B.3 Synthetic Benchmarks

We evaluate ESFM on Simulation Gene, Saturating-Growth Data, Dyngen and a Gaussian mixture benchmark. The aggregate prediction results are reported in Table 2; dataset descriptions and additional results are provided below. In the Dyngen prediction table, entries with $\pm$ report the mean and standard deviation over five inference runs with different random seeds using a fixed trained model.

Simulation Gene. The Simulation Gene dataset is generated from a synthetic gene regulatory network (Zhang et al., 2025a). The expression levels of three genes evolve according to

$$\begin{aligned}
\frac{\mathrm{d}X_1}{\mathrm{d}t} &= \frac{\alpha_1 X_1^2 + \beta}{1 + \alpha_1 X_1^2 + \gamma_2 X_2^2 + \gamma_3 X_3^2 + \beta} - \delta_1 X_1 + \eta_1 \xi_{1,t}, \\
\frac{\mathrm{d}X_2}{\mathrm{d}t} &= \frac{\alpha_2 X_2^2 + \beta}{1 + \gamma_1 X_1^2 + \alpha_2 X_2^2 + \gamma_3 X_3^2 + \beta} - \delta_2 X_2 + \eta_2 \xi_{2,t}, \\
\frac{\mathrm{d}X_3}{\mathrm{d}t} &= \frac{\alpha_3 X_3^2}{1 + \alpha_3 X_3^2} - \delta_3 X_3 + \eta_3 \xi_{3,t}.
\end{aligned} \tag{B.1}$$

Here $X_i(t)$ is the expression level of gene i. Genes X_1 and X_2 form a toggle switch: each gene activates itself and inhibits the other. The external signal β activates both genes, while X_3 inhibits them. The parameter α_i controls self-activation, γ_i controls inhibition and δ_i controls degradation. The term $\eta_i \xi_{i,t}$ represents random noise with strength η_i. Cells divide randomly, with a growth rate determined by X_2:

$$g(\boldsymbol{X}) = \alpha_g \frac{X_2^2}{1 + X_2^2},$$

where α_g controls the growth scale. After division, daughter cells inherit the parent's gene expression state with small random perturbations. The simulation starts from two cell populations: one changes its state and proliferates, while the other remains a steady state. This setting tests whether a model can distinguish population growth from cell state transitions. Snapshots are collected at times 0, 8, 16, 24 and 32. We use X_1 and X_2 as the model inputs and set the hyperparameters to $\delta = 2.0$, $\lambda = 4.0$ and $\kappa = 2.0$. The sensitivity analysis is reported in Table 4. The results show that δ affects prediction performance, whereas the tested values of κ and λ give similar results.

Saturating-Growth Data. This variant follows the same gene regulatory dynamics as Simulation Gene in Equation (B.1), but replaces the growth law with

$$g(\boldsymbol{X}) = \alpha_g \frac{X_2^8}{0.2^8 + X_2^8},$$

where α_g denotes the maximum growth rate. The growth rate increases with X_2 and then stays close to α_g. One population grows at a nearly constant rate, while the other shows almost no growth. This experiment tests whether ESFM can identify which cells grow and accurately predict their growth rates. Snapshots are collected at times 0, 4, 8, 16 and 32. We retain all three gene coordinates for training and use PCA for visualization. We set the hyperparameters to $\delta = 2.0$, $\lambda = 1.0$ and $\kappa = 2.0$. To compute the coefficient of determination R^2 reported in Q1, we pool all $N = 2772$ observed cells across the five snapshots and evaluate each cell at its own observation time t_i. Let $(\boldsymbol{X}_{j,k}, Y_{j,k})$ denote the resulting particles at t_k. At the initial time, the particles are the observed source cells with $Y_{j,0} = 0$.

Physical growth and its particle estimate. The projection in Theorem 4.1 defines the physical density and growth field as

$$\rho_t(\boldsymbol{x}) = \int_{\mathbb{R}} \mathrm{e}^y \hat{\rho}_t(\boldsymbol{x}, y)\, \mathrm{d}y,$$
$$g_t(\boldsymbol{x}) = \frac{\int_{\mathbb{R}} \hat{g}_t(\boldsymbol{x}, y)\hat{\rho}_t(\boldsymbol{x}, y)\mathrm{e}^y\, \mathrm{d}y}{\rho_t(\boldsymbol{x})}.$$

The ODE provides particles $\{(\boldsymbol{X}_j(t), Y_j(t))\}_{j=1}^n$ whose states generally do not coincide with the observed cells. To evaluate growth at an observed state $\boldsymbol{x}$, we use nearby particles to form the local kernel estimate

$$\hat{\rho}_t^h(\boldsymbol{x}, \mathrm{d}y) = \frac{1}{n}\sum_{j=1}^n K_h(\boldsymbol{x} - \boldsymbol{X}_j(t))\delta_{Y_j(t)}(\mathrm{d}y), \qquad K_h(\boldsymbol{z}) = \exp\left(-\frac{\|\boldsymbol{z}\|^2}{2h^2}\right).$$

The kernel acts only on the state coordinates; common normalization factors cancel in the ratio. Substituting this local measure into the projection and using the learned growth field g_ϕ gives

$$\widehat{g}_t^h(\boldsymbol{x}) := \frac{\int_{\mathbb{R}} g_\phi(\boldsymbol{x}, y, t)\mathrm{e}^y \hat{\rho}_t^h(\boldsymbol{x}, \mathrm{d}y)}{\int_{\mathbb{R}} \mathrm{e}^y \hat{\rho}_t^h(\boldsymbol{x}, \mathrm{d}y)}$$
$$= \frac{\sum_j K_h(\boldsymbol{x} - \boldsymbol{X}_j(t))\mathrm{e}^{Y_j(t)} g_\phi(\boldsymbol{x}, Y_j(t), t)}{\sum_j K_h(\boldsymbol{x} - \boldsymbol{X}_j(t))\mathrm{e}^{Y_j(t)}}.$$

We set h to the distance to the 16th nearest generated state, without using growth labels, and include all particles in the estimate. WFR-FM is evaluated directly through its physical growth head at $(\boldsymbol{x}, t_k)$. Let $g_i = g(\boldsymbol{X}_i)$ be its true growth rate and $\widehat{g}_i$ its predicted rate. We compute

$$R^2 = 1 - \frac{\sum_{i=1}^N (\widehat{g}_i - g_i)^2}{\sum_{i=1}^N (g_i - \bar{g})^2}, \qquad \bar{g} = \frac{1}{N}\sum_{i=1}^N g_i.$$

As shown in Figure 2, ESFM predicts growth rates close to the true values for both populations. WFR-FM predicts lower growth rates, including negative values for some cells even though their true growth rates are nonnegative.

Dyngen. The Dyngen dataset is generated using the dyngen single-cell simulator (Cannoodt et al., 2021) and contains 728 cells across five snapshots. The simulated trajectories bifurcate into two branches with unequal cell abundance, providing a test of joint state and mass prediction. We use a five-dimensional PHATE representation (Moon et al., 2019) and set the hyperparameters to $\delta = 2.5$, $\lambda = 6.25$ and $\kappa = 1.5$. The prediction results in Table 5 show that ESFM achieves the lowest distribution and mass errors at most prediction times.

Table 5: Comparison of method performance over time on the Dyngen dataset.

Method	$t=1$		$t=2$		$t=3$		$t=4$	
	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME
MMFM	0.574	—	1.704	—	1.499	—	1.706	—
Metric FM	0.892	—	2.347	—	2.030	—	1.799	—
SF2M	0.637	—	1.266	—	1.415	—	1.790	—
MIOFlow	0.420	—	0.640	—	1.537	—	1.263	—
TIGON	0.446	0.033	0.584	0.060	0.415	0.023	0.603	0.071
DeepRUOT	0.454	0.011	0.481	0.070	0.870	0.104	0.688	0.074
Var-RUOT	0.315	0.128	0.548	0.336	0.630	0.222	0.593	0.023
UOT-FM	0.652	0.008	0.780	0.077	1.252	0.090	2.130	0.213
VGFM	0.335	0.001	0.312	0.073	1.109	0.041	0.634	0.033
WFR-FM	0.110	0.003	0.098	0.007	0.211	0.008	0.121	0.002
USB	**0.109**$_{\pm 0.0002}$	0.002$_{\pm 0.0000}$	**0.093**$_{\pm 0.0002}$	**0.002**$_{\pm 0.0001}$	0.180$_{\pm 0.0013}$	0.015$_{\pm 0.0004}$	0.142$_{\pm 0.0040}$	0.008$_{\pm 0.0008}$
ESFM (ours)	0.135	**0.0002**	**0.093**	0.008	**0.132**	**0.0028**	**0.085**	**0.0003**

Gaussian mixture. The Gaussian mixture benchmark contains two snapshots in 1000 dimensions (Wang et al., 2025). The source contains 500 samples, with 100 in an upper component and 400 in a lower component. In the target, the upper component expands to 1000 samples, while the lower population splits into two components of 200 samples each. The benchmark therefore combines population growth with branching transport in a high dimensional state space. We set the hyperparameters to $\delta = 1.2$, $\lambda = 1.44$ and $\kappa = 4.0$. As shown in Figure 4, the predicted endpoints capture the three target modes, with little movement in the upper population and splitting trajectories in the lower population.

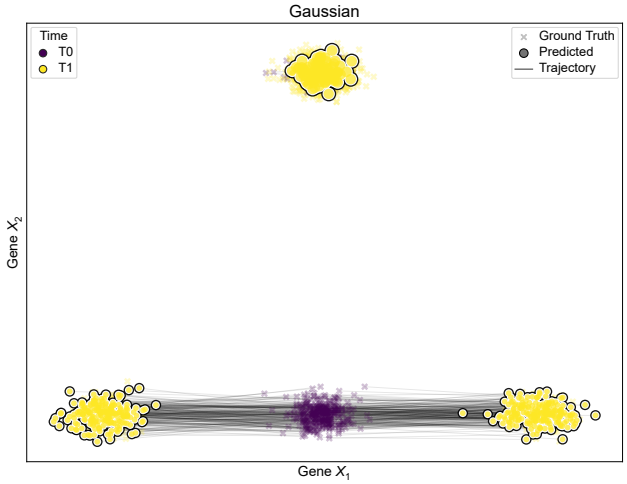


Figure 4: ESFM trajectory prediction on the 1000D Gaussian mixture benchmark.

B.4 Real Single-Cell Benchmarks

We evaluate ESFM on four real single-cell datasets: EMT, embryoid body differentiation (EB), CITE-seq, and mouse hematopoiesis. The aggregate held-out results are shown in Table 3. Detailed dataset descriptions and prediction results are provided below. For the prediction tables below, entries with $\pm$ report the mean and standard deviation over five inference runs with different random seeds using a fixed trained model.

EMT. The epithelial–mesenchymal transition (EMT) dataset contains A549 lung cancer cells collected at four time points during the transition (Cook & Vanderhyden, 2020). We use an autoencoder to reduce the gene expression data to 10 dimensions. These snapshots capture changes in both transcriptional state and observed cell abundance. We evaluate ESFM on this dataset with hyperparameters set to $\delta = 1.5$, $\lambda = 2.25$, and $\kappa = 2.0$. As shown in Table 6, ESFM achieves the lowest distribution error at $t = 2, 3$ and the second-lowest at $t = 1$, while maintaining low mass error.

Table 6: Comparison of method performance over time on the 10D EMT dataset.

Method	$t = 1$		$t = 2$		$t = 3$	
	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME
MMFM	0.258	—	0.287	—	0.310	—
Metric FM	0.261	—	0.297	—	0.305	—
SF2M	0.257	—	0.281	—	0.290	—
MIOFlow	0.244	—	0.267	—	0.284	—
TIGON	0.243	0.002	0.266	0.003	0.285	0.001
DeepRUOT	0.290	0.001	0.319	0.011	0.329	0.002
Var-RUOT	0.254	0.075	0.267	0.014	0.268	0.041
UOT-FM	0.254	0.002	0.270	0.013	0.277	0.010
VGFM	0.235	0.016	0.242	0.011	0.245	0.018
WFR-FM	0.210	0.001	0.227	0.002	0.235	0.001
USB	$\mathbf{0.183}_{\pm 0.000}$	$\mathbf{0.000}_{\pm 0.000}$	$0.216_{\pm 0.000}$	$0.009_{\pm 0.000}$	$0.231_{\pm 0.000}$	$0.009_{\pm 0.000}$
ESFM (ours)	0.197	**0.000**	**0.189**	**0.001**	**0.198**	**0.000**

Embryoid body. The human embryoid body (EB) dataset contains 16,819 cells sampled at five time points over a 27-day differentiation experiment (Moon et al., 2019). This experimental system models early embryonic development and captures the emergence of diverse cell states. We evaluate PCA representations with 5, 50 and 100 dimensions to examine prediction performance at different dimensions. The hyperparameter settings are $\delta = 2.0$, $\lambda = 1.0$ and $\kappa = 2.0$ for 5D; $\delta = 24.0$, $\lambda = 5.0$ and $\kappa = 2.0$ for 50D; and $\delta = 18.0$, $\lambda = 3.0$ and $\kappa = 2.0$ for 100D. The prediction results in Tables 7 to 9 show that ESFM achieves the lowest $\mathcal{W}_1$ across all tested dimensions and prediction times. It also achieves the lowest mass error at three of the four prediction times in the 5D representation.

Table 7: Comparison of method performance over time on the 5D EB dataset.

Method	$t = 1$		$t = 2$		$t = 3$		$t = 4$	
	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME
MMFM	0.477	—	0.554	—	0.781	—	0.872	—
Metric FM	0.449	—	0.552	—	0.583	—	0.597	—
SF2M	0.556	—	0.715	—	0.750	—	0.650	—
MIOFlow	0.442	—	0.585	—	0.651	—	0.670	—
TIGON	0.386	**0.002**	0.502	0.015	0.602	0.021	0.600	0.027
DeepRUOT	0.386	0.005	0.497	0.017	0.591	0.021	0.585	0.030
Var-RUOT	0.416	0.111	0.486	0.144	0.509	0.054	0.511	0.022
UOT-FM	0.544	0.032	0.670	0.029	0.729	0.016	0.852	0.041
VGFM	0.402	0.046	0.494	0.018	0.525	0.035	0.573	0.021
WFR-FM	0.324	0.003	0.401	0.001	0.431	0.005	0.510	0.005
USB	$0.331_{\pm 2 \times 10^{-5}}$	$0.006_{\pm 2 \times 10^{-6}}$	$0.408_{\pm 2 \times 10^{-5}}$	$0.023_{\pm 4 \times 10^{-6}}$	$0.427_{\pm 2 \times 10^{-5}}$	$0.018_{\pm 6 \times 10^{-6}}$	$0.450_{\pm 4 \times 10^{-5}}$	$0.019_{\pm 1 \times 10^{-5}}$
ESFM (ours)	**0.313**	0.005	**0.310**	**0.000**	**0.330**	**0.003**	**0.343**	**0.001**

Table 8: Comparison of method performance over time on the 50D EB dataset.

Method	$t = 1$		$t = 2$		$t = 3$		$t = 4$	
	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME
MMFM	9.124	—	10.474	—	11.022	—	11.480	—
Metric FM	8.506	—	9.795	—	10.621	—	12.042	—
SF2M	9.247	—	10.882	—	11.650	—	12.154	—
MIOFlow	8.447	—	9.229	—	9.436	—	10.123	—
TIGON	8.433	0.067	9.275	0.022	9.802	0.179	10.148	0.101
DeepRUOT	8.169	**0.003**	9.049	0.038	9.378	0.088	9.733	**0.004**
Var-RUOT	9.442	0.128	9.709	0.081	10.482	0.031	10.735	0.030
UOT-FM	8.717	0.063	10.858	0.009	11.813	0.022	12.733	0.018
VGFM	7.951	0.089	8.747	0.042	9.244	0.019	9.620	0.044
WFR-FM	7.664	0.008	8.659	**0.006**	9.182	**0.004**	9.914	**0.004**
USB	$7.992_{\pm 2 \times 10^{-5}}$	$0.008_{\pm 7 \times 10^{-7}}$	$8.973_{\pm 5 \times 10^{-5}}$	$0.019_{\pm 1 \times 10^{-6}}$	$9.421_{\pm 2 \times 10^{-5}}$	$0.008_{\pm 2 \times 10^{-6}}$	$9.963_{\pm 2 \times 10^{-5}}$	$0.017_{\pm 3 \times 10^{-6}}$
ESFM (ours)	**7.419**	0.013	**7.634**	0.040	**7.716**	0.007	**7.918**	0.015

Table 9: Comparison of method performance over time on the 100D EB dataset.

Method	$t=1$		$t=2$		$t=3$		$t=4$	
	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME
MMFM	11.460	—	13.879	—	14.441	—	14.907	—
Metric FM	10.806	—	12.348	—	13.622	—	16.801	—
SF2M	11.333	—	12.982	—	13.718	—	14.945	—
MIOFlow	11.387	—	12.331	—	11.905	—	12.908	—
TIGON	10.547	0.014	12.926	0.052	13.897	0.107	14.945	0.096
DeepRUOT	10.256	**0.002**	11.103	0.074	11.529	0.136	12.406	0.047
Var-RUOT	11.746	0.091	12.237	0.024	12.957	0.150	13.335	0.074
UOT-FM	10.757	0.056	12.799	0.037	13.761	0.044	15.657	0.022
VGFM	10.313	0.048	11.278	0.035	11.703	0.028	12.637	0.066
WFR-FM	9.941	0.009	11.040	**0.006**	11.516	0.008	12.664	0.005
USB	$10.206_{\pm 5\times10^{-5}}$	$0.016_{\pm 5\times10^{-7}}$	$11.297_{\pm 6\times10^{-5}}$	$0.010_{\pm 8\times10^{-7}}$	$11.758_{\pm 9\times10^{-6}}$	$\mathbf{0.002}_{\pm 2\times10^{-6}}$	$12.888_{\pm 5\times10^{-5}}$	$\mathbf{0.003}_{\pm 3\times10^{-6}}$
ESFM (ours)	**9.825**	0.020	**10.728**	0.014	**11.130**	0.009	**11.768**	0.005

CITE-seq. The CITE-seq dataset contains 31,240 cells collected at four time points (Lance et al., 2022). Although the original data include both gene expression and surface protein measurements, our experiments use only the gene expression matrix, reduced to 50 dimensions by PCA. The large number of cells and variation in snapshot size provide a test of prediction on unbalanced single-cell data. We set the hyperparameters to $\delta = 1.5$, $\lambda = 1.0$ and $\kappa = 3.0$. As shown in Table 10, ESFM achieves the lowest distribution error at $t = 2, 3$ and the third-lowest at $t = 1$, although some baselines attain lower mass errors.

Table 10: Comparison of method performance over time on the 50D CITE dataset.

Method	$t=1$		$t=2$		$t=3$	
	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME
MMFM	33.971	—	36.854	—	43.721	—
Metric FM	28.314	—	28.617	—	33.212	—
SF2M	29.543	—	32.655	—	36.265	—
MIOFlow	28.290	—	28.524	—	32.230	—
TIGON	28.196	0.186	27.921	0.545	32.846	0.653
DeepRUOT	28.245	0.168	27.908	0.525	32.950	0.634
Var-RUOT	30.219	0.331	32.702	0.325	40.613	0.486
UOT-FM	33.531	0.009	32.795	0.046	49.751	0.097
VGFM	29.449	0.020	29.722	0.057	33.752	**0.001**
WFR-FM	27.831	0.043	27.478	0.045	34.784	0.022
USB	$\mathbf{27.021}_{\pm 5\times10^{-5}}$	$\mathbf{0.002}_{\pm 1\times10^{-6}}$	$28.081_{\pm 1\times10^{-4}}$	$\mathbf{0.015}_{\pm 2\times10^{-6}}$	$38.447_{\pm 0.003}$	$0.010_{\pm 5\times10^{-5}}$
ESFM (ours)	27.860	0.055	**26.609**	0.095	**28.813**	0.094

Mouse hematopoiesis. The mouse hematopoiesis dataset contains 49,302 lineage-traced cells collected at three time points (Weinreb et al., 2020). Its large cell population and substantial changes in cell abundance make it useful for evaluating both prediction and the computational cost of coupling construction. We use PCA to reduce the gene expression data to 50 dimensions, and set the hyperparameters to $\delta = 15$, $\lambda = 225$ and $\kappa = 1.5$, using mini-batch coupling with a mini-batch size of 2000. The prediction results in Table 11 show that ESFM achieves the lowest $\mathcal{W}_1$ and RME at $t = 2$, while WFR-FM and USB attain lower distribution errors at $t = 1$.

Table 11: Comparison of method performance over time on the 50D Mouse dataset.

Method	$t = 1$		$t = 2$	
	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME
MMFM	7.647	—	10.156	—
Metric FM	7.788	—	11.449	—
SF2M	8.217	—	11.086	—
MIOFlow	6.313	—	6.746	—
TIGON	6.140	0.382	6.973	0.326
DeepRUOT	6.052	0.062	6.757	0.041
Var-RUOT	7.951	0.131	10.862	0.154
UOT-FM	8.114	0.035	9.170	0.011
VGFM	6.274	0.076	6.796	0.070
WFR-FM	**5.486**	**0.012**	<u>6.211</u>	0.011
USB	<u>5.589</u>$_{\pm 1 \times 10^{-6}}$	<u>0.013</u>$_{\pm 6 \times 10^{-8}}$	6.548$_{\pm 4 \times 10^{-6}}$	<u>0.008</u>$_{\pm 4 \times 10^{-7}}$
ESFM (ours)	5.604	0.073	**5.539**	**0.004**

Sensitivity to mini-batch coupling size. We examine the effect of coupling mini-batch size on the Mouse dataset in Table 12. Mini-batch coupling greatly reduces runtime compared with full coupling while maintaining comparable prediction performance. The mini-batch sizes reported in the table yield similar runtime and comparable accuracy in both distribution and mass prediction. Full coupling yields slightly lower distribution error but requires much more computation. These results support mini-batch coupling as an efficient approximation, with the selected batch size providing a balance between prediction quality and computational cost.

Table 12: Sensitivity analysis for the batch size of mini-batch coupling on the 50D Mouse dataset.

Batch Size	$t = 1$		$t = 2$		**Time (s)**
	$\mathcal{W}_1$	RME	$\mathcal{W}_1$	RME	
1000	5.712	0.113	5.583	0.0003	444
1500	5.618	0.076	5.538	0.0007	407
2000	5.604	0.073	5.539	0.004	389
2500	5.597	0.084	5.522	0.003	386
3000	5.611	0.116	5.520	0.003	394
4000	5.560	0.097	5.518	0.002	407
w/o mini-batch	5.483	0.101	5.459	0.001	5570

B.5 Efficiency on the 100D EB dataset

We evaluate the computational efficiency of ESFM on the 100-dimensional embryoid body dataset (EB100) by comparing prediction accuracy, runtime and memory. As shown in Figure 5, ESFM achieves the lowest mean $\mathcal{W}_1$ among the methods in the comparison, reaching $\mathcal{W}_1 = 10.863$ in only 42 seconds with 0.4 GB of memory. This runtime includes coupling computation, endpoint pair presampling and network training. ESFM also has the shortest reported runtime and lowest reported memory usage in the figure, demonstrating an efficient combination of accurate prediction and low computational cost on this large scale dataset. Detailed prediction results, including RME, are provided in Table 9.

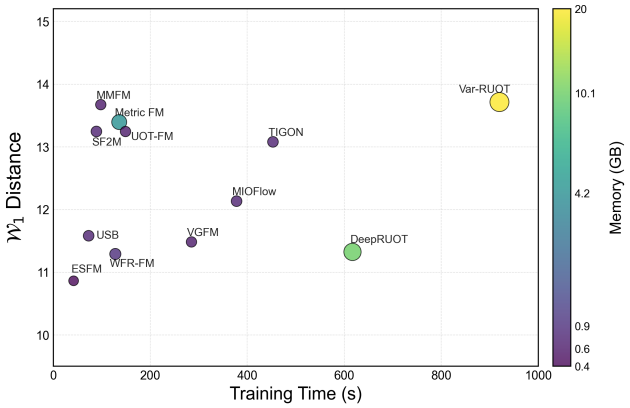


Figure 5: Efficiency comparison on the 100D EB dataset.

B.6 Evaluation of Static and Learned Dynamic Actions

Static action. For an adjacent interval $[t_k, t_{k+1}]$ of duration $T_k = t_{k+1} - t_k$ with n_k source particles, let $\Gamma_{ij}^{(k)}$ and $R_{ij}^{(k)}$ denote the source coupling weight and endpoint mass ratio returned by the entropy-regularized static solver. We compute the geometric cost without the entropy term:

$$A_{\text{stat}}^{(k)} = \frac{1}{n_k T_k} \sum_{i,j} \Gamma_{ij}^{(k)} \left[\|\boldsymbol{x}_j^{k+1} - \boldsymbol{x}_i^k\|^2 + \delta^2 \Phi_M(R_{ij}^{(k)}) \right].$$

Network dynamic action. We initialize every observed source cell at $\boldsymbol{w}_i(t_k) = (\boldsymbol{x}_i^k, 0)$ and integrate $\dot{\boldsymbol{w}}_i(t) = (\boldsymbol{v}_\theta(\boldsymbol{w}_i(t), t), g_\phi(\boldsymbol{w}_i(t), t))$. The action is evaluated along these network-generated trajectories:

$$A_{\text{dyn}}^{(k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} \int_{t_k}^{t_{k+1}} \left[\|\boldsymbol{v}_\theta(\boldsymbol{w}_i(t), t)\|^2 + \delta^2 |g_\phi(\boldsymbol{w}_i(t), t)|^2 \right] \mathrm{d}t.$$

We use forward Euler with 400 steps per interval.

Intervals and numerical checks. For each dataset, we compute the static and dynamic actions separately within each interval between observed time points. The reported actions are obtained by summing over all these intervals:

$$A_{\text{stat}} = \sum_k A_{\text{stat}}^{(k)}, \qquad A_{\text{dyn}} = \sum_k A_{\text{dyn}}^{(k)}, \qquad \text{Gap} = \frac{|A_{\text{dyn}} - A_{\text{stat}}|}{A_{\text{stat}}} \times 100\%.$$

C Relation to Other Unbalanced Flow Matching Methods

C.1 Relation to VGFM

VGFM (Wang et al., 2025) first defines an endpoint coupling by solving a semi-relaxed OT problem in the original state space:

$$\pi^\star \in \arg\min_{\pi \geq 0} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\boldsymbol{x}_0 - \boldsymbol{x}_1\|^2 \, \mathrm{d}\pi(\boldsymbol{x}_0, \boldsymbol{x}_1) + \mathrm{KL}\big(P_\#^0 \pi \,\big\|\, \rho_0\big), \qquad \text{s.t.} \quad P_\#^1 \pi = \rho_1. \tag{C.1}$$

Only the target marginal is constrained exactly. The relaxed source marginal determines the source-wise growth ratio $R(\boldsymbol{x}_0) = \frac{\mathrm{d}P_\#^0 \pi^\star}{\mathrm{d}\rho_0}(\boldsymbol{x}_0)$.

Conditioning on an endpoint pair $(\boldsymbol{x}_0, \boldsymbol{x}_1)$ selected by $\pi^\star$ and taking the initial particle mass to be one, VGFM then prescribes

$$\begin{aligned} \boldsymbol{x}_t &= (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1, \qquad m_t = R(\boldsymbol{x}_0)^t, \\ \boldsymbol{u}_t^\star &= \boldsymbol{x}_1 - \boldsymbol{x}_0, \qquad g_t^\star = \frac{\mathrm{d}\ln m_t}{\mathrm{d}t} = \ln R(\boldsymbol{x}_0). \end{aligned} \tag{C.2}$$

Thus, both the conditional state velocity and growth rate are constant. Once the endpoint masses are fixed, this is the same straight state and log-mass interpolation used by ESFM. VGFM trains $v_\theta(\boldsymbol{x}, t)$ and $g_\omega(\boldsymbol{x}, t)$ to regress these two targets.

VGFM relates Equation (C.1) to a dynamic process with two stages: mass changes first without transport, and transport then occurs without further mass change. In particular, VGFM does not establish that Equations (C.1) and (C.2) arise from a single simultaneous dynamic action. Its final training procedure also adds a Wasserstein loss between generated and observed distributions. Evaluating this loss requires Neural ODE rollouts of positions and weights, so the complete algorithm is not fully simulation free during training. By contrast, ESFM uses the same straight conditional interpolation as VGFM, but defines it as a path in the extended space. Starting from this path, ESFM constructs a dynamic action for which the path is a geodesic, formulates the resulting dynamic OT problem, and derives the corresponding static OT problem and endpoint coupling. This provides a complete derivation from the prescribed path to the action, coupling, and training targets. Consequently, the ESFM conditional path and coupling are consistent.

C.2 Relation to WFR-FM

WFR-FM (Peng et al., 2026b) is designed as a simulation-free flow matching solver for the dynamic Wasserstein–Fisher–Rao optimal transport problem:

$$\begin{aligned} \mathrm{WFR}_\delta^2(\rho_0, \rho_1) := \inf_{\rho_t, \boldsymbol{u}_t, g_t} \quad & \int_0^1 \int_{\mathbb{R}^d} \left(\|\boldsymbol{u}_t(\boldsymbol{x})\|^2 + \delta^2 \, |g_t(\boldsymbol{x})|^2 \right) \rho_t(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \, \mathrm{d}t, \\ \text{s.t.} \quad & \partial_t \rho_t + \nabla_{\boldsymbol{x}} \cdot (\rho_t \boldsymbol{u}_t) = \rho_t g_t, \qquad \rho_{t=0} = \rho_0, \quad \rho_{t=1} = \rho_1. \end{aligned} \tag{C.3}$$

WFR-FM builds its tractable solver from the traveling-Dirac geodesic between two Dirac measures. This geodesic provides the conditional state and mass paths together with their velocity and growth targets. The equivalence between the dynamic and static WFR formulations provides an OET coupling that can be computed with a Sinkhorn solver. Converting the OET coupling into a WFR semi-coupling specifies the paired endpoints and their masses. WFR-FM samples from this semi-coupling and trains its velocity and growth networks with a conditional regression loss weighted by mass.

ESFM reverses the usual design: it specifies a single-particle path first and then derives its action and coupling. The ESFM path enforces monotonic mass interpolation between its endpoints. This can better align with applications in which monotonic population change over an interval is an appropriate prior, whereas WFR geodesics may exhibit nonmonotonic mass evolution. More generally, choosing the WFR Dirac path within the extended space framework recovers the standard dynamic and static WFR formulations; WFR-FM is therefore a special case of the generalized ESFM framework. A detailed derivation is provided in Appendix D.4.

C.3 Relation to USB

Like ESFM, USB (Ying et al., 2026a) learns a growth rate and avoids ODE or SDE trajectory simulation during training. It models the dynamics as a stochastic process:

$$\mathrm{d}\boldsymbol{x}_t = \boldsymbol{u}_t(\boldsymbol{x}_t) \, \mathrm{d}t + \nu \, \mathrm{d}\boldsymbol{W}_t, \qquad \mathrm{d} \ln m_t = g_t(\boldsymbol{x}_t) \, \mathrm{d}t.$$

This model connects USB to the Branching Schrödinger Bridge (BSB) problem. Following Baradat & Lavenant (2025), USB relates BSB to a dynamic RUOT relaxation. Because the corresponding RUOT semi-coupling is not available in a tractable static form, its practical solver approximates the RUOT problem by WFR and computes a WFR/OET semi-coupling (Peng et al., 2026b).

For a sampled endpoint pair $z = (\boldsymbol{x}_0, \boldsymbol{x}_1) \sim \gamma_0$, USB sets $m_0(z) = 1$ and $m_1(z) = \gamma_1(z)/\gamma_0(z)$. It then prescribes the Poisson–Brownian conditional measure path

$$\rho_t(\boldsymbol{x} \mid z) = m_0(z)^{1-t} m_1(z)^t \mathcal{N}\big(\boldsymbol{x} \,\big|\, (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1, \nu^2 t(1-t) I\big) \,,$$

The state component follows a Brownian bridge with diffusion amplitude ν, and the mass varies linearly on the log scale. USB uses neural networks to learn the probability flow velocity $v_\theta(\boldsymbol{x}, t)$, growth rate $g_\theta(\boldsymbol{x}, t)$, and score $s_\theta(\boldsymbol{x}, t)$ through an objective for conditional unbalanced score matching.

The approximation of RUOT by WFR means that USB does not solve the original BSB problem exactly. However, USB uses a WFR semi-coupling together with a Poisson–Brownian conditional path, so coupling–path consistency is not established. By contrast, ESFM does not begin with a preselected OT functional. It first specifies the desired path in the extended state space, constructs an action for which this path is a geodesic, and then derives the corresponding dynamic problem, static OT formulation, and endpoint coupling. Consequently, the ESFM conditional path and coupling are consistent by construction.

D General Lagrangians and Alternative Geodesics

The extended-space formulation also supports other choices of action and conditional path. We present an action-first construction: starting from a Lagrangian, we obtain minimizing paths, the induced endpoint cost and coupling, and the corresponding FM targets. This complements the path-first design of the specific ESFM action in the main text. Appendix D.1 gives the general construction and explains how to obtain minimizing paths. Appendix D.2 considers an action whose geodesics are linear in physical mass, while Appendix D.4 recovers WFR formulation.

D.1 From a Lagrangian to minimizing paths

Single-particle action and paths. Let $\boldsymbol{w} = (\boldsymbol{x}, y) \in \mathcal{W} = \mathbb{R}^d \times \mathbb{R}$, with $y = \ln m$. We start from a nonnegative Lagrangian $L(\boldsymbol{w}, \boldsymbol{v})$, which assigns a cost to the extended velocity $\boldsymbol{v}$ at state $\boldsymbol{w}$. For an absolutely continuous path, define

$$\mathcal{A}_L(\boldsymbol{w}) = \int_0^1 L(\boldsymbol{w}_t, \dot{\boldsymbol{w}}_t)\,\mathrm{d}t, \qquad c_L(\boldsymbol{w}_0, \boldsymbol{w}_1) = \inf_{\boldsymbol{w}:\, \boldsymbol{w}_0 \to \boldsymbol{w}_1} \mathcal{A}_L(\boldsymbol{w}).$$

The conditional path is obtained by minimizing this action with fixed endpoints:

$$\Psi(\boldsymbol{w}_0, \boldsymbol{w}_1) \in \operatorname*{arg\,min}_{\boldsymbol{w}:\, \boldsymbol{w}_0 \to \boldsymbol{w}_1} \mathcal{A}_L(\boldsymbol{w}), \qquad c_L(\boldsymbol{w}_0, \boldsymbol{w}_1) = \mathcal{A}_L(\Psi(\boldsymbol{w}_0, \boldsymbol{w}_1))\,.$$

We refer to these minimizing paths as the geodesics associated with L. When L is the squared velocity norm of a Riemannian metric, they are constant-speed minimizing geodesics of that metric.

Computing the geodesics. For smooth L and a regular minimizing path, the first variation gives the Euler–Lagrange boundary-value problem

$$\frac{\mathrm{d}}{\mathrm{d}t}\partial_{\boldsymbol{v}} L(\boldsymbol{w}, \boldsymbol{v}) = \partial_{\boldsymbol{w}} L(\boldsymbol{w}, \boldsymbol{v}), \qquad \Psi_0 = \boldsymbol{w}_0, \quad \Psi_1 = \boldsymbol{w}_1.$$

Here $\boldsymbol{w} = \Psi_t$ and $\boldsymbol{v} = \partial_t \Psi_t$. This equation identifies candidate paths; minimality must still be verified. For the main ESFM choice $L((\boldsymbol{x}, y), (\boldsymbol{u}, g)) = \|\boldsymbol{u}\|^2 + \delta^2 g^2$, it yields $\partial_t^2 \Psi_t = 0$, hence $\Psi_t = (1-t)\boldsymbol{w}_0 + t\boldsymbol{w}_1$. Cauchy–Schwarz verifies minimality and gives $c_L = \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|^2 + \delta^2 (y_1 - y_0)^2$.

Dynamic and static transport. Integrating the single-particle action over the extended distribution gives

$$\begin{aligned} \inf_{\hat{\rho}, \hat{\boldsymbol{u}}, \hat{g}} \quad & \int_0^1 \int_{\mathbb{R}^d \times \mathbb{R}} L((\boldsymbol{x}, y), (\hat{\boldsymbol{u}}_t, \hat{g}_t))\, \hat{\rho}_t(\boldsymbol{x}, y)\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}y\,\mathrm{d}t, \\ \text{s.t.} \quad & \partial_t \hat{\rho}_t + \nabla_{\boldsymbol{x}} \cdot (\hat{\rho}_t \hat{\boldsymbol{u}}_t) + \partial_y (\hat{\rho}_t \hat{g}_t) = 0, \\ & \hat{\rho}_0(\boldsymbol{x}, y) = \rho_0(\boldsymbol{x}) \delta_0(y), \qquad \int_{\mathbb{R}} \mathrm{e}^y \hat{\rho}_1(\boldsymbol{x}, y)\,\mathrm{d}y = \rho_1(\boldsymbol{x}). \end{aligned} \tag{D.1}$$

The endpoint cost c_L defines the corresponding extended static problem,

$$\begin{aligned} \inf_{\hat{\pi}, \hat{\rho}_1 \geq 0} \quad & \iiiint c_L((\boldsymbol{x}_0, y_0), (\boldsymbol{x}_1, y_1))\, \hat{\pi}(\boldsymbol{x}_0, y_0, \boldsymbol{x}_1, y_1)\,\mathrm{d}\boldsymbol{x}_0\,\mathrm{d}y_0\,\mathrm{d}\boldsymbol{x}_1\,\mathrm{d}y_1, \\ \text{s.t.} \quad & \iint \hat{\pi}(\boldsymbol{x}_0, y_0, \boldsymbol{x}_1, y_1)\,\mathrm{d}\boldsymbol{x}_1\,\mathrm{d}y_1 = \hat{\rho}_0(\boldsymbol{x}_0, y_0) = \rho_0(\boldsymbol{x}_0)\delta_0(y_0), \\ & \iint \hat{\pi}(\boldsymbol{x}_0, y_0, \boldsymbol{x}_1, y_1)\,\mathrm{d}\boldsymbol{x}_0\,\mathrm{d}y_0 = \hat{\rho}_1(\boldsymbol{x}_1, y_1), \\ & \int_{\mathbb{R}} \mathrm{e}^{y_1} \hat{\rho}_1(\boldsymbol{x}_1, y_1)\,\mathrm{d}y_1 = \rho_1(\boldsymbol{x}_1). \end{aligned} \tag{D.2}$$

If L is convex in velocity, admissible dynamics admit trajectory decompositions, and minimizing paths can be selected measurably, Equations (D.1) and (D.2) have the same optimal value. The proof follows the particle-trajectory argument in Appendix A.3, using the convexity of L in place of the quadratic-action inequality.

Conditional FM targets. Given a coupling from the static problem, sample endpoint pairs from its normalized law and set $\boldsymbol{w}_t = \Psi_t(\boldsymbol{w}_0, \boldsymbol{w}_1)$. The regression targets are

$$(\boldsymbol{u}_t^\star, g_t^\star) = \partial_t \Psi_t(\boldsymbol{w}_0, \boldsymbol{w}_1), \qquad g_t^\star = \frac{\mathrm{d}}{\mathrm{d}t} \ln m_t.$$

The networks $(\boldsymbol{v}_\theta, g_\phi)$ are trained with

$$\mathbb{E}\left[\|\boldsymbol{v}_\theta(\boldsymbol{x}_t, y_t, t) - \boldsymbol{u}_t^\star\|^2 + \lambda \, |g_\phi(\boldsymbol{x}_t, y_t, t) - g_t^\star|^2 \right], \qquad \lambda > 0, \tag{D.3}$$

where $t \sim \mathcal{U}[0,1]$.

The construction therefore proceeds as

$$\begin{aligned} \text{Lagrangian and action} &\implies \text{minimizing paths and endpoint cost} \\ &\implies \text{coupling} \implies \text{FM targets.} \end{aligned}$$

An explicit endpoint cost or an efficient solver still depends on the chosen L.

D.2 Example: ESFM with linear physical mass

We now replace the log-mass-linear ESFM geodesic by one that is linear in physical mass. For positive endpoint masses $m_0, m_1 > 0$, define

$$\boldsymbol{x}_t = (1-t)\boldsymbol{x}_0 + t\boldsymbol{x}_1, \qquad m_t = (1-t)m_0 + t m_1, \qquad y_t = \ln m_t.$$

This is the constant-speed straight path in the coordinate $(\boldsymbol{x}, m)$. It minimizes

$$\begin{aligned} \mathcal{A}_{\mathrm{lin}}(\boldsymbol{x}, m) &= \int_0^1 \left(\|\dot{\boldsymbol{x}}_t\|^2 + \delta^2 \, |\dot{m}_t|^2 \right) \mathrm{d}t \\ &= \int_0^1 \left(\|\dot{\boldsymbol{x}}_t\|^2 + \delta^2 \mathrm{e}^{2y_t} \, |\dot{y}_t|^2 \right) \mathrm{d}t, \end{aligned}$$

where the second line rewrites exactly the same action in the ESFM coordinate $y = \ln m$. By Cauchy–Schwarz, its endpoint cost is

$$c_{\mathrm{lin}}(\boldsymbol{w}_0, \boldsymbol{w}_1) = \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|^2 + \delta^2 \, (\mathrm{e}^{y_1} - \mathrm{e}^{y_0})^2 \, .$$

Hence this is still ESFM on $(\boldsymbol{x}, y)$, but with the nonconstant mass-coordinate metric factor e^{2y} instead of the constant metric used in the main text.

As in the main ESFM formulation, normalize the source particle mass to $m_0 = 1$, so $y_0 = 0$. The extended static problem becomes

$$\begin{aligned} \inf_{\pi \geq 0} \quad & \iiint \left[\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|^2 + \delta^2 (\mathrm{e}^{y_1} - 1)^2 \right] \pi(\boldsymbol{x}_0, \boldsymbol{x}_1, y_1) \, \mathrm{d}\boldsymbol{x}_0 \, \mathrm{d}\boldsymbol{x}_1 \, \mathrm{d}y_1, \\ \text{s.t.} \quad & \iint \pi(\boldsymbol{x}_0, \boldsymbol{x}_1, y_1) \, \mathrm{d}\boldsymbol{x}_1 \, \mathrm{d}y_1 = \rho_0(\boldsymbol{x}_0), \\ & \iint \mathrm{e}^{y_1} \pi(\boldsymbol{x}_0, \boldsymbol{x}_1, y_1) \, \mathrm{d}\boldsymbol{x}_0 \, \mathrm{d}y_1 = \rho_1(\boldsymbol{x}_1). \end{aligned}$$

For a fixed spatial pair with $A > 0$, let f be the nonnegative measure of terminal log-mass and set

$$A = \int_{\mathbb{R}} \mathrm{d}f(y_1), \qquad B = \int_{\mathbb{R}} \mathrm{e}^{y_1} \, \mathrm{d}f(y_1), \qquad R = \frac{B}{A}.$$

The local mass problem is

$$\inf_{f \geq 0} \int_{\mathbb{R}} (\mathrm{e}^{y_1} - 1)^2 \, \mathrm{d}f(y_1), \qquad \int_{\mathbb{R}} \mathrm{d}f = A, \quad \int_{\mathbb{R}} \mathrm{e}^{y_1} \, \mathrm{d}f = B.$$

Normalize f to the probability measure μ by $\mathrm{d}\mu(y_1) = \mathrm{d}f(y_1)/A$. The two constraints then give

$$\int_{\mathbb{R}} \mathrm{d}\mu = 1, \qquad \int_{\mathbb{R}} \mathrm{e}^{y_1}\, \mathrm{d}\mu(y_1) = \frac{B}{A} = R.$$

Therefore, applying Jensen's inequality to the strictly convex function $\phi(r) = (r-1)^2$ gives

$$\begin{aligned}\int_{\mathbb{R}} (\mathrm{e}^{y_1} - 1)^2\, \mathrm{d}f(y_1) &= A \int_{\mathbb{R}} \phi(\mathrm{e}^{y_1})\, \mathrm{d}\mu(y_1) \\ &\geq A\phi\bigg(\int_{\mathbb{R}} \mathrm{e}^{y_1}\, \mathrm{d}\mu(y_1)\bigg) = A(R-1)^2.\end{aligned}$$

Since ϕ is strictly convex, equality holds if and only if $\mathrm{e}^{y_1} = R$, or equivalently $y_1 = \ln R$. Hence

$$f^\star = A\delta_{\ln R}.$$

Unlike the capped log-mass quadratic cost in the main ESFM model, this local problem does not require a two-point optimizer: all local mass is assigned the deterministic terminal value $y_1 = \ln R$.

Defining the semi-couplings γ_0 and γ_1 as in the main text, with $R = \mathrm{d}\gamma_1/\mathrm{d}\gamma_0 > 0$, eliminates y_1 and gives

$$\begin{aligned}\inf_{\gamma_0,\gamma_1 \geq 0} \quad & \iint \left[\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|^2 + \delta^2 \left(\frac{\mathrm{d}\gamma_1}{\mathrm{d}\gamma_0} - 1 \right)^2 \right] \mathrm{d}\gamma_0, \\ \text{s.t.} \quad & \int \gamma_0(\boldsymbol{x}_0, \boldsymbol{x}_1)\, \mathrm{d}\boldsymbol{x}_1 = \rho_0(\boldsymbol{x}_0), \\ & \int \gamma_1(\boldsymbol{x}_0, \boldsymbol{x}_1)\, \mathrm{d}\boldsymbol{x}_0 = \rho_1(\boldsymbol{x}_1).\end{aligned} \tag{D.4}$$

The only change from the main ESFM reduced problem is therefore the local growth penalty: $\Phi_M(R)$ is replaced by $(R-1)^2$.

D.3 Entropic coupling and conditional targets for the example

For discrete measures with source weights a_i, target weights b_j, and $C_{ij} = \left\|\boldsymbol{x}_i^0 - \boldsymbol{x}_j^1\right\|^2$, an entropic version of Equation (D.4) is

$$\begin{aligned}\min_{\Gamma_{ij} \geq 0,\, R_{ij} > 0} \quad & \sum_{i,j} \left[C_{ij}\Gamma_{ij} + \delta^2 \Gamma_{ij}(R_{ij} - 1)^2 + \varepsilon \Gamma_{ij}(\ln \Gamma_{ij} - 1) \right], \\ \text{s.t.} \quad & \sum_j \Gamma_{ij} = a_i, \qquad \sum_i R_{ij}\Gamma_{ij} = b_j.\end{aligned}$$

Introducing row and column dual variables u_i and v_j, elimination of R uses the one-dimensional conjugate

$$H_j(v_j) = \sup_{R>0} \left\{ v_j R - \delta^2 (R-1)^2 \right\}.$$

On the interior branch, its maximizer and value are

$$R_j = 1 + \frac{v_j}{2\delta^2}, \qquad H_j = \delta^2(R_j^2 - 1). \tag{D.5}$$

The row update has the familiar ESFM form

$$u_i = \varepsilon \ln a_i - \varepsilon \ln \sum_j \exp\left(\frac{H_j - C_{ij}}{\varepsilon} \right).$$

For fixed u, define

$$K_j = \frac{b_j}{\sum_i \exp((u_i - C_{ij})/\varepsilon)}.$$

The column constraint and Equation (D.5) reduce the column update to the scalar equation

$$\ln R_j + \frac{\delta^2}{\varepsilon}(R_j^2 - 1) = \ln K_j. \tag{D.6}$$

Its left-hand side is strictly increasing on $R_j > 0$, so Equation (D.6) has a unique positive solution. The semi-couplings are recovered as

$$\Gamma^{(0)}_{ij} = \exp\left(\frac{u_i + H_j - C_{ij}}{\varepsilon}\right), \qquad \Gamma^{(1)}_{ij} = R_j \Gamma^{(0)}_{ij}.$$

Thus changing the geodesic changes the scalar growth update, but preserves the same coupling architecture used by ESFM.

Finally, sample (i, j) from the normalized base coupling $\Gamma^{(0)}$, set $m_0 = 1$ and $m_1 = R_j$, and use

$$\begin{aligned}
&\boldsymbol{x}_t = (1-t)\boldsymbol{x}^0_i + t\boldsymbol{x}^1_j, \\
&m_t = 1 + t(R_j - 1), \qquad y_t = \ln(1 + t(R_j - 1)), \\
&\boldsymbol{u}^\star_t = \boldsymbol{x}^1_j - \boldsymbol{x}^0_i, \qquad g^\star_t = \dot{y}_t = \frac{R_j - 1}{1 + t(R_j - 1)} = (R_j - 1)\mathrm{e}^{-y_t}.
\end{aligned}$$

Substitution of these targets into Equation (D.3) gives a simulation-free ESFM training objective for the linear-mass geodesic.

D.4 Example: recovering WFR from the WFR geodesic

The general construction in Appendix D.1 also contains WFR as a particular choice of extended space geometry. Keep the coordinate $y = \ln m$, but replace the main ESFM action by the microscopic action

$$\mathcal{A}_{\mathrm{WFR}}(\boldsymbol{w}) = \int_0^1 \mathrm{e}^{y_t} \left(\|\dot{\boldsymbol{x}}_t\|^2 + \delta^2 \left|\dot{y}_t\right|^2\right) \mathrm{d}t. \tag{D.7}$$

Because $m_t = \mathrm{e}^{y_t}$ and $\dot{y}_t = \dot{m}_t / m_t$, this is exactly the microscopic WFR action. Its minimizing Dirac-to-Dirac paths are the WFR geodesics, including the traveling Dirac in the moving regime. Define

$$\cos_+(s) = \cos\left(\min\left\{s, \frac{\pi}{2}\right\}\right), \qquad s \geq 0.$$

The endpoint cost induced by Equation (D.7) is

$$c_{\mathrm{WFR}}(\boldsymbol{w}_0, \boldsymbol{w}_1) = 4\delta^2 \left[\mathrm{e}^{y_0} + \mathrm{e}^{y_1} - 2\cos_+\left(\frac{\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|}{2\delta}\right) \mathrm{e}^{(y_0+y_1)/2}\right].$$

Apply the extended static construction in Equation (D.2) with $y_0 = 0$. At a fixed spatial pair, let f denote the terminal log-mass measure and set

$$A = \int \mathrm{d}f(y_1), \qquad B = \int \mathrm{e}^{y_1} \, \mathrm{d}f(y_1).$$

The Cauchy–Schwarz inequality gives

$$\left(\int \mathrm{e}^{y_1/2} \, \mathrm{d}f(y_1)\right)^2 \leq AB,$$

and the bound is attained by $f^* = A\delta_{\ln(B/A)}$. Consequently, the optimized local cost is

$$4\delta^2 \left[A + B - 2\cos_+\left(\frac{\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|}{2\delta}\right)\sqrt{AB}\right].$$

Defining the semi-couplings $\gamma_0 = \int \pi \, \mathrm{d}y_1$ and $\gamma_1 = \int \mathrm{e}^{y_1} \pi \, \mathrm{d}y_1$ and integrating the local value over spatial pairs therefore yields

$$\begin{aligned}
\inf_{(\gamma_0,\gamma_1)\in\Gamma(\rho_0,\rho_1)} 4\delta^2 \iint \Big[&\gamma_0(\boldsymbol{x}_0, \boldsymbol{x}_1) + \gamma_1(\boldsymbol{x}_0, \boldsymbol{x}_1) \\
&- 2\cos_+\left(\frac{\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|}{2\delta}\right)\sqrt{\gamma_0(\boldsymbol{x}_0, \boldsymbol{x}_1)\gamma_1(\boldsymbol{x}_0, \boldsymbol{x}_1)}\Big] \, \mathrm{d}\boldsymbol{x}_0 \, \mathrm{d}\boldsymbol{x}_1,
\end{aligned}$$

where $\Gamma(\rho_0, \rho_1)$ imposes the source marginal on γ_0 and the target marginal on γ_1. This is the standard static WFR semi-coupling functional corresponding to Equation (C.3), equivalent to the formulation used in WFR-FM (Peng et al., 2026b, Equation (3.7)). Our action omits their overall factor of $1/2$, which leaves the optimal couplings unchanged. Using the same geodesic to construct conditional targets and retaining $\mathrm{e}^{y_t} = m_t$ as the regression weight gives the mass-weighted conditional loss used by WFR-FM. Hence WFR-FM is a special case of the generalized extended space framework when the WFR geodesic and action are selected.

E Intuition and Limitations of Minibatch OT

For large single-cell datasets, ESFM may use mini-batch or chunked coupling computations for memory reasons. This is a practical approximation to the full reduced coupling problem in Equation (4.8). As in WFR-FM, the theoretical convergence of mini-batch unbalanced or extended space couplings to the full coupling is not established here. Empirically, chunking is used as an engineering device.

We also conducted an ablation study on the choice of OT batch size on the Mouse dataset (see Table 12). Prediction performance remains broadly stable across the tested batch sizes, suggesting that ESFM is robust to the choice of mini-batch coupling size in this setting.

F Pseudocode for ESFM Inference

Given the trained field $(\boldsymbol{v}_\theta, g_\phi)$, inference starts from $(\boldsymbol{x}_0, 0)$ and solves

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\boldsymbol{x}_t\\ y_t\end{pmatrix} = \begin{pmatrix}\boldsymbol{v}_\theta(\boldsymbol{x}_t, y_t, t)\\ g_\phi(\boldsymbol{x}_t, y_t, t)\end{pmatrix}.$$

The predicted physical density is recovered by

$$\rho_1^{\mathrm{pred}}(\boldsymbol{x}) = \int \mathrm{e}^y \hat{\rho}_1^{\mathrm{pred}}(\boldsymbol{x}, y)\,\mathrm{d}y.$$

Algorithm 3 summarizes the inference procedure using forward Euler steps.

Algorithm 3 ESFM inference workflow

Input: initial samples $\{\boldsymbol{x}_0^{(i)}\}_{i=1}^N$, trained field $\boldsymbol{s}_{\theta,\phi} = (\boldsymbol{v}_\theta, g_\phi)$, final time T, step size Δt.
Initialize $\boldsymbol{x}^{(i)} \leftarrow \boldsymbol{x}_0^{(i)}$, $y^{(i)} \leftarrow 0$, and $t \leftarrow 0$ for all i.
while $t < T$ **do**
  Set $\Delta t_{\mathrm{step}} \leftarrow \min(\Delta t, T - t)$.
  for $i = 1, \ldots, N$ **do**
    Evaluate $(\boldsymbol{v}, g) = \boldsymbol{s}_{\theta,\phi}(\boldsymbol{x}^{(i)}, y^{(i)}, t)$.
    Update $\boldsymbol{x}^{(i)} \leftarrow \boldsymbol{x}^{(i)} + \boldsymbol{v}\,\Delta t_{\mathrm{step}}$.
    Update $y^{(i)} \leftarrow y^{(i)} + g\,\Delta t_{\mathrm{step}}$.
  end for
  Set $t \leftarrow t + \Delta t_{\mathrm{step}}$.
end while
Output: evolved extended particles $\{(\boldsymbol{x}^{(i)}, y^{(i)})\}_{i=1}^N$ and physical weights $\{\mathrm{e}^{y^{(i)}}\}_{i=1}^N$.