CURLY FLOW MATCHING FOR LEARNING NON-GRADIENT FIELD DYNAMICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Modeling the transport dynamics of natural processes from population-level observations is a ubiquitous problem that arises in the natural sciences. A key challenge in these settings is making important modeling assumptions over the scientific process at hand that enable faithful learning of governing dynamics that mimic actual system behavior. Traditionally, the de-facto assumption present in approaches relies on the principle of least action that result in gradient field dynamics, that lead to trajectories that minimize an energy functional between two probability measures. However, many real world systems such as cell cycles in single-cell RNA are known to exhibit non-gradient, periodic behavior, which *fundamentally* cannot be captured by current state-of-the-art methods such as optimal transport based conditional flow matching. In this paper, we introduce CURLY FLOW MATCHING (CURLY-FM), a novel approach that is capable of learning non-gradient field dynamics by designing and solving a Schrödinger bridge problem with a reference process with non-zero drift—in stark contrast from zero-drift reference processes—which is constructed using *inferred velocities* in addition to population snapshot data. We instantiate CURLY-FM by solving the single-cell trajectory inference problem with approximate velocities inferred using RNA velocity. We demonstrate that CURLY-FM can learn trajectories that match both RNA velocity and population marginals. CURLY-FM expands flow matching models beyond the modeling of populations and towards the modeling known periodic behavior observed in cells.

1 INTRODUCTION

Learning the underlying governing dynamics of cells through the phases of cell division is a core problem in cellular biology. Due to important advances in single-cell RNA sequencing (scRNA-seq) measurements (Macosko et al., 2015; Lähnemann et al., 2020), cellular dynamics modeling is informed by high-resolution measurements of snapshots through the various phases of cell state. Despite the advances in scRNA-seq the process is inherently destructive such that data is predominantly available at the population level rather than the sample level. This presents the inference of individual cell trajectories—the so-called *trajectory inference* problem—given snapshots in time as a significant problem of interest towards the accurate modeling of cellular dynamics.

Table 1: Overview of the properties of different approaches.

Method	$\mathbb{P}_\theta$		$\mathbb{Q}$		Models Curl
	v_t	σ_t	f_t	g_t	
DSBM	✓	✓	✗	✓	✗
OT-CFM	✓	✗	✗	$\lim_{g_t \to 0}$	✗
CURLY-FM (ours)	✓	✓	✓	$\lim_{g_t \to 0}$	✓

The dominant paradigm in solving the trajectory inference problem for single-cell RNA involves leveraging tools from computational optimal transport (OT) Peyré & Cuturi (2019) to learn neural dynamical systems, e.g. NeuralODE (Chen et al., 2018), such that sampled trajectories under the model optimize a notion of likeliness of being observed (Bunne et al., 2024). In a nutshell, such methods broadly follow a pipeline that first infers "optimal" cell trajectories which follow the gradient of some potential function (often termed a Waddington Landscape (Waddington, 1942)), before searching for important regulators of some biological process in development (Schiebinger et al., 2019; Shahan et al., 2022) or disease (Tong et al., 2023a; Klein et al., 2025). Despite the ability to produce approximately optimal trajectories with respect to the energy landscape, such methods are limited in their ability to model gradient-field dynamics. Consequently, trajectories inferred under the model are

not realistic and fail to model crucial system dynamics such as periodic behavior that arises during cell cycling. More precisely, cells are known to exhibit periodic behavior at several timescales with periodic behavior at the scale of months, days, hours, and minutes. These behaviors cannot be captured by current OT-based methods as periodic behavior cannot be captured by gradient field dynamics.

Present work. In this paper, we tackle the modeling of cellular dynamics under the presence of non-gradient field dynamics. We introduce CURLY FLOW MATCHING (CURLY-FM), a novel approach to learning non-gradient field dynamics by solving a Schrödinger bridge problem that is equipped with a designed reference process that induces the learning of periodic behavior. Specifically, we consider reference processes with non-zero drift—in stark contrast to zero drift processes in approaches such as Diffusion Schrödinger Bridges (De Bortoli et al., 2021b; Shi et al., 2024). Such a modification elevates the established (entropic optimal) mass transport problem to a new class of problems that require matching the reference drift while also transporting mass between time marginals associated with observations. As a result, solutions to this Schrödinger bridge problem are capable of learning non-gradient field dynamics and exhibit behaviors such as periodicity as found in cell cycling.

In addition to conventional population snapshot data, we design CURLY-FM by leveraging approximate velocity information which is used to construct the drift of a reference process. Consequently, to model periodic dynamics CURLY-FM solves the Schrödinger bridge problem by decomposing into a two-stage algorithm. The first stage learns a neural path interpolant by regressing against the drift of our constructed reference process. Unlike straight paths in optimal transform conditional flow matching trajectories under the neural path interpolant exhibit cyclic behavior due to the optimization objective of matching a constructed reference drift. In stage two we learn in a simulation-free manner to construct the generative process that solves the mass transport problem as a mixture of conditional bridges built using optimal transport-based couplings that minimize the length of the velocity field of the neural path interpolant. The combination of our two-stage approach enables CURLY-FM to learn capture dynamics that *do not* affect the population of cells, but do affect individual cells which include periodic behavior found in the cell cycle of scRNA-seq.

We instantiate CURLY-FM on modeling scRNA-seq data by constructing a reference drift using estimates of RNA-velocity (La Manno et al., 2018). More precisely, RNA-velocity is a technique that allows for the estimate of velocity at measured cells by exploiting our knowledge of the underlying system where we can sequence both old (spliced) and new (unspliced) versions of many genes to estimate the rate of change in transcription. This gives an approximation of the instantaneous velocity of a cell. Using RNA-velocity and population snapshots CURLY-FM solves the Schrödinger bridge problem by searching for a bridge that matches the reference process as closely as possible, but also matches the marginal distributions at both timepoints of the cellular dynamics.

We empirically validate CURLY-FM we consider a cell cycle system (Riba et al., 2022), with a single snapshot where the cells are known to be captured in multiple stages of the cell cycle. We show previous flow matching approaches fail to model this cyclic behavior as they are not able to take advantage of the additional RNA-velocity information. While there exist simulation-based methods that are in principle able to learn the correct dynamics (Tong et al., 2020), we show that in practice CCFM performs significantly better both in terms of accuracy to matching the reference drift due to its simulation-free training algorithm. We summarize our main contributions as follows:

1. We define the RNA-velocity regularized trajectory-inference problem, a principled Schrödinger bridge problem with non-zero reference drift that solves the trajectory inference.
2. We introduceCURLY-FM, a simulation-free training method for approximating solutions to the RNA-velocity informed Schroödinger bridge problem.
3. We investigate CURLY-FM on the Deep Cycle dataset with known periodic behavior, demonstrating the effective modeling of accurate cell cycles that cannot be modeled with prior approaches.

2 BACKGROUND AND PRELIMINARIES

Given two distributions μ_0 and μ_1 the distributional matching problem seeks to find a push-forward map that $\psi : \mathbb{R}^d \to \mathbb{R}^d$ transports the initial distribution to the desired endpoint $\mu_1 = [\psi]_\#(\mu_0)$. Such a problem setup is pervasive in many areas of machine learning and notably encompasses the standard generative modeling and optimal transport settings (De Bortoli et al., 2021a; Peyré et al., 2019). In this paper, we consider the setting where each distribution μ_0 and μ_1 is an

empirical distribution that is accessible through a dataset of observations $\{x_0^i\}_{i=1}^N \sim \mu_0(x_0)$ and $\{x_1^j\}_{j=1}^N \sim \mu_1(x_1)$. Thus the modeling task is to learn the (approximate) transport map ψ.

2.1 Continuous normalizing flows and flow matching

One common choice for modeling ψ is a deterministic dynamical system with a time-dependent generator $\psi_t : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$. The solution to this dynamical system is an ordinary differential equation (ODE) and the learned transport map is known as a continuous normalizing flow (CNF). A CNF is a time-indexed neural transport map ψ_t, for all time $t \in [0,1]$ that is trained to pushforward samples from prior μ_0 to a desired target μ_1. Specifically, a CNF models the following ODE $\frac{d}{dt}\psi_t(x) = f_t(\psi_t(x))$ with initial conditions $\psi_0(x_0) = x_0$ and $f_t : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ is the time dependent vector field associated to the ODE and transports samples from $\mu_0 \to \mu_1$.

The most scalable way to train CNFs is to utilize a *simulation-free* training objective which regresses a learned neural vector field $v_{t,\theta}(x_t) : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ to the desired target vector field $f_t(x_t)$ for all time. This technique is commonly known as flow-matching (Liu, 2022; Albergo & Vanden-Eijnden, 2023; Lipman et al., 2023; Tong et al., 2023c) and has the neural transport map $\psi_{t,\theta}$ which is obtained through a neural differential equation (Chen et al., 2018) $\frac{d}{dt}\psi_{t,\theta}(x) = v_{t,\theta}(\psi_{t,\theta}(x))$. Specifically, flow-matching regresses $v_{t,\theta}(x_t)$ to the target *conditional* vector field $f_t(x_t|z)$ associated to the target flow $\psi_t(x_t|z)$. We say that this conditional vector field $f_t(x_t|z)$, *generates* the target density $\mu_1(x_1)$ by interpolating along the probability path $\mu_t(x_t|z)$ in time. We often do not have closed-form access to the generating marginal vector field $f_t(x_t)$. Still, with conditioning, e.g. $z = (x_0, x_1)$, we can obtain a simple analytic expression of a conditional vector field that achieves the same goals. The conditional flow-matching (CFM) objective can then be stated as a simple simulation-free regression,

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t,q(z),\mu_t(x_t|z)} \|v_{t,\theta}(t, x_t) - f_t(x_t|z)\|_2^2. \tag{1}$$

The conditioning distribution $q(z)$ can be chosen from any valid coupling, for instance, the independent coupling $q(z) = \mu_0(x_0)\mu_1(x_1)$. To generate samples and their corresponding log density according to the CNF we may solve the following flow ODE numerically with initial conditions $x_0 = \psi_0(x_0)$ and $c = \log \mu_0(x_0)$, which is the log density under the prior:

$$\frac{d}{dt}\begin{bmatrix} \psi_{t,\theta}(x_t) \\ \log \mu_t(x_t) \end{bmatrix} = \begin{bmatrix} v_{t,\theta}(t, x_t) \\ -\nabla \cdot v_{t,\theta}(t, x_t) \end{bmatrix}. \tag{2}$$

In the next section, we outline a different methodology to build a transport map leveraging *stochastic* dynamics. This allows us to frame the mass transport problem as a Schrödinger bridge with non-zero reference drift, which is well suited to modeling noisy measurements found in single-cell evolution.

3 Schrödinger Bridge with non-zero reference field

The dynamic nature of cell evolution can be captured as a mass transport problem under a prescribed reference process. Specifically, we model the time evolution of cells using a parametrized stochastic differential equation (SDE), with drift $v_{t,\theta} : \mathbb{R}^d \to \mathbb{R}^d$ and parameters θ, diffusion coefficient $g_t > 0$:

$$\mathrm{d}\mathbf{X}_t = v_{t,\theta}(\mathbf{X}_t)\,\mathrm{d}t + g_t \mathrm{d}\mathbf{B}_t, \quad \mathbf{X}_0 \sim \mu_0, \mathbf{X}_1 \sim \mu_1, \tag{3}$$

where $\mathbf{B}_t$ is a standard Brownian motion and by convention time $t \in [0,1]$ flows from $t = 0$ to $t = 1$ such that marginal distribution at the endpoints are μ_0 and μ_1. These endpoints are provided as empirical distributions and represent endpoint observations along a transport trajectory. The SDE in eq. (3) induces a path measure in the space of Markov path measures $(\mathbb{P}_{t,\theta})_{t\in[0,1]} \in \mathcal{P}(C[0,1], \mathbb{R}^d)$ such that the marginal density p_t evolves according to the following Fokker-Plank equation:

$$\frac{\partial p}{\partial t} = -\nabla \cdot (v_{t,\theta}(\mathbf{X}_t), p_t(\mathbf{X}_t)) + \frac{g_t^2}{2}\Delta p_t(\mathbf{X}_t), \quad p_0 = \mu_0, p_1 = \mu_1. \tag{4}$$

In addition, our modeling of cellular dynamics is informed by a reference process which is defined by the following SDE with corresponding drift $f_t : \mathbb{R}^d \to \mathbb{R}^d$ and diffusion coefficient $\sigma_t > 0$:

$$\mathrm{d}\mathbf{X}_t = f_t(\mathbf{X}_t)\,\mathrm{d}t + \sigma_t d\mathbf{B}_t. \tag{5}$$

We denote the induced path measure of eq. (5) as $(\mathbb{Q}_t)_{t\in[0,1]} \in \mathcal{P}(C[0,1], \mathbb{R}^d)$.

We now aim to solve the Schrödinger bridge problem which finds an optimal path measure $\mathbb{P}^*$ that is the solution to the following KL-divergence minimization problem over path measures:

$$\mathbb{P}^* = \underset{\theta}{\operatorname{argmin}} \left[\mathrm{KL}\left(\mathbb{P}_\theta || \mathbb{Q}\right) : \mathbb{P}_0 = \mu_0, \mathbb{P}_1 = \mu_1\right] \tag{6}$$

In settings where eq. (5) is driftless and constant diffusion coefficient—i.e. $\mathrm{d}\mathbf{X}_t = \sigma \mathbf{B}_t$—the Schrödinger bridge problem (Schrödinger, 1932) devolves into the *Diffusion Schrödinger Bridge* problem (De Bortoli et al., 2021a; Bunne et al., 2023). In this special case, the Schrödinger bridge problem admits a unique solution and is linked to the entropic optimal transport plan through the seminal result of Föllmer (1988). Specifically, $\mathbb{P}^*$ is a mixture of conditional Brownian bridges $\mathbb{Q}_t(\cdot|x_0, x_1)$ weighted by the entropic OT-plan $\pi^* \in \Pi(\mu_0 \otimes \mu_1)$ which is a valid coupling in the product measure $\mu_0 \otimes \mu_1$, in other words $\int \pi(x_0, \cdot) = \mu_0(x_0), \int \pi(\cdot, x_1) = \mu(x_1)$,

$$\mathbb{P}^* = \int \mathbb{Q}_t\left(\cdot|x_0, x_1\right) d\pi^*(x_0, x_1) \tag{7}$$

$$\pi^*(\mu_0, \mu_1) = \underset{\pi \in \Pi(\mu_0 \otimes \mu_1)}{\operatorname{argmin}} \int c(x_0, x_1) d\pi(x_0, x_1) + 2\sigma^2 \mathrm{KL}(\pi || \mu_0 \otimes \mu_1). \tag{8}$$

Operationally, the conditional Brownian bridges take the form of a Normal distribution $\mathbb{Q}_t(\cdot|x_0, x_1) = \mathcal{N}(x_t; tx_1 + (1-t)x_0, t(1-t)\sigma^2)$ with the mean given as an interpolation between two endpoints. Furthermore, when $\sigma \to 0$ the entropic OT problem reduces to the regular OT problem. We note that this Schrödinger bridge problem can be reinterpreted as a stochastic optimal control problem where the control cost is the drift $v_{t,\theta}$. That is the stochastic optimal control perspective minimizes *average kinetic energy*[1] of the learned process which leads to the following optimization problem:

$$v_\theta^* = \left\{\min_\theta \int \mathbb{E}_{\mathbb{P}_t}\left[\frac{1}{2}\|v_{t,\theta}(\mathbf{X}_t)\|_2^2\right] dt : \mathrm{d}\mathbf{X}_t = v_{t,\theta}(\mathbf{X}_t)\,\mathrm{d}t + g_t \mathrm{d}\mathbf{B}_t, \mathbb{P}_0 = \mu_0,\, \mathbb{P}_1 = \mu_1\right\}. \tag{9}$$

Computationally, $\mathbb{P}^*$ can be approximated using mini-batch optimal transport Fatras et al. (2020; 2021) and simulation-free matching algorithms (Tong et al., 2023c;b; Pooladian et al., 2023), iterative proportional and Markov fitting (De Bortoli et al., 2021a; Shi et al., 2024), and generalized Schrödinger bridge matching (Liu et al., 2023a).

3.1 Schrödinger Bridges with non-zero drift

We now consider the more general case where the drift of the reference process $\mathbb{Q}$ is non-zero. In this case, existing computational approaches no longer apply. More precisely, we are unable to exploit the key property that the conditional bridges are a mixture of Brownian bridges. However, we make the observation when $\sigma \to 0$, the OT coupling must minimize the kinetic energy of interpolants between the marginals μ_0 and μ_1. Consequently, we aim to minimize the *average relative kinetic energy* as a stochastic control cost under the prescribed reference vector field $f_t(x_t)$:

$$v_\theta^* = \left\{\min_\theta \int \mathbb{E}_{\mathbb{P}_t}\left[\frac{1}{2}\|v_{t,\theta}(x_t) - f_t(x_t)\|_2^2\right] : \mathrm{d}\mathbf{X}_t = v_{t,\theta}(\mathbf{X}_t)\,\mathrm{d}t + g_t \mathrm{d}\mathbf{B}_t, \mathbb{P}_0 = \mu_0,\, \mathbb{P}_1 = \mu_1\right\},$$

with the key distinction that the reference process reduces to an ODE: $\mathrm{d}x_t = f_t(x_t)\,\mathrm{d}t$. We highlight the solution is still a mixture of conditional bridges $\mathbb{P}_t$, which need to be constructed. We approximate $\mathbb{P}_t$ by designing a neural path interpolant $\varphi_{t,\eta}$ with parameters η that we use to learn conditional flows:

$$x_{t,\eta} = tx_1 + (1-t)x_0 + t(1-t)\varphi_{t,\eta}(x_0, x_1). \tag{10}$$

We optimize $\varphi_{t,\eta}$ by minimizing the following simulation-free objective of the relative kinetic energy:

$$\mathcal{L}(\eta) = \mathbb{E}_{t\sim\mathcal{U}[0,1], x_0\sim\mu_0(x_0), x_1\sim\mu_1(x_1)}\left[\left\|\alpha\frac{\partial x_{t,\eta}}{\partial t} - f_t(x_{t,\eta})\right\|_2^2\right], \quad f_t(x_{t,\eta}) = \kappa(x_{t,\eta}, x_0) f_0(x_0).$$

Here α is a learned global scaling and $f_t(x_t)$ is estimated using a smooth nearest neighbor based distance kernel $\kappa(x_{t,\eta}, x_0) = \|x_{t,\eta} - x_0^i\|_2 / \sum_i^N \|x_{t,\eta} - x_0^i\|_2$ with respect to N closest points from μ_0. The time derivative of the neural path-interpolant can be computed using automatic differentiation:

$$\frac{\partial x_{t,\eta}}{\partial t} = x_1 - x_0 + t(1-t)\frac{\partial \varphi_{t,\eta}}{\partial t}(x_0, x_1) + (1-2t)\varphi_{t,\eta}(x_0, x_1). \tag{11}$$

[1] Schrödinger bridges minimize the relative entropy w.r.t. to $\mathbb{Q}$ and kinetic energy in the deterministic case.

Algorithm 1 Training algorithm for neural path interpolant network

Require: Time marginals $\mu_0(x_0)$ and $\mu_1(x_1)$, neural path interpolant φ_η, reference field f_t

while Training **do**

  Sample $(x_0, x_1) \sim \mu_0(x_0)\mu_1(x_1)$ and $t \sim \mathcal{U}(0,1)$

  $x_{t,\eta} \leftarrow (1-t)x_0 + tx_1 + t(1-t)\varphi_{t,\eta}(x_0, x_1)$

  $\frac{\partial x_{t,\eta}}{\partial t} \leftarrow x_1 - x_0 + t(1-t)\frac{\partial \varphi_{t,\eta}(x_0,x_1)}{\partial t} + (1-2t)\varphi_{t,\eta}(x_0, x_1)$

  $\mathcal{L}(\eta) \leftarrow \left\| \alpha \frac{\partial x_{t,\eta}}{\partial t} - f_t(x_{t,\eta}) \right\|_2^2, \quad f_t(x_{t,\eta}) = \kappa(x_{t,\eta}, x_0)$

  Update η using gradient $\nabla_\eta \mathcal{L}(\eta)$

return (approximate) optimal interpolants parametrized by $\varphi_{t,\eta}$

Algorithm 2 Marginal Flow Matching

Require: Time marginal $\mu_0(x_0)$, trained network $\varphi_{t,\eta}$, vector field network $v_{t,\theta}$.

while Training **do**

  Sample $(x_0, x_1) \sim q$ and $t_{ij} \sim \mathcal{U}(0,1)$

  $C_\eta^{ij}(x_0^i, x_1^j) \leftarrow \mathbb{E}_t \left[\left\| \frac{\partial x_{t,\eta}}{\partial t} \right\|_2^2 \right]$

  $\pi(x_0, x_1) \leftarrow \text{OT}(x_0, x_1, C_\eta)$

  $x_0, x_1 \sim \pi(x_0, x_1)$

  $t \sim \mathcal{U}(0,1)$

  $\mathcal{L}(\theta) \leftarrow \left\| v_{t,\theta}(x_{t,\eta}) - \left(\frac{\partial x_{t,\eta}}{\partial t} \right) \texttt{.detach()} \right\|_2^2$

  $\theta \leftarrow \text{Update}(\theta, \nabla_\theta \mathcal{L}(\theta))$

return v_θ

The pseudocode for learning the neural path is presented in algorithm 1. To approximate $\mathbb{P}_t$ we next learn to approximate the optimal mixture of conditional bridges $\mathbb{P}_t^* = \mathbb{E}_{x_0, x_1 \sim \pi^*(x_0, x_1)} [x_{t,\eta}]$. However, this necessitates the feasibility of computing the OT-plan which for this problem is defined below:

$$\pi^*(\mu_0, \mu_1) = \underset{\pi \in \Pi(\mu_0 \otimes \mu_1)}{\text{argmin}} \int c(x_0, x_1) d\pi(x_0, x_1), \quad c(x_0, x_1) = \int_0^1 \left\| \frac{\partial x_{t,\eta}}{\partial t} \right\|_2^2 dt$$
$$\text{s.t.} \int \pi(x_0, \cdot) = \mu_0(x_0), \int \pi(\cdot, x_1) = \mu(x_1).$$

Instead of computing the optimal transport cost $c(x_0, x_1)$ through simulation we leverage an unbiased stochastic estimator with K samples:

$$c(x_0, x_1) = \mathbb{E}_{t \sim \mathcal{U}[0,1]} \left[\left\| \frac{\partial x_{t,\eta}}{\partial t} \right\|_2^2 \right] = \frac{1}{K} \sum_i^K \left\| \frac{\partial x_{t,\eta}^i}{\partial t} \right\|_2^2. \tag{12}$$

We use the cost in eq. (12) to estimate a transport plan $\pi(x_0, x_1)$ which we then use to construct the approximated mixture of conditional bridges $\mathbb{P}_t$. Using this we next pose the learning problem for learning the drift $v_{t,\theta}(x_t)$ of the SDE in eq. (3) evolves cells:

$$\mathcal{L}(\theta) = \int \mathbb{E}_{\mathbb{P}_t} \left[\frac{1}{2} \left\| v_{t,\theta}(x_{t,\eta}) - \frac{\partial x_{t,\eta}}{\partial t} \right\|_2^2 \right], \quad \mathbb{P}_t = \mathbb{E}_{x_0, x_1 \sim \pi(x_0, x_1)} [x_{t,\eta}].$$

The pseudocode for this marginal (flow) matching objective is presented in algorithm 2.

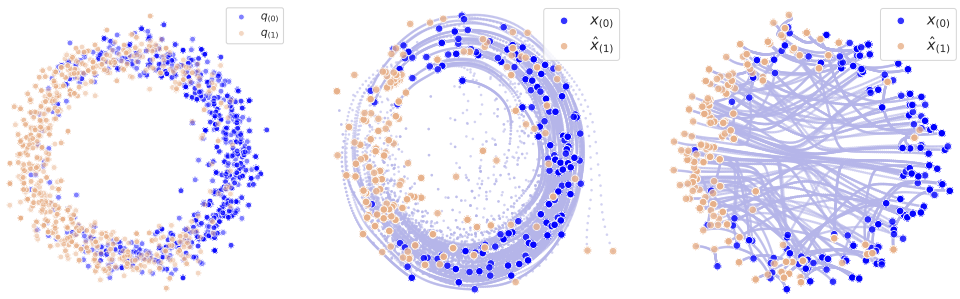
Remark 1. *We highlight that while $\mathcal{L}(\theta)$ seeks to match $v_{t,\theta}$ to velocity of the neural path-interpolant $\frac{\partial x_{t,\eta}}{\partial t}$ the optimal velocity $v_t^* \neq f_t(x_t)$. This is because the reference process $\mathbb{Q}$ does not necessarily transport μ_0 to μ_1. More precisely, the reference process does not have constraints at the endpoints that $\mathbb{Q}_0 = \mu_0$ and $\mathbb{Q}_1 = \mu_1$ which is required from our learned process $\mathbb{P}_\theta$ and its drift $v_{t,\theta}$.*

4 EXPERIMENTS

We investigate the application of CURLY-FM on synthetic toy data as well as a real-world dataset based on the different stages of cell cycles found in single-cell trajectories. Specifically, we benchmark on the Deep Cycle dataset (Riba et al., 2022) which contains single-cell trajectories with RNA-velocity data on samples $x_0 \sim \mu_0(x_0)$ but not elsewhere in both space and time. Through these experiments we test the efficacy of our CURLY-FM approach in generalizing across previously unseen population data *and* learning cycling dependencies which more faithfully capture underlying cell cycle dynamics. We use this RNA-velocity to estimate the velocity of the vector field at intermediate time steps $f_t(x_t)$ using our weighted nearest neighbor kernel as outlined in algorithm 1.

Baselines. For baselines, we rely on flow based generative models that have been applied to single-cell and trajectory inference problems. Specifically, we compare CURLY-FM to TrajectoryNet (Tong et al., 2020) which learns dynamics in single-cell data from static single-cell RNA sequencing snapshots by combining neural ODEs learned through maximum likelihood training with optimal transport to model transitions between time points. We further investigate simulation-free alternatives to TrajectoryNet in Conditional Flow Matching (CFM) (Liu, 2022; Lipman et al., 2023; Tong et al., 2023c) that also learns vector fields as a neural ODE between the two-time marginals but using a flow matching-based regression objective. We further include Optimal-Transport Conditional Flow Matching (OT-CFM) (Tong et al., 2023c) which enhances conditional flow-matching with mini-batch optimal transport-based couplings that minimize the kinetic energy of the *target conditional vector-field*. We note that neither of these simulation-free methods are able to incorporate velocity information.

Metrics. We report Maximum Mean Discrepancy (MMD) and Wasserstein-2 ($\mathcal{W}_2$) distance between the ground truth x_1 and the predicted $\hat{x}_1$ after simulating the respective neural ODE's of each method. These quantitative metrics assess the ability of each method in performing trajectory inference by reconstructing the x_1 samples given the associated starting sample x_0. To assess the ability of each method to model cell cycle dynamics we compute the cosine distance of the RNA-velocity of the reference process $\mathbb{Q}$ averaged over the time of integration—i.e. from $t = 0$ to $t = 1$. A lower cosine distance indicates better adherence to the vector field of the reference process f_t and is correlated with more faithful modeling of the underlying cell cycle dynamics of single-cell trajectories. We use cosine distance following Tong et al. (2020) because RNA-velocity is known to have inaccurate magnitudes.



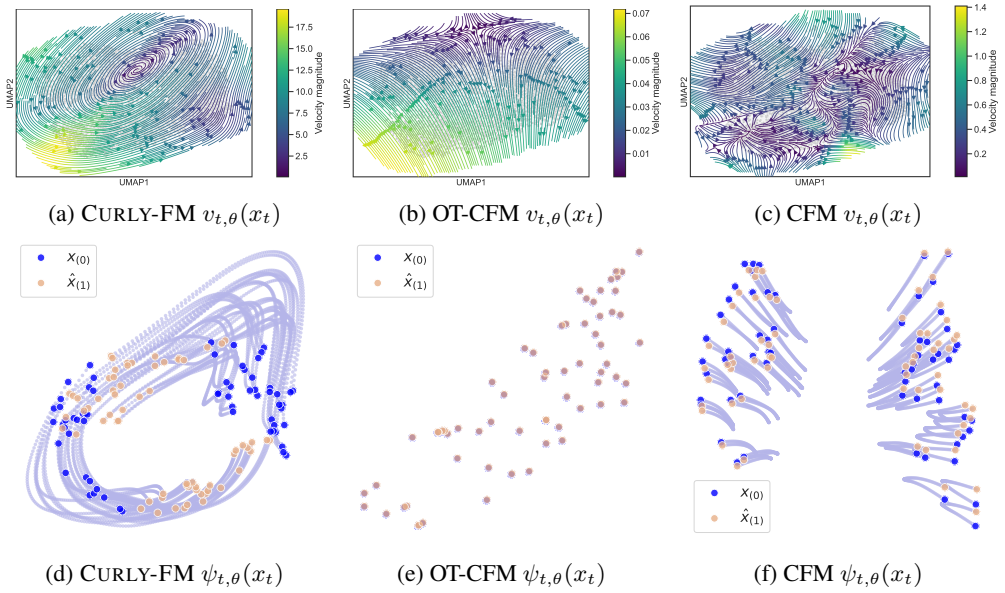
(a) Asymmetric `circles` (b) CURLY-FM (c) CFM

Figure 1: Trajectories generated between samples drawn from asymmetric `circles` distribution at $t = 0$ and $t = 1$ and respective to underlying reference field $f_t(x_t, \omega)$. CURLY-FM is capable of learning cycling dependencies unlike other flow based models such as CFM.

4.1 SYNTHETIC EXPERIMENTS

We consider a simple task of learning trajectories from population-level observed populations that exhibit non-gradient, periodic behavior. To examine this, we construct simple source and target distributions on asymmetrically arranged `circles` (Figure 1a), each with higher particle population density on one side. Given a circular reference velocity field $f_t(x_t, \omega)$ with constant rotational speed $\omega = \pi$, the goal is to learn velocity-field $v_{t,\theta}(x_t)$ and trajectories $\psi_{t,\theta}(x_t)$ between $t = 0$ and $t = 1$.

Results. We compare CURLY-FM with CFM, demonstrating that standard approaches such as CFM are not able to capture cycling patterns of the reference field dynamics. Figures 1b and 1c clearly show that using methods such as CFM with zero-reference field f_t results in straight paths between source and target distributions, thereby failing to capture cycling patterns in the underlying data.



(a) CURLY-FM $v_{t,\theta}(x_t)$ (b) OT-CFM $v_{t,\theta}(x_t)$ (c) CFM $v_{t,\theta}(x_t)$

(d) CURLY-FM $\psi_{t,\theta}(x_t)$ (e) OT-CFM $\psi_{t,\theta}(x_t)$ (f) CFM $\psi_{t,\theta}(x_t)$

Figure 2: Trajectories $\psi_{t,\theta}(x_t)$ and RNA-velocity fields $v_{t,\theta}(x_t)$ learned using CURLY-FM (figures 2a and 2d) and conditional flow matching (figures 2c, 2d 2b, 2e). Results show that unlike other methods CURLY-FM learns cycling patterns matching the learned behavior to that of cell cycle.

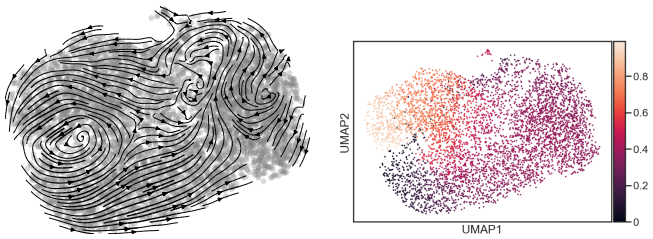
Table 2: Quantitative results for cell cycle trajectory inference task. We report the mean result for a metric with standard deviation over three seeds. CURLY-FM performs the best across matching inferred velocity field to the reference process (cosine distance) while maintaining comparable performance on other metrics.

Datasets →	$d = 5$			$d = 10$			$d = 20$		
Algorithm ↓	$\mathcal{W}_2$ ↓	MMD ↓	Cos. Dist ↓	$\mathcal{W}_2$ ↓	MMD ↓	Cos. Dist ↓	$\mathcal{W}_2$ ↓	MMD ↓	Cos. Dist ↓
CFM	0.294 ± 0.030	0.493 ± 0.110	1.065 ± 0.080	0.606 ± 0.059	0.120 ± 0.022	1.001 ± 0.037	1.227 ± 0.013	0.031 ± 0.003	1.007 ± 0.010
OT-CFM	0.248 ± 0.030	0.387 ± 0.079	0.800 ± 0.309	0.586 ± 0.041	0.118 ± 0.025	1.008 ± 0.039	1.183 ± 0.015	0.024 ± 0.004	0.978 ± 0.125
TrajectoryNet	0.531 ± 0.021	0.714 ± 0.061	1.077 ± 0.031	0.853 ± 0.059	0.238 ± 0.018	0.979 ± 0.064	–	–	–
CURLY-FM (Ours)	0.944 ± 0.255	0.914 ± 0.193	$\mathbf{0.387 \pm 0.145}$	0.972 ± 0.044	0.214 ± 0.005	$\mathbf{0.343 \pm 0.105}$	4.263 ± 0.535	0.091 ± 0.014	$\mathbf{0.364 \pm 0.088}$

4.2 EXPERIMENTS ON SINGLE-CELL RNA-VELOCITY DATA

To show that CURLY-FM is effective in learning cycling behavior in single-cell data, we leverage a cell cycle dataset of human fibroblasts (Riba et al., 2022) to learn cell trajectories by considering RNA-velocities (see figure 3a) during cell cycle. As a result, cells rotate in space, with rotation of a cell $\gamma \in [0, 1]$ during cell cycling phase shown in figure 3b).

We consider a single snapshot data of RNA velocities and construct source and target distributions μ_0 and μ_1. We leverage RNA-velocity field to estimate reference velocity field $f_t(x_t)$ by smoothing over $N = 20$ nearest neighbors drawn from from μ_0 using $\kappa(x_{t,\eta}, x_0) = \|x_{t,\eta} - x_0^i\|_2 / \sum_i^N \|x_{t,\eta} - x_0^i\|_2$. Further, we consider an arbitrary



(a) RNA-Velocity Field (b) Cell Cycles

Figure 3: Data

d-dimensional set-up, where d is the number of top ranked genes based on their variability. Our task is to learn underlying non-gradient field dynamics and cycling behavior.

Traditional flow-based models cannot capture cycling patterns. Figure 2 shows an overview of learned velocity fields $v_{t,\theta}(x_t)$ and trajectories $\psi_{t,\theta}(x_t)$ between cell cycle distributions at $t = 0$ and $t = 1$. In table 2, we show results on trajectory inference task comparing CURLY-FM to CFM, OT-CFM, and TrajectoryNet. Given underlying cell cycle process, the aim to learn circular trajectories resulting from a divergent-free velocity field. While traditional methods are successful in generating end points in close proximity to ground truth, they fail at learning cyclic patterns as shown in figures 2f.

CURLY-FM demonstrates learning non-gradient field dynamics. Our results show that considering non-zero reference field and velocity inference captures non-gradient dynamics patterns in data. Figures 2a and 2d show learned behavior using CURLY-FM. We observe that trajectory $\psi_{t,\theta}(x_t)$ inferred with CURLY-FM closely matches expected cycling patterns in the fibroblast dataset, in contrast to trajectories inferred using CFM and OT-CFM. This is quantified in Table 2 where we can see the cosine distance to the reference field is significantly lower for CURLY-FM.

5 RELATED WORK

Flow-matching for efficient continuous normalizing flows. Flow-matching (Lipman et al., 2023), also known as rectified flows (Liu, 2022; Liu et al., 2023b) or stochastic interpolants (Albergo & Vanden-Eijnden, 2023; Albergo et al., 2023), has emerged as the default method for training continuous normalizing flow (CNF) models (Chen et al., 2018; Grathwohl et al., 2019). However, unlike maximum likelihood training, flow matching training requires you to pre-specify conditional probability paths, which can make it less flexible in modeling dynamics, therefore many works attempt to derive methods for using minimum energy (Tong et al., 2023c; Pooladian et al., 2023) more flexible conditional paths (Neklyudov et al., 2024; Kapuśniak et al., 2024).

Schrödinger bridges with deep learning. To tackle the Schrödinger bridge problem in high dimensions many methods propose simulation-based (De Bortoli et al., 2021b; Chen et al., 2022; Koshizuka & Sato, 2023; Liu et al., 2022) and simulation-free (Shi et al., 2024; Tong et al., 2023b; Pooladian & Niles-Weed, 2024; Liu et al., 2023a) methods with various additional components incorporating variable growth rates (Zhang et al., 2024; Pariset et al., 2023; Sha et al., 2024), stochasticity, and manifold structure Huguet et al. (2022) have been proposed based on neural ODE and neural SDE (Li et al., 2020; Kidger et al., 2021) frameworks. However, very few methods are able to incorporate approximate velocity data, and either match marginals using simulation (Tong et al., 2020), or do not attempt to match marginals (Qiu et al., 2022).

RNA-velocity methods on discrete manifolds. A common strategy to regularize and interpret RNA-velocity (La Manno et al., 2018; Bergen et al., 2020) is to restrict it to a Markov process on a graph of cells representing a discrete manifold or compute higher level statistics on it (Qiu et al., 2022). However, these approaches are not equipped to match the marginal cell distribution over time. CURLY-FM can be seen as a method that unites these approaches with marginal-matching approaches.

6 CONCLUSION

In this work we introduced CURLY-FM, a method capable of learning non-gradient field dynamics by solving a Schrödinger bridge problem with a non-zero reference process drift. In contrast to prior work which leverages RNA-velocity, CURLY-FM is simulation-free, greatly improving numerical stability and efficiency. We showed the utility of this method in learning more accurate dynamics in a cell cycle system with known periodic behavior. CURLY-FM opens up the possibility of moving beyond modeling population-dynamics with simulation-free training methods and towards reconstructing the underlying governing dynamics (Xing, 2022).

Nevertheless, CURLY-FM is currently limited in its ability to discover the underlying dynamics by accurate inference of the reference field, which is an inherently difficult problem especially over longer timescales. Exciting directions for future work involve additional verification of trajectories through lineage tracing (McKenna & Gagnon, 2019; Wagner & Klein, 2020), and improved modeling across non-stationary populations with the additional incorporation of unbalanced transport or multiomics datatypes (Baysoy et al., 2023).

REFERENCES

Michael S. Albergo and Eric Vanden-Eijnden. Building normalizing flows with stochastic interpolants. *International Conference on Learning Representations (ICLR)*, 2023. (Cited on pages 3 and 8)

Michael S. Albergo, Nicholas M. Boffi, and Eric Vanden-Eijnden. Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint 2303.08797*, 2023. (Cited on page 8)

Alev Baysoy, Zhiliang Bai, Rahul Satija, and Rong Fan. The technological landscape and applications of single-cell multi-omics. *Nature Reviews Molecular Cell Biology*, 24(10):695–713, 2023. (Cited on page 8)

Volker Bergen, Marius Lange, Stefan Peidli, F Alexander Wolf, and Fabian J Theis. Generalizing rna velocity to transient cell states through dynamical modeling. *Nature biotechnology*, 38(12): 1408–1414, 2020. (Cited on page 8)

Charlotte Bunne, Ya-Ping Hsieh, Marco Cuturi, and Andreas Krause. The schrödinger bridge between gaussian measures has a closed form. In *International Conference on Artificial Intelligence and Statistics*, pp. 5802–5833. PMLR, 2023. (Cited on page 4)

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. (Cited on page 1)

Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural ordinary differential equations. *Neural Information Processing Systems (NeurIPS)*, 2018. (Cited on pages 1, 3, and 8)

Tianrong Chen, Guan-Horng Liu, and Evangelos A. Theodorou. Likelihood training of Schrödinger bridge using forward-backward SDEs theory. *International Conference on Learning Representations (ICLR)*, 2022. (Cited on page 8)

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion schrödinger bridge with applications to score-based generative modeling. *Advances in Neural Information Processing Systems*, 34:17695–17709, 2021a. (Cited on pages 2 and 4)

Valentin De Bortoli, James Thornton, Jeremy Heng, and Arnaud Doucet. Diffusion Schrödinger bridge with applications to score-based generative modeling. *Neural Information Processing Systems (NeurIPS)*, 2021b. (Cited on pages 2 and 8)

Kilian Fatras, Younes Zine, Rémi Flamary, Rémi Gribonval, and Nicolas Courty. Learning with minibatch Wasserstein: Asymptotic and gradient properties. *Artificial Intelligence and Statistics (AISTATS)*, 2020. (Cited on page 4)

Kilian Fatras, Younes Zine, Szymon Majewski, Rémi Flamary, Rémi Gribonval, and Nicolas Courty. Minibatch optimal transport distances; analysis and applications. *arXiv preprint 2101.01792*, 2021. (Cited on page 4)

Hans Föllmer. Random fields and diffusion processes. *Lect. Notes Math*, 1362:101–204, 1988. (Cited on page 4)

Will Grathwohl, Ricky T. Q. Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *International Conference on Learning Representations (ICLR)*, 2019. (Cited on page 8)

Guillaume Huguet, D. S. Magruder, Alexander Tong, Oluwadamilola Fasina, Manik Kuchroo, Guy Wolf, and Smita Krishnaswamy. Manifold interpolating optimal-transport flows for trajectory inference. *Neural Information Processing Systems (NeurIPS)*, 2022. (Cited on page 8)

Kacper Kapuśniak, Peter Potaptchik, Teodora Reu, Leo Zhang, Alexander Tong, Michael Bronstein, Avishek Joey Bose, and Francesco Di Giovanni. Metric flow matching for smooth interpolations on the data manifold, 2024. URL https://arxiv.org/abs/2405.14780. (Cited on page 8)

Patrick Kidger, James Foster, Xuechen Li, Harald Oberhauser, and Terry Lyons. Neural SDEs as Infinite-Dimensional GANs. *International Conference on Machine Learning*, 2021. (Cited on page 8)

Dominik Klein, Giovanni Palla, Marius Lange, Michal Klein, Zoe Piran, Manuel Gander, Laetitia Meng-Papaxanthos, Michael Sterr, Lama Saber, Changying Jing, et al. Mapping cells through time and space with moscot. *Nature*, pp. 1–11, 2025. (Cited on page 1)

Takeshi Koshizuka and Issei Sato. Neural lagrangian schrödinger bridge: Diffusion modeling for population dynamics, 2023. URL https://arxiv.org/abs/2204.04853. (Cited on page 8)

Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E Kastriti, Peter Lönnerberg, Alessandro Furlan, et al. Rna velocity of single cells. *Nature*, 560(7719):494–498, 2018. (Cited on pages 2 and 8)

David Lähnemann, Johannes Köster, Ewa Szczurek, Davis J. McCarthy, Stephanie C. Hicks, Mark D. Robinson, Catalina A. Vallejos, Kieran R. Campbell, Niko Beerenwinkel, Ahmed Mahfouz, Luca Pinello, Pavel Skums, Alexandros Stamatakis, Camille Stephan-Otto Attolini, Samuel Aparicio, Jasmijn Baaijens, Marleen Balvert, Buys De Barbanson, Antonio Cappuccio, Giacomo Corleone, Bas E. Dutilh, Maria Florescu, Victor Guryev, Rens Holmer, Katharina Jahn, Thamar Jessurun Lobo, Emma M. Keizer, Indu Khatri, Szymon M. Kielbasa, Jan O. Korbel, Alexey M. Kozlov, Tzu-Hao Kuo, Boudewijn P.F. Lelieveldt, Ion I. Mandoiu, John C. Marioni, Tobias Marschall, Felix Mölder, Amir Niknejad, Lukasz Raczkowski, Marcel Reinders, Jeroen De Ridder, Antoine-Emmanuel Saliba, Antonios Somarakis, Oliver Stegle, Fabian J. Theis, Huan Yang, Alex Zelikovsky, Alice C. McHardy, Benjamin J. Raphael, Sohrab P. Shah, and Alexander Schönhuth. Eleven grand challenges in single-cell data science. *Genome Biology*, 21(1):31, 2020. ISSN 1474-760X. doi: 10.1186/s13059-020-1926-6. (Cited on page 1)

Xuechen Li, Ting-Kam Leonard Wong, Ricky T. Q. Chen, and David Duvenaud. Scalable gradients for stochastic differential equations. *International Conference on Artificial Intelligence and Statistics*, 2020. (Cited on page 8)

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *International Conference on Learning Representations (ICLR)*, 2023. (Cited on pages 3, 6, and 8)

Guan-Horng Liu, Tianrong Chen, Oswin So, and Evangelos A Theodorou. Deep generalized Schrödinger bridge. *Neural Information Processing Systems (NeurIPS)*, 2022. (Cited on page 8)

Guan-Horng Liu, Yaron Lipman, Maximilian Nickel, Brian Karrer, Evangelos A Theodorou, and Ricky TQ Chen. Generalized schr\" odinger bridge matching. *arXiv preprint arXiv:2310.02233*, 2023a. (Cited on pages 4 and 8)

Qiang Liu. Rectified flow: A marginal preserving approach to optimal transport. *arXiv preprint arXiv:2209.14577*, 2022. (Cited on pages 3, 6, and 8)

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. *International Conference on Learning Representations (ICLR)*, 2023b. (Cited on page 8)

Evan Z Macosko, Anindita Basu, Rahul Satija, James Nemesh, Karthik Shekhar, Melissa Goldman, Itay Tirosh, Allison R Bialas, Nolan Kamitaki, Emily M Martersteck, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214, 2015. (Cited on page 1)

Aaron McKenna and James A Gagnon. Recording development with single cell dynamic lineage tracing. *Development*, 146(12):dev169730, 2019. (Cited on page 8)

Kirill Neklyudov, Rob Brekelmans, Alexander Tong, Lazar Atanackovic, Qiang Liu, and Alireza Makhzani. A computational framework for solving wasserstein lagrangian flows, 2024. URL https://arxiv.org/abs/2310.10649. (Cited on page 8)

Matteo Pariset, Ya-Ping Hsieh, Charlotte Bunne, Andreas Krause, and Valentin De Bortoli. Unbalanced diffusion schrödinger bridge, 2023. URL https://arxiv.org/abs/2306.09099. (Cited on page 8)

Gabriel Peyré and Marco Cuturi. Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019. (Cited on page 1)

Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. (Cited on page 2)

Aram-Alexandre Pooladian and Jonathan Niles-Weed. Plug-in estimation of schrödinger bridges, 2024. URL https://arxiv.org/abs/2408.11686. (Cited on page 8)

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. *International Conference on Learning Representations (ICLR)*, 2023. (Cited on pages 4 and 8)

Xiaojie Qiu, Yan Zhang, Jorge D Martin-Rufino, Chen Weng, Shayan Hosseinzadeh, Dian Yang, Angela N Pogson, Marco Y Hein, Kyung Hoi Joseph Min, Li Wang, et al. Mapping transcriptomic vector fields of single cells. *Cell*, 185(4):690–711, 2022. (Cited on page 8)

Andrea Riba, Attila Oravecz, Matej Durik, Sara Jiménez, Violaine Alunni, Marie Cerciat, Matthieu Jung, Céline Keime, William M Keyes, and Nacho Molina. Cell cycle gene regulation dynamics revealed by rna velocity and deep-learning. *Nature communications*, 13(1):2865, 2022. (Cited on pages 2, 6, 7, and 12)

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. (Cited on page 1)

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

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. (Cited on page 8)

Rachel Shahan, Che-Wei Hsu, Trevor M Nolan, Benjamin J Cole, Isaiah W Taylor, Laura Greenstreet, Stephen Zhang, Anton Afanassiev, Anna Hendrika Cornelia Vlot, Geoffrey Schiebinger, et al. A single-cell arabidopsis root atlas reveals developmental trajectories in wild-type and cell identity mutants. *Developmental cell*, 57(4):543–560, 2022. (Cited on page 1)

Yuyang Shi, Valentin De Bortoli, Andrew Campbell, and Arnaud Doucet. Diffusion schrödinger bridge matching. *Advances in Neural Information Processing Systems*, 36, 2024. (Cited on pages 2, 4, and 8)

Alexander Tong, Jessie Huang, Guy Wolf, David van Dijk, and Smita Krishnaswamy. Trajectorynet: A dynamic optimal transport network for modeling cellular dynamics, 2020. URL https://arxiv.org/abs/2002.04461. (Cited on pages 2, 6, and 8)

Alexander Tong, Manik Kuchroo, Shabarni Gupta, Aarthi Venkat, Beatriz Perez San Juan, Laura Rangel, Brandon Zhu, John G Lock, Christine Chaffer, and Smita Krishnaswamy. Learning transcriptional and regulatory dynamics driving cancer cell plasticity using neural ode-based optimal transport. *bioRxiv preprint 2023.03.28.534644*, 2023a. (Cited on page 1)

Alexander 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. *AISTATS*, 2023b. (Cited on pages 4 and 8)

Alexander Tong, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Kilian Fatras, Guy Wolf, and Yoshua Bengio. Improving and generalizing flow-based generative models with minibatch optimal transport. *arXiv preprint arXiv:2302.00482*, 2023c. (Cited on pages 3, 4, 6, and 8)

C. H. Waddington. The epigenotype. *Endeavour*, 1:18–20, 1942. (Cited on page 1)

Daniel E Wagner and Allon M Klein. Lineage tracing meets single-cell omics: opportunities and challenges. *Nature Reviews Genetics*, 21(7):410–427, 2020. (Cited on page 8)

Jianhua Xing. Reconstructing data-driven governing equations for cell phenotypic transitions: integration of data science and systems biology. *Physical Biology*, 19(6):061001, 2022. (Cited on page 8)

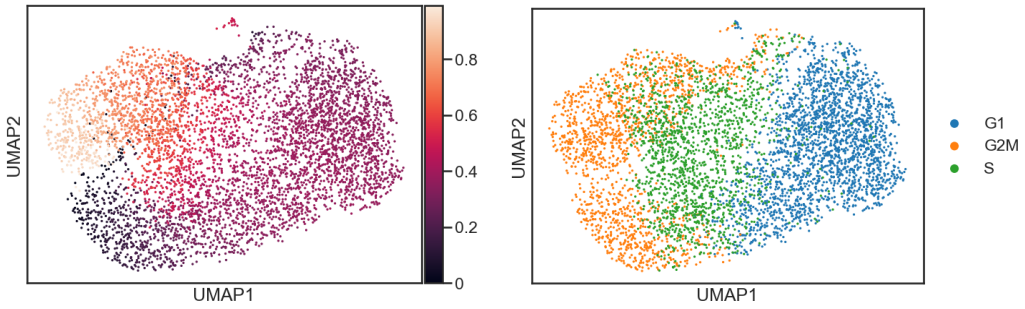
Zhenyi Zhang, Tiejun Li, and Peijie Zhou. Learning stochastic dynamics from snapshots through regularized unbalanced optimal transport, 2024. URL https://arxiv.org/abs/2410.00844. (Cited on page 8)

APPENDIX

A HUMAN FIBROBLASTS DATASET

We consider human fibroblasts dataset (Riba et al., 2022) that contains genomic information about 5,367 cells observed across a cell cycle. Cell data further contains information about cycling genes, and more specifically their RNA-velocities which we use to estimate reference velocity field $f_t(x_t)$.

Pre-processing. Data is pre-processed by selecting top d variable genes from the data. Further, we use `scvelo` package to compute imputed unspliced (Mu) and spliced (Ms) expressions as well velocity graph. Figure 4 shows a distribution of cell rotations and phases during a cell cycle process.



(a) Cell Rotations γ (b) Cell Phases

Figure 4: Human Fibroblasts Dataset.

B EXPERIMENTAL DETAILS

$\varphi_{t,\eta}(x_0, x_1)$ and $v_{t,\theta}(x_t, \eta)$ design. Both $\varphi_{t,\eta}(x_0, x_1)$ and $v_{t,\theta}(x_t, \eta)$ are designed as MLP models with 3 layers. We select MLP dimensions on the basis of number of chosen highly variable genes d:

- For $\varphi_{t,\eta}(x_0, x_1)$ we choose $d_{in} = 2 \times d$ and $d_{out} = d$
- For $v_{t,\theta}(x_t, \eta)$, we choose dimensions of $d_{in} = d$.

Dataset is split as $[80\%, 10\%, 10\%]$ across training, validation and test.

Training. All CURLY-FM and baseline experiments are run using $lr = 10^{-4}$ learning rate and Adam optimizer with default $\beta_1, \beta_2, and \epsilon$ values across three seeds and with 1,000 epochs split into 500 epochs to train $\varphi_{t,\eta}$ followed by 500 epochs to train $v_{t,\theta}$.

TrajectoryNet was run with 250 epochs with the Euler integrator with 20 timesteps per timepoint. We use 250 epochs to limit the experimental time to limit the number of function evaluations to roughly 5x that of simulation-free methods. We use a batch size of 256 samples. We use a Dormand-Prince 4-5 (dopri5) adaptive step size ODE solver to sample trajectories with absolute and relative tolerances of 10^{-4}. All experiments were conducted using a mixture of CPUs and A10 GPUs.