

---

# Efficient Trajectory Inference in Wasserstein Space Using Consecutive Averaging

---

**Amartya Banerjee**

Dept. of Computer Science  
UNC Chapel Hill

**Harlin Lee**

SDSS  
UNC Chapel Hill

**Nir Sharon**

Dept. of Applied Math  
Tel Aviv University

**Caroline Moosmüller**

Dept. of Mathematics  
UNC Chapel Hill

## Abstract

Capturing data from dynamic processes through cross-sectional measurements is seen in many fields, such as computational biology. Trajectory inference deals with the challenge of reconstructing continuous processes from such observations. In this work, we propose methods for B-spline approximation and interpolation of point clouds through consecutive averaging that is intrinsic to the Wasserstein space. Combining subdivision schemes with optimal transport-based geodesic, our methods carry out trajectory inference at a chosen level of precision and smoothness, and can automatically handle scenarios where particles undergo division over time. We prove linear convergence rates and rigorously evaluate our method on cell data characterized by bifurcations, merges, and trajectory splitting scenarios like *supercells*, comparing its performance against state-of-the-art trajectory inference and interpolation methods. The results not only underscore the effectiveness of our method in inferring trajectories but also highlight the benefit of performing interpolation and approximation that respect the inherent geometric properties of the data.

## 1 INTRODUCTION

Many dynamic processes yield cross-sectional observations at different time steps, which can be represented as a sequence of *point clouds*, or discrete probability measures. To understand the underlying process, we

need to first interpolate or approximate a likely path between these measurements. This problem of reconstructing continuous paths is called *trajectory inference*, which is seen in applications such as computational biology (Sha et al. (2023); Tong et al. (2020); Huguet et al. (2022); Schiebinger et al. (2019); Sae-lens et al. (2019)), computer graphics (Huang et al. (2022)), and control theory (Howe et al. (2022); Craig et al. (2024)). This paper considers the case where we only observe the state of a population as a whole, i.e. points in a point cloud do not have an inherent order. It may be the case that each sample is equivalent and interchangeable (e.g. drones in a swarm), or that there is no preservation of samples across time (e.g. cell measurement).

Recently, *optimal transport (OT)* (Peyré et al. (2019); Ambrosio et al. (2013); Villani (2008)) has shown to be effective in this problem, as it can provide a matching between two point clouds. By using OT to infer which sample at time step  $t_i$  could correspond to which one at  $t_j$ , trajectories can be constructed in a way that minimize cost or effort. In the seminal work by Schiebinger et al. (2019), a piecewise linear OT interpolation method is proposed to infer cell trajectories. Higher order piecewise polynomials (e.g. cubic splines) were introduced in Chen et al. (2018); Benamou et al. (2019) and extended by Chewi et al. (2021); Clancy and Suarez (2022); Justiniano et al. (2024). Deep learning-based methods that try to fit smoother paths include TrajectoryNet (Tong et al. (2020)) and MIOFlow (Huguet et al. (2022)), which train neural ODEs with OT-informed tools, and CFM (Tong et al. (2024)), which uses flow-based methods for constructing trajectories. In a similar spirit, WLF (Neklyudov et al. (2023)) proposes a unified approach for optimal transport based trajectory inference by optimizing a Lagrangian action functional in the space of densities, allowing one to incorporate problem-specific priors directly into the learned dynamics.

Our paper fits in this line of work in the following sense. We propose a set of algorithms for trajectory in-

ference (either interpolatory or approximate) on point clouds using consecutive OT-based geodesic averaging in the Wasserstein space. The methods are flexible and efficient as they allow us to tune the trajectories' smoothness and accuracy. Furthermore, they are *intrinsic to the geometry of the Wasserstein space*, which naturally handles mass-splitting phenomena.

We first propose the *Wasserstein Lane-Riesenfeld (WLR)* algorithm to approximate B-splines in the Wasserstein space. We leverage the algorithm by Lane and Riesenfeld (1980), which efficiently defines B-splines in  $\mathbb{R}^d$  through consecutive averaging. It has been shown that this algorithm can be extended to Riemannian manifolds in Wallner and Dyn (2005), and our method extends this idea to the Wasserstein space by using a OT-based geodesic.

While the WLR algorithm is the main method we propose, we observe that consecutive averaging can be applied more broadly beyond B-splines, and following ideas from Wallner and Dyn (2005), we showcase an interpolatory trajectory inference method based on the 4-point scheme by Dyn et al. (1987). We further discuss how our idea extends to any iterative method that uses consecutive averaging (e.g. any subdivision scheme).

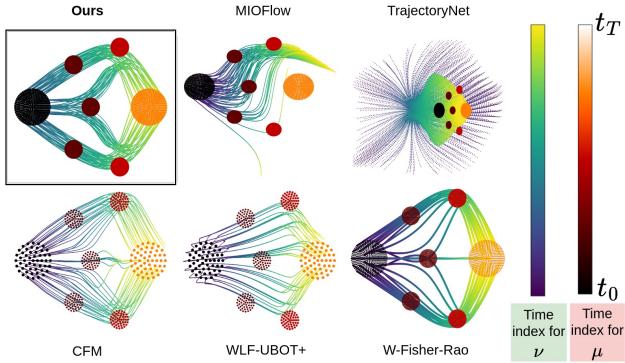


Figure 1: Our proposed WLR successfully performs trajectory inference on Converging Gaussian dataset, which has different number of points per time step. While others maintain a fixed number of trajectories from initialization to termination, our method can split trajectories automatically. The fact that WLR can naturally deal with mass splitting phenomena is one of its major benefits. See Sec. 5 for details.

Our contributions are:

1. Algorithmically, we define curves intrinsic to the Wasserstein geometry that do not depend on splines in  $\mathbb{R}^d$  as required in Chewi et al. (2021). Unlike Chewi et al. (2021), which only works on uniform data in  $\mathbb{R}^n$  by applying cubic splines af-

ter a fixed OT matching, our WLR approach handles non-uniform or unbalanced masses and easily adapts to arbitrary support spaces. Moreover, while methods like Benamou et al. (2019); Chen et al. (2018) formulate a global cubic spline minimization in Wasserstein space (with existence results largely demonstrated on Gaussian examples), solving such an optimization is not tractable in general. In contrast, WLR is a simple iterative scheme that remains fully intrinsic to Wasserstein geometry but is still straightforward to implement for real data. Our algorithms can therefore naturally deal with “trajectory splitting”, which is needed when mass is non-uniformly distributed over point clouds. This makes our methods comparable to other intrinsic Wasserstein spline methods, such as Chen et al. (2018); Benamou et al. (2019). Furthermore, our proposed algorithms can flexibly handle both approximation (via B-splines) and exact interpolation (via the 4-point scheme), and can be easily extended to other subdivision schemes.

2. Theoretically, our approach is consistent with existing theory in degenerate cases. When each point cloud consists of just one point, we recover classical B-spline theory (on  $\mathbb{R}^d$ ). When there are only two point clouds, we recover the OT assignment between those point clouds. Importantly, unlike prior spline-subdivision work on finite-dimensional Riemannian manifolds (e.g., Wallner and Dyn (2005) and Dyn and Sharon (2017a)), we prove convergence for an iterative subdivision scheme in the *infinite-dimensional* Wasserstein space.
3. Experimentally, WLR can accurately infer individual trajectories between uniform point clouds as for example needed in cell trajectories. Moreover, it can handle complex scenarios involving weighted mass splitting as in the case of *supercell* data. Subject to user’s choice of parameters, WLR is fast and stable compared to neural ODE-based methods such as Tong et al. (2020); Huguet et al. (2022), which can fail under stiff dynamics or unstable optimization. WLR is however slightly slower than extrinsic methods such as Chewi et al. (2021), as we need to solve more OT problems.

We also mention the concurrent publication Baccou and Liandrat (2024), which, similar to this paper, combines subdivision schemes with optimal transport. We present a more comprehensive comparison to other trajectory inference methods and provide new applications to cell data, including supercells.

In the following sections, we describe background on

optimal transport and splines (Sec. 2), state (Sec. 3) then demonstrate (Sec. 5) our proposed algorithms on various numerical experiments.

## 2 PRELIMINARIES

We define relevant notations and formulate the trajectory inference problem, then present background on optimal transport and splines in the context of that problem. We view optimal transport as a special case of trajectory inference where there are only two point clouds, and splines in  $\mathbb{R}^d$  as another edge case where there is only one point per point cloud.

### 2.1 Problem formulation

We consider probability measures  $\mu_t$  over  $\mathbb{R}^d$  depending on a continuous time parameter  $t$ . For the purpose of implementation,  $\mu_t$  is a discrete probability measure or *point cloud*, i.e. it is defined by a set of support points  $x_t = \{x_{t,i}\}_{i=1}^{n_t} \subset \mathbb{R}^d$ , and a probability vector  $a_t \in \mathbb{R}_+^{n_t}$  with  $\sum_{i=1}^{n_t} a_{t,i} = 1$ . Mathematically, this can be described as  $\mu_t = \sum_{i=1}^{n_t} a_{t,i} \delta_{x_{t,i}}$ ; see Peyré et al. (2019). Here,  $a_{t,i}$  describes the amount of mass located at  $x_{t,i}$ , and we allow for non-uniformly distributed mass. Note that the number of points  $n_t$  may change at different times  $t$ , and while the points in  $x_t$  are inherently unordered, we give them an arbitrary yet consistent ordering and represent them as a matrix  $x_t \in \mathbb{R}^{d \times n_t}$  in a slight abuse of notation.

Given  $T+1$  observations  $\mu_{t_0}, \dots, \mu_{t_T}$ , where  $t_0 < \dots < t_T$ , the aim of *trajectory inference* is to find probability measures  $\nu_t$  such that for  $j = 0, \dots, T$ ,  $\nu_{t_j} \approx \mu_{t_j}$  for approximation, or such that  $\nu_{t_j} = \mu_{t_j}$  for interpolation.

### 2.2 Discrete optimal transport and the Wasserstein space

In this section, we consider a single pair of point clouds  $(\mu_{t_0}, \mu_{t_1})$  as defined in Section 2.1. Discrete optimal transport (OT) studies the problem of transporting  $\mu_{t_0}$  to  $\mu_{t_1}$  in an “optimal” manner. Given a cost matrix  $C = [C_{ij}] \in \mathbb{R}_+^{n_{t_0} \times n_{t_1}}$  that describes how much effort is needed to move mass from one point in  $\mu_{t_0}$  to another in  $\mu_{t_1}$ , OT aims to find a coupling matrix  $\gamma \in \mathbb{R}_+^{n_{t_0} \times n_{t_1}}$  that attains

$$f := \min_{\gamma} \langle \gamma, C \rangle, \text{ s.t. } \gamma \mathbf{1} = a_{t_0}, \gamma^\top \mathbf{1} = a_{t_1}, \gamma \geq 0. \quad (1)$$

Sometimes called the Kantorovich formulation, this equation by Kantorovich (2006) is often solved via linear programming. This has been crucial in applications including economics, inverse problems, image processing, computational biology, and most re-

cently, machine learning (e.g. Courty et al. (2017)); see Peyré et al. (2019) for an excellent overview on computational OT. A popular choice for cost is  $C_{ij} = \|x_{t_0,i} - x_{t_1,j}\|^p$  for  $p \geq 1$ , in which case  $W_p(\mu_{t_0}, \mu_{t_1})$ , the  $p$ -Wasserstein distance between  $\mu_{t_0}$  and  $\mu_{t_1}$ , is defined as  $f^{1/p}$  using the minimum value attained in eq. (1).

When the space of probability measures on  $\mathbb{R}^d$  with finite  $p$ -th moment is equipped with  $W_p$ , we call it *Wasserstein space*. The Wasserstein space is a metric space and carries Riemannian-like structure. It can formally be considered an infinite dimensional Riemannian manifold (Otto (2001); Lott (2008)). Representing the point clouds to be on the Wasserstein space allows us to define useful geometric properties such as barycenters (Aguech and Carlier (2011)). Of particular interest to trajectory inference is the *geodesic* in the Wasserstein space. Once  $\gamma$  is obtained, the constant speed geodesic between  $\mu_{t_0}$  and  $\mu_{t_1}$  is given by the pushforward measure:

$$\mu_\alpha = (\pi_\alpha)_\sharp \gamma, \quad \text{where} \quad (2)$$

$$\pi_\alpha(x_{t_0,i}, x_{t_1,j}) = (1 - \alpha)x_{t_0,i} + \alpha x_{t_1,j} \quad \text{and } \alpha \in [0, 1].$$

$\mu_\alpha$  is the shortest path between  $\mu_{t_0}$  and  $\mu_{t_1}$  in the Wasserstein space. See Villani (2008); Santambrogio (2015); Ambrosio et al. (2013) for more background on Wasserstein geometry.

### 2.3 Splines through consecutive averaging

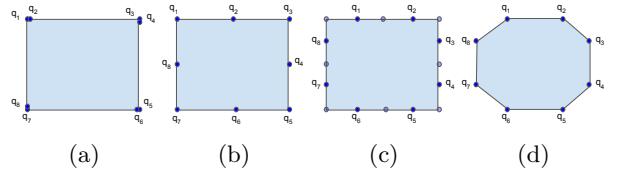


Figure 2: Illustration of the classical Lane-Riesenfeld (Alg. 4) with  $M = 2$ ,  $R = 1$ . This refinement step can be repeated to approximate a smooth cubic B-spline. (a) Doubling points; (b) First averaging; (c) Second averaging; (d) Refined points.

We discuss another edge case of Section 2.1, namely where each point cloud only has a single point, that is,  $n_t = 1$ . Splines are popular methods that fit piecewise polynomial curves through such data points in  $\mathbb{R}^d$ . In this work, we limit our discussion of splines from the viewpoint of *subdivision schemes*. These are iterative refinement methods that approximate continuous or smooth curves based on a sequence of initial data points, which is also called knots. See Appendix A for details.

We focus on one such method by Lane and Riesenfeld (1980) in this work (Appendix A Algorithm 4).

The Lane-Riesenfeld algorithm depends on two parameters—the degree  $M$ , which is related to the smoothness of the limit curve, and the refinement level  $R$ , which flexibly handles the accuracy of the approximation curve. As  $R \rightarrow \infty$ , we obtain a B-spline of degree  $M$  in  $\mathbb{R}^d$  which is  $C^{M-1}$  (Cavaretta et al. (1989)).

The algorithm contains two main phases: doubling of points and computing averages between consecutive points. An illustration is provided in Figure 2 with degree  $M = 2$  and refinement level  $R = 1$ . The algorithm is applied to a closed curve with four initial points, refining the shape iteratively.

Following Wallner and Dyn (2005); Dyn and Sharon (2017b), we can generalize the averaging phase via a linear averaging operator between two points  $y_i, y_j \in \mathbb{R}^d$  and  $\alpha \in [0, 1]$ :

$$\text{av}_\alpha(y_i, y_j) = (1 - \alpha)y_i + \alpha y_j. \quad (3)$$

With this operator, step (13) of Algorithm 4 becomes  $y_j^{(m)} \leftarrow \text{av}_{1/2}(y_j^{(m-1)}, y_{j+1}^{(m-1)})$ . It has been observed that the Lane-Riesenfeld algorithm can be further adapted to work on finite-dimensional Riemannian manifolds by replacing  $\text{av}_\alpha$  with the geodesic average on the Riemannian manifold (Wallner and Dyn (2005)).

### 3 WASSERSTEIN CURVE APPROXIMATION THROUGH CONSECUTIVE AVERAGING

We have now gathered two ingredients for trajectory inference. On one hand, we have a way to calculate the shortest path between two point clouds in the Wasserstein space (Section 2.2). On the other hand, we have a way to efficiently fit a curve through many points (not point *clouds*) in  $\mathbb{R}^d$  from Section 2.3. Our key insight is that if we replace the linear averaging operator in eq. (3) with the Wasserstein geodesic in eq. (2), we can use the consecutive averaging approach described in Section 2.3 to define smooth approximating curves through the point clouds in the Wasserstein space.

We propose the Wasserstein Lane-Riesenfeld (WLR) algorithm (c.f. Algorithm 1), which parallels the ideas of Dyn and Sharon (2017a,b) and incorporates an OT-based averaging step (c.f. Algorithm 2). This adjustment preserves the algorithm’s simplicity and scalability at a given refinement level, while effectively tackling the task of B-spline approximation. The defining characteristic of WLR is that the resulting curve is *intrinsic* to the Wasserstein geometry since the geodesic averaging step stays within the Wasserstein space. This is similar to the computationally heavier algorithms of Chen et al. (2018), but different from e.g.

#### Algorithm 1 Wasserstein Lane-Riesenfeld

---

```

1: procedure WLR( $[\mu_{t_j}]_{j=0}^T, R, M$ )
2:   input Point clouds to be refined  $[\mu_{t_j}]_{j=0}^T$ 
3:   input Refinement Level  $R \in \mathbb{Z}_+$ 
4:   input B-Spline degree  $M$ 
5:    $\nu^{(M)} \leftarrow [\mu_{t_j}]_{j=0}^T$ 
6:    $\triangleright$  Initialize point clouds to be doubled
7:   for  $r = 1$  to  $R$  do
8:     for  $j = 0$  to  $|\nu^{(M)}|$  do
9:        $\nu_{2j}^{(0)} \leftarrow \nu_j^{(M)}$   $\triangleright$  Doubling point clouds
10:       $\nu_{2j+1}^{(0)} \leftarrow \nu_j^{(M)}$ 
11:     end for
12:     Duplicate  $\nu_0^{(0)}$  and  $\nu_T^{(0)}$   $m$  times.
13:     for  $m = 1$  to  $M$  do
14:       for  $j = 0$  to  $|\nu^{(m-1)}|$  do
15:          $\nu_j^{(m)} \leftarrow \text{OT-av}(\nu_j^{(m-1)}, \nu_{j+1}^{(m-1)}, \frac{1}{2})$ 
16:          $\triangleright$  Repeated OT averaging
17:       end for
18:     end for
19:   end for
20:   return Refined point clouds  $\nu^{(M)}$ 
21: end procedure

```

---

Chewi et al. (2021), which is *extrinsic* in the sense that it constructs splines in  $\mathbb{R}^d$ . An important consequence is that unlike prior work, WLR has the ability to automatically handle non-uniform mass, mass splitting and bifurcations.

Before discussing numerical experiments, we illustrate our algorithm’s behavior on two dimensional toy data. Figure 3 plots three point clouds ordered (in time) from left to right. In the top row, all points have uniform mass. In the bottom row, mass is non-uniformly distributed in the second and third time steps. This demonstrates the following capacities of WLR.

**Trajectory inference** WLR produces natural trajectories between point clouds. In our implementation of WLR, we repeat the end point clouds  $\mu_{t_0}$  and  $\mu_{t_T}$  before averaging (see (11) in Algorithm 1) so that they are interpolated. Therefore, when a sequence of point clouds  $\nu$  is returned by WLR, we can set the first point cloud as  $\nu_{t_0} = \mu_{t_0}$ , the last point cloud as  $\nu_{t_T} = \mu_{t_T}$ , and evenly space out the remaining point clouds in between  $t_0, t_1, \dots, t_T$ . In addition, if we compute OT between consecutive pairs of point clouds in  $\nu$ , we are able to trace how mass from a single point in  $\mu_{t_0}$ , i.e. mass of weight  $a_{t_0, i}$  supported on  $x_{t_0, i}$ , travels from  $t_0$  to  $t_T$ . In other words, we are able to trace trajectories of individual samples in  $\mathbb{R}^d$  over time.

**Approximation and bending behavior** B-splines in  $\mathbb{R}^d$  do not interpolate given data. Rather, they find

**Algorithm 2** OT Averaging via Wasserstein Space Geodesic in eq. (2)

```

1: procedure OT-AV( $\mu_A, \mu_B, \alpha$ )
2:   input Point cloud  $\mu_A$ :  $(x_A, a_A)$ 
3:   input Point cloud  $\mu_B$ :  $(x_B, a_B)$ 
4:   input Averaging parameter  $\alpha \in [0, 1]$ 
5:    $\gamma \leftarrow \text{OT}(\mu_A, \mu_B)$ 
6:    $\triangleright$  Solve eq. (1) to obtain  $\gamma \in \mathbb{R}^{n_A \times n_B}$ .
7:    $x, a \leftarrow \text{null}$ 
8:   for every  $\gamma_{ij} > \epsilon$  do
9:     Concatenate  $(1 - \alpha) \cdot x_{A,i} + \alpha \cdot x_{B,j}$  to  $x$ 
10:    Append  $\gamma_{ij}$  to  $a$ 
11:   end for
12:    $a \leftarrow a/\text{sum}(a)$   $\triangleright$  Renormalization
13:   return Point cloud  $\mu_\alpha$  defined by  $(x, a)$ 
14: end procedure

```

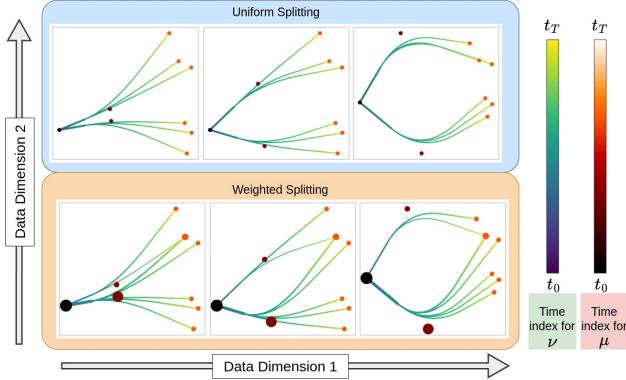


Figure 3: We provide an illustrative example in  $\mathbb{R}^2$  to demonstrate how WLR infers trajectories and handles both uniform (Top) and non-uniform (Bottom) masses. The size of the point is proportional to  $a_{t_j,i}$ , the mass of point  $i$  at time step  $t_j$ . In all cases, the trajectories respect the geometry of the Wasserstein space.

an approximating path between the knots. Figure 3 shows that this is still true for WLR and how the points in the second time step “bend” the trajectories like gravitational objects, and the bending is weighted by the mass located in each of the points.

**Splitting trajectories** This phenomenon naturally occurs in point clouds with non-uniform weights. Our algorithm can accommodate this automatically through the geodesic averaging procedure.

Lastly, we note that OT-based geodesic averaging can be applied to subdivision schemes beyond B-splines. Following ideas from Wallner and Dyn (2005), we describe an interpolatory trajectory inference method based on the 4-point scheme by Dyn et al. (1987) in Section 4.

### 3.1 Complexity analysis

WLR’s runtime depends on the B-spline degree  $M$ , number of initial point clouds  $T+1$ , refinement level  $R$ , and number of points per point cloud  $n_t$ . If the number of points stay the same, i.e.  $n_t = n$ , and there is a one-to-one mapping at each OT step, the overall complexity of WLR is  $\mathcal{O}(2^R \cdot (T+M) \cdot n^3 \cdot \log n)$ . Naturally, computation increases as  $M$  (indicating smoothness) and  $R$  (indicating approximation accuracy) increase. It is important to note that in the subdivision scheme literature,  $R \leq 8$  and  $M = 3$  suffice for most visual purposes. Still we provide a careful study of the interplay between these parameters for WLR runtime in Appendix B.

The most time consuming procedure is computing discrete OT in (5) of Algorithm 2, for which we use linear programming implemented in Python OT (Flamary et al. (2021)). In practice, we also vectorize lines (7–14) in Algorithm 2. Since the averaging operations only happen between consecutive point clouds, it is theoretically possible to parallelize WLR, but we leave that for future work.

### 3.2 Convergence analysis

For the purpose of the convergence proof, we consider the data  $\mu_{t_j}, j = 0, \dots, T$  in the  $p$ -Wasserstein space over  $\mathbb{R}^d$ ,  $\mathcal{W}_p(\mathbb{R}^d)$ , which is the space of all probability measure with finite  $p$ -th moment. This space is a metric space with distance given by the  $p$ -Wasserstein distance  $W_p$  in Section 2.2.

The  $p$ -Wasserstein distance satisfies the geodesic property: for any two measures,  $\mu_i$  and  $\mu_j$  we have for  $\alpha \in [0, 1]$ ,

$$W_p(\text{OT-av}_\alpha(\mu_i, \mu_j), \mu_j) = (1 - \alpha) W_p(\mu_i, \mu_j). \quad (4)$$

In addition,  $\mathcal{W}_p(\mathbb{R}^d)$  is a separable complete metric space since  $\mathbb{R}^d$  has these properties (Villani (2008)). Therefore, any Cauchy sequence in  $\mathcal{W}_p(\mathbb{R}^d)$  converges, which is the main property we need for the following convergence result.

Our subdivision scheme WLR generates sequences of point clouds, starting from the initial data  $\nu^{(0)} = [\nu_{t_j}^{(0)}]_{j=0}^T$  with  $\nu_{t_j}^{(0)} = \mu_{t_j}$ , and generating, by repeated refinements, the sequences  $\nu^{(R)} = [\nu_j^{(R)}]_{j=0}^{2^R T + 1}$ ,  $R \in \mathbb{N}$ . Therefore, informally, the limit  $\nu = \lim_{R \rightarrow \infty} \nu^{(R)}$  is the object we seek in Section 2.1. Technically, we associate each refinement level’s data  $[(j2^{-R}, \nu_j^{(R)})]_{j=0}^{2^R T + 1}$  with the piecewise geodesic interpolant in  $\mathcal{W}_p(\mathbb{R}^d)$ , which is defined using the Wasserstein geodesic of eq. (2) and is calculated by Algorithm 2. The piece-

wise geodesic interpolant is

$$N^{(R)}(t) = \text{OT-av}_t(\nu_j^{(R)}, \nu_{j+1}^{(R)}), \quad t \in [j2^{-R}, (j+1)2^{-R}).$$

Then, the refinement scheme (e.g. WLR) is called convergent if the sequence of piecewise geodesic interpolants  $\{N^{(R)}(t)\}_{R \in \mathbb{Z}_+}$  converges uniformly (in  $t$ ) to a limit curve, where the curve takes values in the Wasserstein space  $\mathcal{W}_p(\mathbb{R}^d)$ . This is called the *limit of the subdivision scheme*.

The result for our WLR schemes reads:

**Theorem 1.** Consider a sequence of measures  $\mu_{t_0}, \dots, \mu_{t_T}$  with  $t_0 < \dots < t_T$ . Define the initial data by  $\nu^{(0)} = [\nu_{t_j}^{(0)}]_{j=0}^T$  with  $\nu_{t_j}^{(0)} = \mu_{t_j}$ , and choose a smoothness degree  $M$ . Apply WLR (i.e. call Algorithm 1 for the initial data, smoothness  $M$ , and refinement level  $R$ ) to obtain a  $R$ -times refined sequence of measures  $\nu^{(R)}$ . As  $R \rightarrow \infty$ , WLR converges to a continuous limit curve  $\nu$  in  $\mathcal{W}_p(\mathbb{R}^d)$ , where convergence is in the sense of subdivision schemes defined above.

For the obtained piecewise geodesic connecting the sequence of measures in the  $R$ -th iteration step ( $N^{(R)}(t)$ ), we obtain a linear convergence rate:

$$W_p(N^{(\infty)}(t), N^{(R)}(t)) \leq C \frac{1}{2^{R-1}} = \mathcal{O}(2^{-R}), \quad (5)$$

where constant  $C$  is a constant independent of  $R$ .

*Proof.* The proof follows from adapting the results of (Dyn and Sharon, 2017a, Corollary 3.3) for finite-dimensional Riemannian manifolds to the Wasserstein space.

First, we recall the notion

$$\Delta\nu^{(R)} := \sup_{j \in \mathbb{Z}} W_p(\nu_j^{(R)}, \nu_{j+1}^{(R)}), \quad (6)$$

and assume that the initial data is bounded, that is,  $\Delta\nu^{(0)} < \infty$ . Following the arguments in the proof of Dyn and Sharon (2017a), we get that there exist  $\eta \in (0, 1)$  such that

$$\Delta\nu^{(R)} \leq \eta \Delta\nu^{(R-1)}, \quad R \in \mathbb{N}.$$

We term the factor  $\eta$  *contractivity factor*. Using the geodesic property of eq. (4), one can actually show that WLR has a contractivity factor of  $\eta = 1/2$ . In fact, this value equals the contractivity factor of the original linear Lane-Riesenfeld scheme, which is the optimal contractivity for linear subdivision schemes, see Dyn and Sharon (2024).

To follow the proof of Dyn and Sharon (2017a), we also need an additional mild condition, the *displacement-safe property*, which is immediately satisfied for WLR,

because it is satisfied for the standard linear Lane-Riesenfeld algorithm. This, together with the contractivity property, implies that there exists a constant  $C$ , independent of  $R$ , such that  $W_p(N^{(R+1)}(t), N^R(t)) \leq C\eta^R$ . And so the convergence follows, since for any  $\ell \in \mathbb{N}$ , we have

$$\begin{aligned} W_p(N^{(R+\ell)}(t), N^R(t)) &\leq C\eta^R(1 + \eta + \dots + \eta^\ell) \\ &\leq \frac{C}{1 - \eta}\eta^R. \end{aligned}$$

Namely,  $\{N^{(R)}(t)\}_{R \in \mathbb{Z}_+}$  is a Cauchy sequence in the Wasserstein space, and since this space is complete, the sequence converges to a limit. By letting  $\ell \rightarrow \infty$ , we obtain a bound on the difference between the limit curve and the  $R$ -th iteration:

$$W_p(N^{(\infty)}(t), N^{(R)}(t)) \leq \frac{C}{2^{R-1}} = \mathcal{O}(2^{-R}). \quad \square$$

## 4 EXTENSION OF WLR AND THE FOUR-POINT SCHEME

Although the Wasserstein LaneRiesenfeld algorithm (Algorithm 1, an extension of Algorithm 4) is our primary focus, the approach introduced in Section 3 naturally applies to any subdivision scheme. In particular, replacing each linear average with an OT-av results in a broad family of both interpolatory and approximating schemes in Wasserstein space. We choose WLR to illustrate this idea because it can be tuned to different smoothness levels through the parameter  $M$ , and it exhibits strong convergence properties Dyn and Sharon (2017b,a), which we extend in Section 3.2.

As an interpolatory alternative, we also adapt the classical four-point scheme of Dyn et al. (1987) by substituting all linear averages with OT-av. The resulting Wasserstein four-point scheme is described in Algorithm 3 and demonstrated in Figure 4 on two-dimensional circular, non-uniform Gaussian data. Notably, its contractivity factor is  $\eta = 4\omega + \frac{1}{2}$  Dyn and Sharon (2017c). For the parameter  $\omega = \frac{1}{16}$ , we have  $\eta = \frac{3}{4}$ . By an argument analogous to the convergence proof of WLR in Section 3.2, this contractivity guarantees convergence of the four-point scheme in Wasserstein space.

**Algorithm 3** 4-Point Scheme in Wasserstein Space

```

1: procedure FOURPOINTSCHEME( $[\mu_{t_j}]_{j=0}^T, R$ )
2:   input Point clouds to be refined  $[\mu_{t_j}]_{j=0}^T$ .
3:   input Refinement level  $R \in \mathbb{Z}_+$ .
4:    $\nu^{(0)} \leftarrow [\mu_{t_j}]_{j=0}^T$ 
5:    $w \leftarrow \frac{1}{16}$   $\triangleright w$  is the weighting parameter
6:   for  $r = 1$  to  $R$  do
7:     for  $j = 2$  to  $(|\nu^{(r-1)}| - 1)$  do
8:        $\nu_{2j}^{(r-1)} \leftarrow \nu_j^{(r-1)}$ 
9:     end for
10:    for  $j = 2$  to  $(|\nu^{(r-1)}| - 2)$  do
11:       $X_A \leftarrow \text{OT-AV}(\nu_j^{(r-1)}, \nu_{j-1}^{(r-1)}, -2w)$ 
12:       $X_B \leftarrow \text{OT-AV}(\nu_{j+1}^{(r-1)}, \nu_{j+2}^{(r-1)}, -2w)$ 
13:       $\nu_{2j+1}^{(r)} \leftarrow \text{OT-AV}(X_A, X_B, \frac{1}{2})$ 
14:    end for
15:     $\nu^{(r)} \leftarrow [\nu_4^{(r)}, \dots, \nu_{2(|\nu^{(r-1)}|-1)}^{(r)}]$ 
16:  end for
17:  return Refined point clouds  $\nu^{(R)}$ .
18: end procedure

```

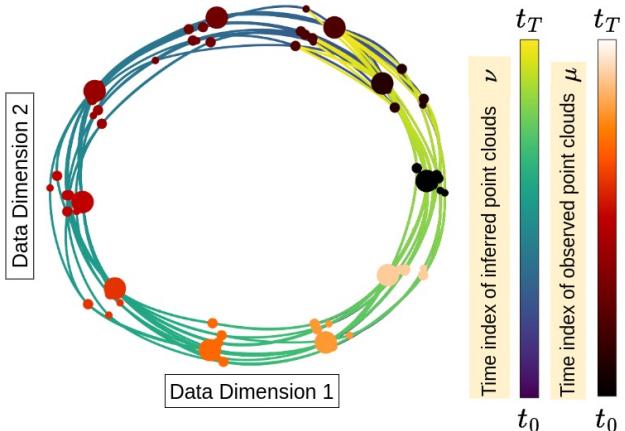


Figure 4: Illustrative example in  $\mathbb{R}^2$  demonstrating exact interpolation via the Wasserstein four-point scheme on weighted circular Gaussians.

## 5 EXPERIMENTS

Section 5.1 highlights settings where mass splitting occurs or where point clouds across time steps have different number of points, which WLR is uniquely suited for. We then evaluate the performance of WLR against other trajectory inference methods on uniformly distributed mass in Section 5.2. In both settings, performance is evaluated with the leave-one-out experiment as suggested in Huguet et al. (2022). For each algorithm, we withhold data at a common intermediate time step during the ‘training’ phase. Then, in

the ‘inference’ phase, we predict the outcomes for the omitted time step and document the 1-Wasserstein ( $W_1$ ) distance between the predicted and the omitted point clouds. We compared WLR against six state-of-the-art algorithms: MIOFlow (Huguet et al. (2022)), TrajectoryNet (Tong et al. (2020)), Conditional Flow Matching (Tong et al. (2024)), F & S algorithm (Chewi et al. (2021)), Wasserstein-Fisher-Rao (Clancy and Suarez (2022)), and Wasserstein Lagrangian Flow (Neklyudov et al. (2023)). We repeated each experiment by subsampling within each dataset using five different random seeds. When possible, we optimized code and tuned parameters of other methods for fair comparison. See Appendix D for all experimental setup and dataset details. Our Python code is available at <https://github.com/amartya21/Wasserstein-Trajectory-Inference>.

### 5.1 Splitting trajectories and non-uniform weights in supercells

In this section, we designed two experiments to demonstrate the effectiveness of WLR in handling these mass splitting scenarios. The first uses the Converging Gaussian data by Clancy and Suarez (2022), which has different number of points with uniformly distributed mass per time step. The inferred trajectories from the use of all time steps in the data sets are visualized in Figure 1, and the leave-one-out results are reported in Table 1, in which WLR performs best or second best numerically.

Method	Runtime (sec) ↓	Leave-one-out $W_1 \downarrow$	Mean $W_1 \downarrow$
WLR (Ours)	$63.50 \pm 118.3$	<b><math>0.70 \pm 0.0</math></b>	$0.29 \pm 0.3$
WFR	<b><math>2.72 \pm 0.85</math></b>	$0.92 \pm 0.24$	$0.44 \pm 0.13$
CFM	$167.84 \pm 9.37$	$0.72 \pm 0.01$	<b><math>0.27 \pm 0.29</math></b>
MIOFlow	$32.25 \pm 0.37$	$2.78 \pm 2.64$	$2.20 \pm 1.89$
TrajectoryNet	$1310.13 \pm 55.54$	$1.94 \pm 0.03$	$2.22 \pm 0.51$
WLF-SB	$92.87 \pm 10.10$	$25.64 \pm 19.04$	$25.61 \pm 18.89$
WLF-UBOT+	$78.94 \pm 9.00$	$0.95 \pm 0.07$	$1.04 \pm 0.05$

Table 1: WLR performs the **best** or second best on Converging Gaussian, which has a different number of points per time step with uniformly distributed mass. See Figure 1 for corresponding visualizations.

The second experiment is on CITE-Seq supercells, where the mass distribution is not uniform. In single-cell genomics, the concept of *supercells* or *metacells* has gained popularity in recent years as a way to simplify scRNA-seq data (Baran et al. (2019); Chen et al. (2021); Bilous et al. (2022)). Supercells are formed by merging transcriptionally similar cells by creating a  $k$ -nearest neighbor graph based on gene expression. By aggregating similar cells, supercells can mitigate the effects of technical noise and dropout events, provide a more robust representation of the underlying bio-

logical state, and capture rare cell types or transient states that might be overlooked in individual cell analysis. This simplification aids in downstream analyses, such as visualization and differential expression analysis and allows for a more comprehensive understanding of the cellular heterogeneity within a sample.

Creation of supercells naturally gives rise to data with different numbers of points per time step, i.e.  $n_t$  that changes over  $t$ , which then leads to splitting trajectories. Existing trajectory inference methods struggle to accurately model such scenarios, as they typically rely on unbalanced transport and assume uniformly distributed mass at each time step. These assumptions do not adequately capture the complexity of real-world systems where masses can split unevenly or follow diverse paths. For example, it is intuitive to assign weights to supercells such that the weights are proportional to the number of cells that have been merged to create each supercell.

Existing trajectory inference methods typically struggle with such scenarios as they rely on unbalanced transport and assume uniform mass distributions. To the best of our knowledge, there are currently no neural-network based or spline-based methods in the literature that address trajectory inference with weighted mass distributions by accounting for splitting. As a result, Figure 5 and Table 2 report standalone values.

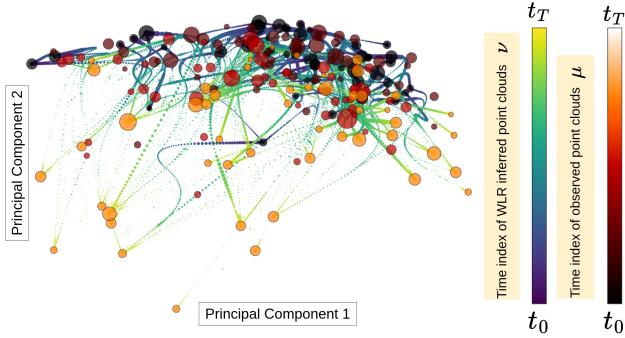


Figure 5: WLR produces smooth trajectories on CITE-seq supercells with non-uniformly distributed mass by automatically splitting trajectories. The size of the points is proportional to  $a_{t_j,i}$ , the mass of point  $i$  at time step  $t_j$ . Only a subset ( $n_t = 50$ ) is visualized for clarity.

## 5.2 Constant number of points with uniform mass distribution

This section applies WLR to the setting where the number of points  $n_t$  stays constant with uniform mass distribution. Our primary objective in these experi-

$R$	Time	Leave-one-out $W_1 \downarrow$
4	25.17 mins	23.90701
5	5.77 hours	23.56288
6	1.19 days	21.49304

Table 2: WLR is capable of trajectory inference on CITE-seq supercells at varying refinement levels.

ments is not strictly to surpass the existing state-of-the-art methods but rather to demonstrate that our approach is competitive and provides distinct advantages when dealing with more complex scenarios such as in the case of supercells.

We use a wide range of simulated and real datasets for this task: Diverging Gaussian, Petal (Huguet et al. (2022)), Dyngen Tree and Cycle (Cannoodt et al. (2021)), and CITE-seq (Stoeckius et al. (2017)). These datasets capture the natural dynamics observed in cellular differentiation such as bifurcations and merges.

These experiments are split up into two categories. In the first set of results in Table 4, we compare B-spline based approximation methods and neural network approaches, both of which provide smooth approximations rather than exact interpolations of the input point clouds. We also perform trajectory inference using all time steps in the datasets and visualize WLR trajectories in Figure 6; visualization of other methods are deferred to Figure 10 in Appendix E. In the second set of results in Table 5, we compare the interpolatory methods for trajectory inference, namely the original F & S method proposed by Chewi et al. (2021) and our four point scheme described in Algorithm 3.

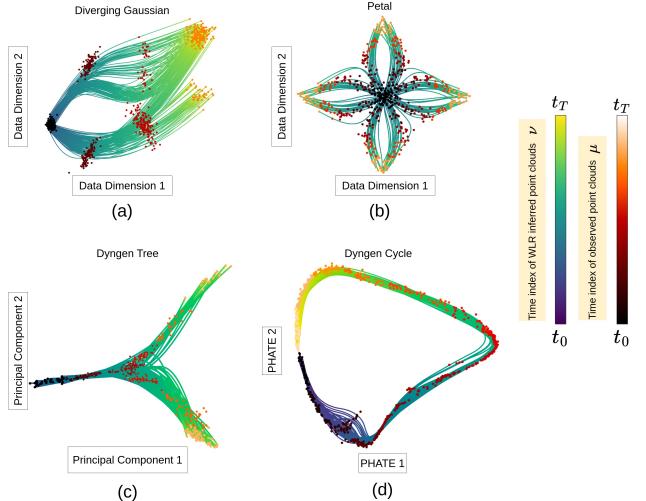


Figure 6: WLR produces smooth and visually convincing trajectories on four synthetic datasets that mimic cell differentiation.

### 5.3 Discussion

Our experiments demonstrate that WLR performs consistently well across diverse datasets in terms of runtime, leave-one-out  $W_1$  distance, leave-one-out MSE, and qualitative trajectory smoothness. WLR can dependably model complex dynamics such as bifurcations, merges, and trajectory splitting with both uniform and non-uniform weights. In comparison, deep learning methods tend to require long training times and exhibit optimization instability. Deep learning-based methods also have inherent limitations when it comes to handling mass splitting phenomena. These approaches maintain a fixed number of trajectories from initialization to termination, determined by the initial configuration, making it challenging to naturally model scenarios where particle masses split or merge. The only way these methods can approximate mass splitting behavior is by working in the unbalanced optimal transport setting. However, this is an indirect workaround that does not fully capture the underlying splitting dynamics. In contrast, our method can directly handle mass splitting phenomena without requiring such approximations, making it more suitable for applications where trajectory branching is a natural part of the system dynamics. Note that quantitative performance does not always align with qualitative plausibility of the trajectories; see Figure 10 and Figure 1. F&S produces results similar to—often identical to—WLR and offers faster runtimes. However, F&S is not designed to handle trajectory splitting or varying number of points across time steps. The Wasserstein-Fisher-Rao splines improve upon F&S via unbalanced transport, but optimizing for a different geometry can lead to crossing trajectories or omitted time steps.

## 6 CONCLUSIONS

We present a simple framework for trajectory inference on point clouds that respects the inherent geometry in the data. Our method builds on subdivision scheme methods with optimal transport-based geodesic to efficiently and provably approximate (or interpolate) curves in the Wasserstein space. Unlike prior work based on neural ODEs, which may fail under stiff dynamics or rely heavily on hyperparameters, our method consistently provides smooth and accurate trajectories in a user-controlled way. While our current implementation of WLR uses a straightforward linear programming approach to solve OT problems sequentially, we believe that parallelization could substantially reduce runtime. Since WLR requires solving OT only between consecutive time steps, these subproblems could, in principle, be run in parallel, although

special care would be needed at the boundaries where sub-trajectories meet. In future work, improving scalability and efficiency could greatly extend WLRs applicability to large-scale datasets. We demonstrate WLR’s capability to handle high-dimensional complex biological datasets like CITE-seq supercells and show the ability to carry out trajectory inference in uniform and weighted mass splitting scenarios. Moreover, based on our preliminary results in Appendix C, we posit that employing our framework as an initialization step could enhance the convergence rates of various neural network-based trajectory inference methods.

### 6.1 Acknowledgements

We thank the authors of Chewi et al. (2021) and Justiniano et al. (2024) for sharing their code with us. We also thank Natalie Stanley for helpful discussions and introducing us to supercells. NS is partially supported by the NSF-BSF award 2019752 and the DFG award 514588180. CM is partially supported by NSF awards DMS-2306064 and DMS-2410140, and by a seed grant from the School of Data Science and Society at UNC. There are no competing interests to declare.

### References

- Aguech, M. and Carlier, G. (2011). Barycenters in the wasserstein space. *SIAM Journal on Mathematical Analysis*, 43(2):904–924.
- Ambrosio, L., Bressan, A., Helbing, D., Klar, A., Zuazua, E., Ambrosio, L., and Gigli, N. (2013). A users guide to optimal transport. *Modelling and Optimisation of Flows on Networks: Cetraro, Italy 2009*, Editors: Benedetto Piccoli, Michel Rascle, pages 1–155.
- Baccou, J. and Liandrat, J. (2024). Subdivision scheme for discrete probability measure-valued data. *Applied Mathematics Letters*, 158:109233.
- Baran, Y., Bercovich, A., Sebe-Pedros, A., Lubling, Y., Giladi, A., Chomsky, E., Meir, Z., Hoichman, M., Lifshitz, A., and Tanay, A. (2019). Metacell: analysis of single-cell rna-seq data using k-nn graph partitions. *Genome biology*, 20:1–19.
- Benamou, J.-D., Gallouët, T. O., and Vialard, F.-X. (2019). Second-order models for optimal transport and cubic splines on the Wasserstein space. *Found Comp Math*, 19:113–1143.
- Bilous, M., Tran, L., Cianciaruso, C., Gabriel, A., Michel, H., Carmona, S. J., Pittet, M. J., and Gfeller, D. (2022). Metacells untangle large and complex single-cell transcriptome networks. *BMC bioinformatics*, 23(1):336.

- Cannoodt, R., Saelens, W., Deconinck, L., and Saeys, Y. (2021). Spearheading future omics analyses using dyngen, a multi-modal simulator of single cells. *Nature Communications*, 12(1):3942.
- Cavaretta, A., Dahmen, W., and Micchelli, C. (1989). *Stationary Subdivision*. Preprint / A: Fachbereich Mathematik. Freie Universität Berlin. Fachbereich Mathematik.
- Chen, L., He, Q., Zhai, Y., and Deng, M. (2021). Single-cell rna-seq data semi-supervised clustering and annotation via structural regularized domain adaptation. *Bioinformatics*, 37(6):775–784.
- Chen, Y., Conforti, G., and Georgiou, T. T. (2018). Measure-valued spline curves: An optimal transport viewpoint. *SIAM Journal on Mathematical Analysis*, 50(6):5947–5968.
- Chewi, S., Clancy, J., Le Gouic, T., Rigollet, P., Stepaniants, G., and Stromme, A. (2021). Fast and smooth interpolation on Wasserstein space. In *International Conference on Artificial Intelligence and Statistics*, pages 3061–3069. PMLR.
- Clancy, J. and Suarez, F. (2022). Wasserstein-Fisher-Rao splines. arXiv:2203.15728.
- Courty, N., Flamary, R., Habrard, A., and Rakotomamonjy, A. (2017). Joint distribution optimal transportation for domain adaptation. *Advances in neural information processing systems*, 30.
- Craig, K., Elamvazhuthi, K., and Lee, H. (2024). A blob method for mean field control with terminal constraints. *arXiv preprint arXiv:2402.10124*.
- Dyn, N., Levin, D., and Gregory, J. A. (1987). A 4-point interpolatory subdivision scheme for curve design. *Computer Aided Geometric Design*, 4(4):257–268.
- Dyn, N. and Sharon, N. (2017a). A global approach to the refinement of manifold data. *Mathematics of Computation*, 86(303):pp. 375–395.
- Dyn, N. and Sharon, N. (2017b). Manifold-valued subdivision schemes based on geodesic inductive averaging. *Journal of Computational and Applied Mathematics*, 311:54–67.
- Dyn, N. and Sharon, N. (2017c). Manifold-valued subdivision schemes based on geodesic inductive averaging. *Journal of Computational and Applied Mathematics*, 311:54–67.
- Dyn, N. and Sharon, N. (2024). Improving the convergence analysis of linear subdivision schemes. *arXiv preprint arXiv:2405.09414*.
- Flamary, R., Courty, N., Gramfort, A., Alaya, M. Z., Boisbunon, A., Chambon, S., Chapel, L., Corenflos, A., Fatras, K., Fournier, N., et al. (2021). Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8.
- Howe, N., Dufort-Labbé, S., Rajkumar, N., and Bacon, P.-L. (2022). Myriad: a real-world testbed to bridge trajectory optimization and deep learning. *Advances in Neural Information Processing Systems*, 35:29801–29815.
- Huang, X., Wang, Y., Guizilini, V., Ambrus, R., Gaidon, A., and Solomon, J. (2022). Representation learning for object detection from unlabeled point cloud sequences. In *Conference on Robot Learning*.
- Huguet, G., Magruder, D. S., Tong, A., Fasina, O., Kuchroo, M., Wolf, G., and Krishnaswamy, S. (2022). Manifold interpolating optimal-transport flows for trajectory inference. *Advances in Neural Information Processing Systems*, 35:29705–29718.
- Justiniano, J., Rumpf, M., and Erbar, M. (2024). Approximation of splines in wasserstein spaces. *ESAIM: Control, Optimisation and Calculus of Variations*, 30:64.
- Kantorovich, L. V. (2006). On the translocation of masses. *Journal of Mathematical Sciences*, 133(4):1381–1382.
- Lane, J. M. and Riesenfeld, R. F. (1980). A theoretical development for the computer generation and display of piecewise polynomial surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2(1):35–46.
- Lott, J. (2008). Some geometric calculations on Wasserstein space. *Comm. Math. Phys.*, 277:423–437.
- Moon, K. R., Van Dijk, D., Wang, Z., Gigante, S., Burkhardt, D. B., Chen, W. S., Yim, K., Elzen, A. v. d., Hirn, M. J., Coifman, R. R., et al. (2019). Visualizing structure and transitions in high-dimensional biological data. *Nature biotechnology*, 37(12):1482–1492.
- Neklyudov, K., Brekelmans, R., Tong, A., Atanackovic, L., Liu, Q., and Makhzani, A. (2023). A computational framework for solving wasserstein lagrangian flows.
- Otto, F. (2001). The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174.
- Peyré, G., Cuturi, M., et al. (2019). Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607.
- Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nature biotechnology*, 37(5):547–554.

- Santambrogio, F. (2015). Optimal transport for applied mathematicians. *Birkhäuser, NY*, 55(58-63):94.
- Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., et al. (2019). Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943.
- Sha, Y., Qiu, Y., Zhou, P., and Nie, Q. (2023). Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. *Nature Machine Intelligence*, pages 1–15.
- Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P. K., Swerdlow, H., Satija, R., and Smibert, P. (2017). Large-scale simultaneous measurement of epitopes and transcriptomes in single cells. *Nature methods*, 14(9):865.
- Tong, A., FATRAS, K., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Wolf, G., and Bengio, Y. (2024). Improving and generalizing flow-based generative models with minibatch optimal transport. *Transactions on Machine Learning Research*. Expert Certification.
- Tong, A., Huang, J., Wolf, G., Van Dijk, D., and Krishnaswamy, S. (2020). TrajectoryNet: A dynamic optimal transport network for modeling cellular dynamics. In *International conference on machine learning*, pages 9526–9536. PMLR.
- Villani, C. (2008). *Optimal Transport: Old and New*, volume 338. Springer Science & Business Media.
- Wallner, J. and Dyn, N. (2005). Convergence and  $C^1$  analysis of subdivision schemes on manifolds by proximity. *Computer Aided Geometric Design*, 22(7):593–622.

## Checklist

1. For all models and algorithms presented, check if you include:
  - (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
  - (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
  - (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]
2. For any theoretical claim, check if you include:
  - (a) Statements of the full set of assumptions of all theoretical results. [Yes]

- (b) Complete proofs of all theoretical results. [Yes]
- (c) Clear explanations of any assumptions. [Yes]
3. For all figures and tables that present empirical results, check if you include:
  - (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
  - (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
  - (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
  - (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
  - (a) Citations of the creator If your work uses existing assets. [Yes]
  - (b) The license information of the assets, if applicable. [Yes]
  - (c) New assets either in the supplemental material or as a URL, if applicable. [Yes]
  - (d) Information about consent from data providers/curators. [Yes]
  - (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]
5. If you used crowdsourcing or conducted research with human subjects, check if you include:
  - (a) The full text of instructions given to participants and screenshots. [Not Applicable]
  - (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
  - (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

---

## Supplementary Material: Efficient Trajectory Inference in Wasserstein Space Using Consecutive Averaging

---

### A SUBDIVISION SCHEMES AND THE LANE RIESENFIELD ALGORITHM

**Algorithm 4** B-spline Approximation Algorithm by Lane and Riesenfeld (1980)

---

```

1: procedure LANERIESENFELD( $[x_j]_{j \in \mathbb{Z}}, R, M$ )
2:   input Points to be refined  $[x_j]_{j \in \mathbb{Z}}$ .  $x_j \in \mathbb{R}^d$ .
3:   input Refinement level  $R \in \mathbb{Z}_+$ .
4:   input Degree  $M$  of B-splines to be approximated.
5:    $y^{(M)} \leftarrow [x_j]_{j \in \mathbb{Z}}$ 
6:        $\triangleright$  Initializing points to be doubled.
7:   for  $r = 1$  to  $R$  do
8:     for  $j \in \mathbb{Z}$  do
9:        $y_{2j}^{(0)} \leftarrow y_j^{(M)}$   $\triangleright$  Doubling points.
10:       $y_{2j+1}^{(0)} \leftarrow y_j^{(M)}$ 
11:    end for
12:    for  $m = 1$  to  $M$  do
13:      for  $j \in \mathbb{Z}$  do
14:         $y_j^{(m)} \leftarrow \frac{1}{2}(y_j^{(m-1)} + y_{j+1}^{(m-1)})$ 
15:         $\triangleright$  Repeated averaging.
16:      end for
17:    end for
18:  end for
19:  return Refined points  $[y_j^{(M)}]_{j \in \mathbb{Z}}$ .
20: end procedure

```

---

We briefly describe subdivision schemes and their adaptation to the nonlinear domain via nonlinear averaging. Given an ordered sequence of points,  $\mathbf{p} = \{p_i\}_{i=0}^T \subset \mathbb{R}^d$ , we consider a linear *refinement rule* of the form

$$S(\mathbf{p})_j = \sum_{i \in \mathbb{Z}} a_{j-2i} p_i, \quad j \in \mathbb{Z}, \quad (7)$$

where  $\{a_i\}_{i \in I} \subset \mathbb{R}$  is the *mask* of the refinement rule  $S$ . We assume the mask has finitely many non-zero elements, that is,  $|I| < \infty$ . Note that to avoid any technical boundary treatment, we extend the sequence  $\mathbf{p}$  to all of  $\mathbb{Z}$  by repeating the starting point  $p_0$  and the endpoint  $p_T$ , i.e., we consider  $\dots, p_0, p_0, p_1, \dots, p_{T-1}, p_T, p_T, \dots$

To obtain finer and finer sequences of  $\mathbf{p}$ , we consider  $S^k(\mathbf{p})$ , i.e.  $S$  is applied  $k$ -times to  $\mathbf{p}$ . The repeated application of refinement rules is called a subdivision scheme. A subdivision scheme is called *convergent* if the piecewise linear interpolant to the data  $(i2^{-k}, S^k(\mathbf{p})_i)$  converges uniformly to a continuous limit curve with values in  $\mathbb{R}^d$ . See Cavaretta et al. (1989) for more details.

It has been observed in (Wallner and Dyn, 2005, Theorem 1) that any affinely invariant ( $\sum_i a_{j-2i} = 1$  for all  $j$ ) refinement rule as in (equation 7) can be (non-uniquely) expressed via the averaging operator  $\text{av}$  from eq. (3). Wallner and Dyn (2005) further observed that this fact can be used to define refinement rules for sequences  $\mathbf{p}$  in Riemannian manifolds by replacing  $\text{av}$  by the geodesic average.

## B DEPENDENCE ON PARAMETERS

The runtime results in Figure 7 were obtained by running WLR for varying degrees ( $M$ ) and refinement levels ( $R$ ) on the Diverging Gaussian dataset. It is evident that the runtime increases significantly as  $R$  increases, particularly beyond  $R = 7$ . While  $M$  also influences runtime, its impact is less pronounced. This is consistent with our analysis in Section 3.1 that complexity grows exponentially with  $R$  and linearly with  $M$ . In practice,  $R = 7, M = 3$  is more than enough to achieve smooth trajectories in most cases.

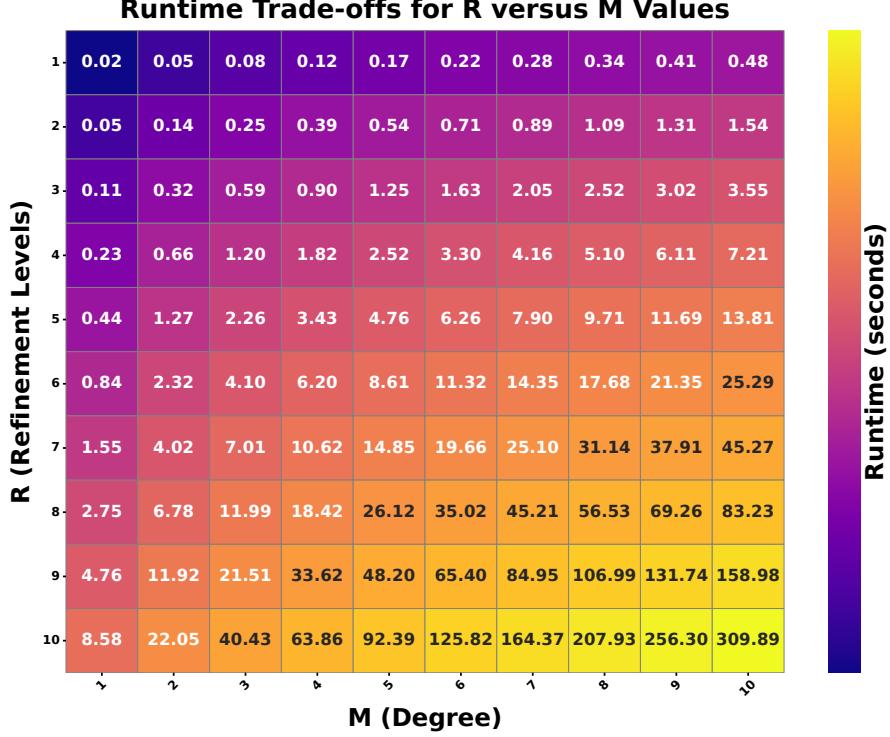


Figure 7: WLR runtimes on the Diverging Gaussian dataset with varying degrees and refinement levels.

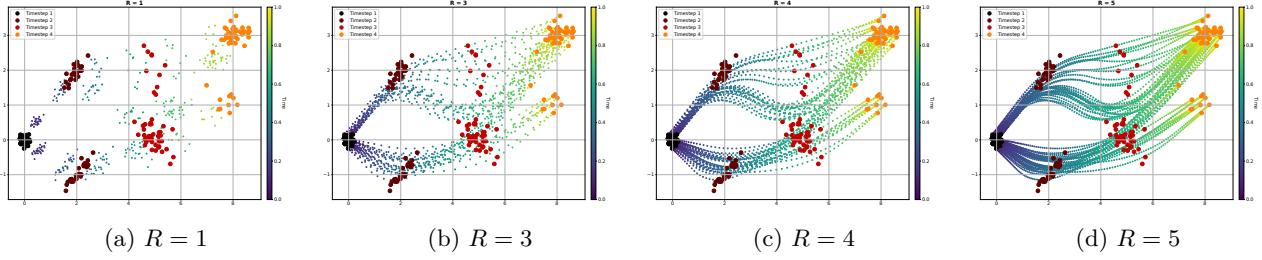


Figure 8: WLR on Diverging Gaussians at various values of  $R$ .

Figure 8 shows the application of WLR on the Diverging Gaussian data at various refinement levels, with lower refinement levels offering reduced computational time. This highlights the algorithms adaptability to user-defined refinement levels. In our implementation, the number of refined points returned by WLR also depends on  $R$  and  $M$  as  $|\nu^{(M)}| = 2^R(T + M - 1) + 2 - M$ , which follows from Algorithm 1.

## C PRELIMINARY RESULTS: WLR AS INITIALIZATION FOR DEEP LEARNING-BASED METHODS

While a comprehensive study of WLR initialization for deep learning methods remains a promising direction for future work, our preliminary experiments suggest some qualitative benefits. To demonstrate these advantages, we

compare two scenarios: training MIOFlow from scratch and training MIOFlow initialized with WLR trajectories, both with a fixed computational budget of 10 epochs. The qualitative results can be seen in Figure 9. MIOFlow alone struggles to capture the bifurcating behavior present in the diverging Gaussian dataset, producing simplified trajectories without branching as seen in (Figure 9a). In contrast, when initialized with WLR trajectories, MIOFlow successfully maintains the bifurcating structure while benefiting from the neural network’s capacity to refine the trajectories (Figure 9b). This result hints that WLR initialization could provide geometric guidance that helps in capturing complex trajectories.

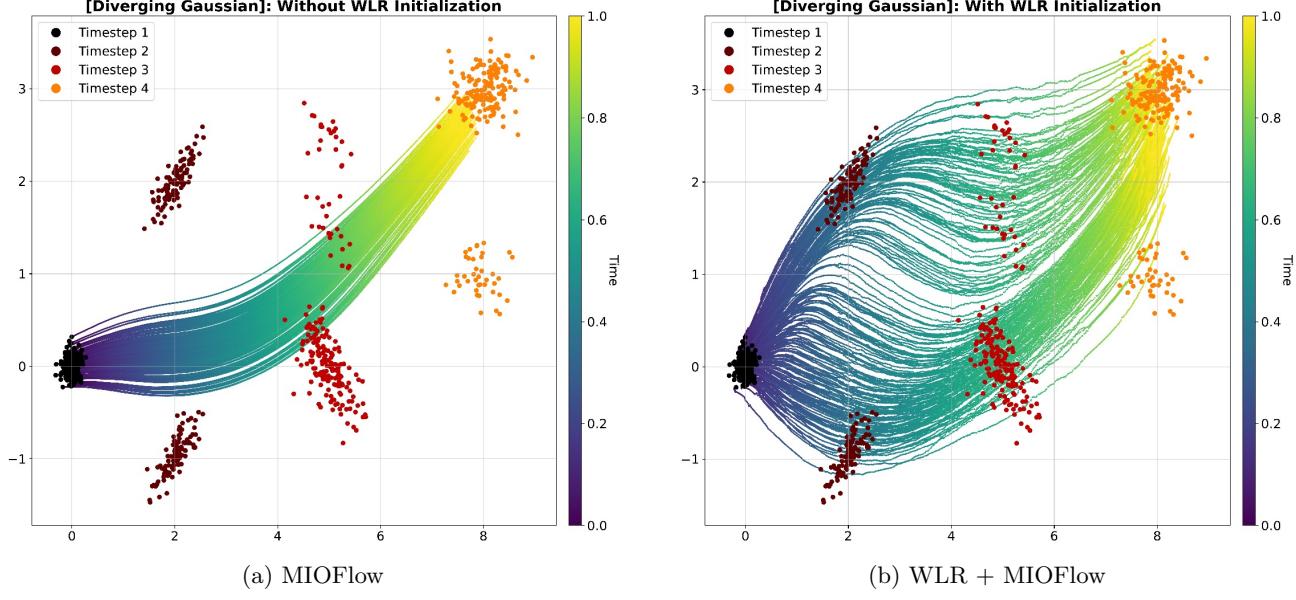


Figure 9: On the left, Figure 9a shows MIOFlow trajectories without any prior initialization. Figure 9b shows the trajectories after initialization with WLR on Diverging Gaussian Data. In both cases MIOFlow was run with a fixed budget of 10 epochs

## D EXPERIMENTAL DETAILS

We provide details on the experiments in Section 5.

### D.1 Datasets

We use a wide range of simulated and real datasets that capture the natural dynamics observed in cellular differentiation. This includes phenomena such as bifurcations and merges, as depicted in Figure 6.

**Diverging Gaussian** We designed this dataset (Figure 6a) to initiate from a singular point cluster and gradually disperse as time progresses. The points all have uniform mass.

**Petal** The Petal dataset (Figure 6b), as defined by Huguet et al. (2022), presents a significant challenge due to its inherent geometric structure. We resampled it to make  $n_t$  consistent. The points all have uniform mass.

**Dyngen Tree and Cycle** Utilizing Dyngen by Cannoodt et al. (2021), we simulated a single-cell RNA sequencing (scRNA-seq) dataset to model a dynamic cellular process. We generated both tree (Figure 6d) and cycle (Figure 6e) data, using 100 transcription factors for each. They were then embedded into a 10-dimensional space via PCA and PHATE (Moon et al. (2019)), respectively.

**CITE-seq** This dataset from the Multimodal Single-Cell Integration challenge at NeurIPS 2022 Stoeckius et al. (2017) consists of time-series measurements of CD34+ hematopoietic stem and progenitor cells (HSPCs)

collected at four distinct time points: days 2, 3, 4, and 7 post-isolation. To prepare the data for trajectory inference, we normalized each feature dimension to ensure comparable scales across features. We then computed the first 100 principal components of the data to capture the most salient features of cellular variation and to maintain a higher dimensionality compared to other datasets.

**Converging Gaussian** This dataset (Figure 6c) was crafted by Clancy and Suarez (2022) to simulate the intricate dynamics characteristic of cellular differentiation. The points are sampled from Gaussian distributions with intermediate stages having cellular division processes. Changing  $n_t$  across intermediate time steps makes this dataset challenging. F&S cannot handle this setup, so we skip it. On the other hand, we were not able to adapt code by Clancy and Suarez (2022) to run on any other datasets, so we report results from Wasserstein-Fisher-Rao only on Converging Gaussian. We also do not calculate MSE since the definition becomes ambiguous as the number of points between the predicted point clouds and the held out point clouds can differ.

**CITE-seq Supercells** We used Bilous et al. (2022)<sup>1</sup> on the full CITE-seq dataset to create supercell (metacells), which are aggregates of individual cells that share similar transcriptomic profiles. SuperCells package constructs a  $k$ -nearest neighbor ( $k$ -NN) graph based on the most variable genes, then partitions this graph to form metacells. Gene expression profiles for these metacells are computed by averaging the expression levels of their constituent cells. We used a graining level of  $\gamma = 20$  with  $k = 5$ .

## D.2 Parameters and Setup

All experiments including runtime calculations were carried out on x86\_64 machine with AMD EPYC 7713 64-Core Processor with NVIDIA A100-PCIE-40GB GPU. WLR, F&S and Wasserstein-Fisher-Rao did not require any GPU for training. Our paper used two software packages: Python OT<sup>2</sup> version 0.9.3 with MIT license from Flamary et al. (2021) and DynGen<sup>3</sup> version 1.0.5 with MIT license (CC-BY) from Cannoodt et al. (2021). Dataset details are in Table 3. For the experiments in Table 4, data subsampling was carried out using the following seeds: 1783345717, 3621289926, 3137468471, 1383545203, 3849164755.

Dataset	$d$	$T + 1$	$j$	Original # of points	$n_t$	Author
Diverging Gaussian	2	4	2	200	-	Us
Petal	2	5	2	135	-	Huguet et al. (2022)
Converging Gaussian	2	4	2	126, 378, 252, 126	-	Clancy and Suarez (2022)
Dygen Tree	10	7	3	319, 665, 992, 993, 743, 460, 328	319	Cannoodt et al. (2021)
Dygen Cycle	10	15	8	52	-	Cannoodt et al. (2021)
CITE-seq	100	4	2	6071, 7643, 8485, 7195	1000	Stoeckius et al. (2017)
CITE-seq Supercells	100	4	2	304, 382, 424, 360	-	Us

Table 3: Dimensions  $d$ , number of time steps  $T + 1$ , held out time step  $j$ , original number of points, and uniformly sampled (if applicable) number of points  $n_t$  of the datasets used in Section 5.

**Wasserstein Lane-Riesenfeld (WLR)** We used  $M = 10$  for Petal and  $M = 2$  for the rest.  $R = 7$  and  $\epsilon = 10^{-5}$  in all experiments in Section 5.2. For our experiments on supercells, we used  $R = 4, 5, 6$  and  $\epsilon = 10^{-8}$ .

**TrajectoryNet** For all experiments, we used TrajectoryNet<sup>4</sup> with 1000 iterations and `whiten=True` with all other default optimization parameters in Tong et al. (2020).

**MIOFlow** We observed the best trajectory inference results without training the geodesic autoencoder but utilizing the density loss for all of our experiments. We used the following hyperparameters for MIOFlow<sup>5</sup>. Despite our best efforts, we were unable to reproduce their published figure on Petal. After multiple attempts, we report their *best* outcome in Table 4 and Figure 10.

<sup>1</sup><https://github.com/GfellerLab/SuperCell>

<sup>2</sup><https://pythonot.github.io/>

<sup>3</sup><https://github.com/dynverse/dyngen>

<sup>4</sup><https://github.com/KrishnaswamyLab/TrajectoryNet>

<sup>5</sup><https://github.com/KrishnaswamyLab/MIOFlow>

- Diverging Gaussian: sample\_size=60, n\_local\_epochs = 50, n\_epochs = 0, n\_post\_local\_epochs = 0,  $\lambda_e = 0.001$ ,  $\lambda_d = 10$ ,  $\epsilon = 0.5$  n\_points = 200, n\_trajectories = 200, n\_bins = 500, n\_epochs\_emb = 1000
- Converging Gaussian: sample\_size=30, n\_local\_epochs = 50, n\_epochs = 0, n\_post\_local\_epochs = 0,  $\lambda_e = 0.001$ ,  $\lambda_d = 35$ ,  $\epsilon = 0.5$  n\_points = 64, n\_trajectories = 64, n\_bins = 500, n\_epochs\_emb = 1000
- Petal: n\_local\_epochs = 50, n\_points = 135, n\_trajectories = 135, n\_bins = 500. The rest are default hyperparameters mentioned in Huguet et al. (2022).
- Dyngen Tree: sample\_size=30, n\_local\_epochs = 5, n\_epochs = 0, n\_post\_local\_epochs = 0,  $\lambda_e = 0.001$ ,  $\lambda_d = 15$ ,  $\epsilon = 0.5$  n\_points = 100, n\_trajectories = 100, n\_bins = 500, n\_epochs\_emb = 1000
- Dyngen Cycle: sample\_size=30, n\_local\_epochs = 10, n\_epochs = 0, n\_post\_local\_epochs = 0,  $\lambda_e = 0.001$ ,  $\lambda_d = 15$ ,  $\epsilon = 0.5$  n\_points = 50, n\_trajectories = 50, n\_bins = 500, n\_epochs\_emb = 1000
- CITE-seq: sample\_size=100, n\_local\_epochs = 50, n\_epochs = 0, n\_post\_local\_epochs = 0,  $\lambda_e = 0.001$ ,  $\lambda_d = 15$ ,  $\epsilon = 0.5$  n\_points = 50, n\_trajectories = 1000, n\_bins = 500, n\_epochs\_emb = 1000, use\_density\_loss = True

**Fast and Smooth (F & S)** The original authors of Chewi et al. (2021) proposed a general algorithm for performing interpolation using cubic splines in  $\mathbb{R}^d$  after determining an ordering through successive optimal transport computations on the initially observed point clouds. To ensure a fair comparison, we wrote our own code to adapt their methodology to approximate B-splines in  $\mathbb{R}^d$  via the Lane-Riesenfeld algorithm and carried out comparisons with both the versions (B-splines and cubic splines) of this algorithm. We used the same  $M$  and  $R$  as WLR. As mentioned in Appendix D.1, F&S is not able to run on Converging Gaussian.

**Wasserstein-Fisher-Rao** The authors of Clancy and Suarez (2022) proposed a method for computing splines for measures of differing masses using the notion of unbalanced optimal transport. We used their method for computing splines on the converging Gaussian data using all default parameters with  $\eta_{OT} = 0.001$  for 200 iterations and 32 interpolating splines from the first to the last time step. As mentioned in Appendix D.1, we were able to run their code<sup>6</sup> only on this dataset.

**Conditional Flow Matching** We used batch\_size = 256,  $\sigma = 0.1$ ,  $w = 64$  and learning\_rate = 0.0001, epochs = 10000 and n\_traj\_bins = 500 for all our experiments. The code can be found here<sup>7</sup>.

- Model Architecture:
  - Network: Multi-Layer Perceptron (MLP) with default layers and size and activations
  - Input Dimension: Concatenation of position coordinates (data dimensions) and time
  - Time-Varying: Yes
  - Output Dimension: Matching data dimensionality
- Training Parameters:
  - Batch size: 256
  - Number of iterations: 10,000
  - Learning rate: 1e-4 (Adam optimizer)
  - Sigma ( $\sigma$ ): 0.1
  - Weight Decay: 0 (default)
- Loss Function:
  - Type: Mean Squared Error (MSE) between predicted and true conditional flows
  - Additional Terms: None
- Trajectory Generation:

<sup>6</sup>[https://github.com/felipesua/WFR\\_splines](https://github.com/felipesua/WFR_splines)

<sup>7</sup><https://github.com/atong01/conditional-flow-matching>

- Number of Trajectory Bins: 500
- Trajectory Sampling:
  - \* ODE Solver: Dormand-Prince (dopri5)
  - \* Sensitivity Method: Adjoint

**Wasserstein Lagrangian Flow (WLF)** We implemented the algorithm proposed by Neklyudov et al. (2023) and compared their method using two loss variants, namely unbalanced optimal transport with physical potential (UBOT+) and Schrödinger Bridge (SB). For all experiments, we used the following hyperparameters:

- Model Architecture:
  - Score Network (S): 4-layer MLP with 512 hidden units and Swish activation
  - Sampling Network (Q): 4-layer MLP with 128 hidden units and Swish activation
  - Both networks use skip connections and dropout rate of 0.2
- Training Parameters:
  - Batch size: 512
  - Number of iterations: 1000
  - Learning rate: 1e-3 (Adam optimizer)
  - Step size: 0.05
- Loss-Specific Parameters:
  - SB variant: noise level  $\sigma = 0.1$
  - UBOT+ variant: balance parameter  $\lambda = 1.0$
- Sampling:
  - Latent dimension for sampling network: 10
  - Number of simulation steps: 500

## E ADDITIONAL RESULTS

We present full visual comparison between WLR and other methods in Figure 10. Trajectory inference is done on all time steps with no hold-out point clouds. Table 4 contains experimental results on datasets having point clouds with uniform mass.

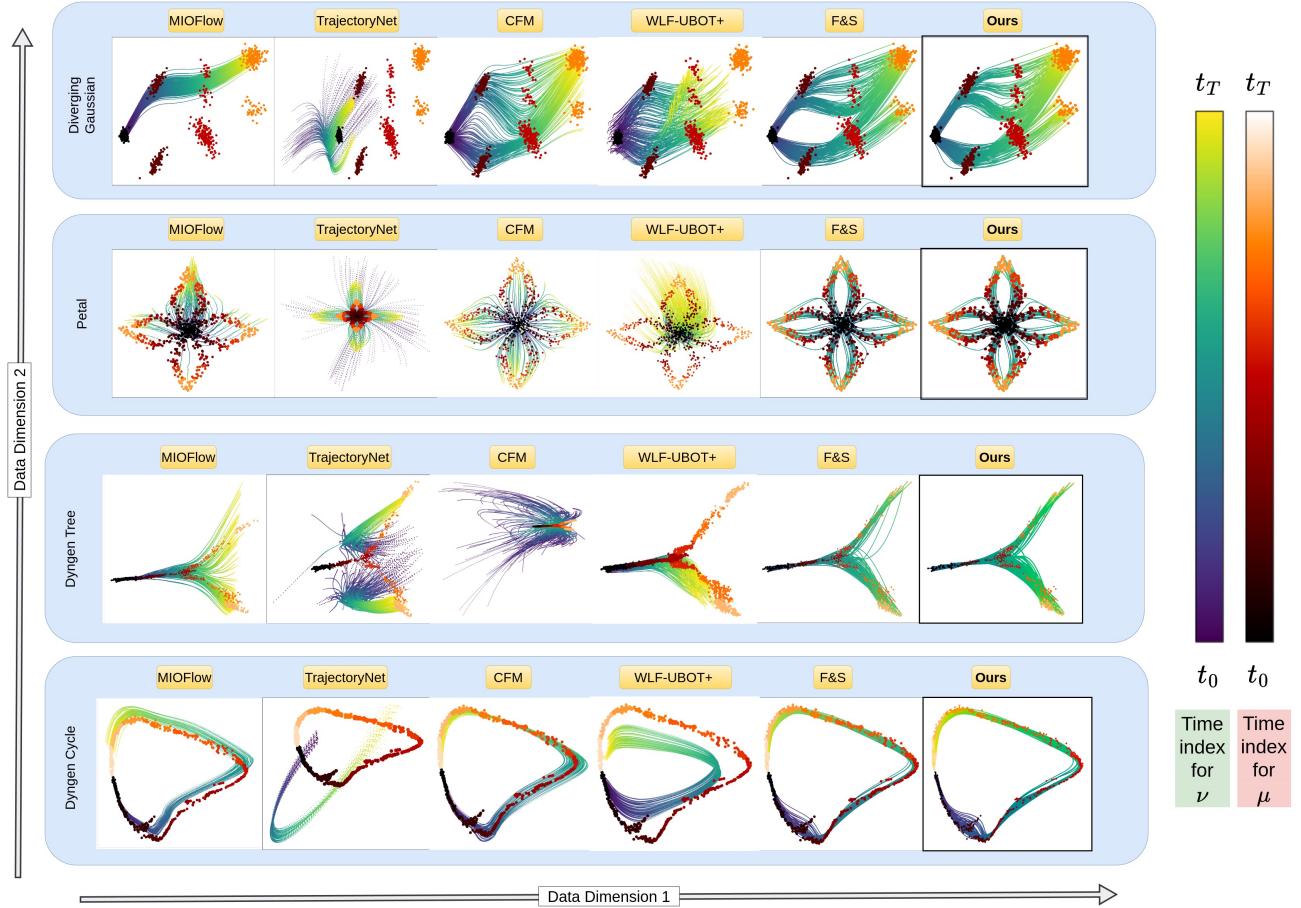


Figure 10: Full-scale visual comparison between WLR and other methods with point clouds having uniform mass and equal points per time step. Trajectory inference is done on all time steps with no hold-out point clouds. Note that we used F & S with B-splines to ensure a fair comparison with WLR, as both are approximation schemes.

Dataset	Method	Runtime (sec) ↓	Leave-one-out		Mean	
			MSE ↓	$W_1 \downarrow$	MSE ↓	$W_1 \downarrow$
Diverging Gaussian	WLR (m=2) (Ours)	$3.99 \pm 0.03$	<b><math>1.73 \pm 0.06</math></b>	<b><math>1.16 \pm 0.02</math></b>	<u><math>1.16 \pm 0.81</math></u>	<b><math>0.61 \pm 0.61</math></b>
	CFM	$182.07 \pm 0.67$	$1.92 \pm 0.06$	$1.45 \pm 0.02$	<b><math>1.12 \pm 0.81</math></b>	$0.66 \pm 0.57$
	MIOFlow	$16.85 \pm 0.16$	$2.18 \pm 1.69$	$1.58 \pm 0.76$	$1.61 \pm 1.15$	$1.19 \pm 0.76$
	TrajectoryNet	$1340.30 \pm 88.83$	$14.04 \pm 0.21$	$5.30 \pm 0.04$	$12.02 \pm 9.42$	$4.37 \pm 2.11$
	F & S (B Spline)	<b><math>1.14 \pm 0.01</math></b>	<b><math>1.73 \pm 0.06</math></b>	<b><math>1.16 \pm 0.02</math></b>	<u><math>1.16 \pm 0.81</math></u>	<b><math>0.61 \pm 0.61</math></b>
	WLF-SB	$103.08 \pm 10.52$	$2758.34 \pm 5987.36$	$41.16 \pm 68.71$	$2799.99 \pm 6063.34$	$41.98 \pm 68.79$
	WLF-UBOT+	$89.71 \pm 12.08$	$1.96 \pm 0.33$	$1.16 \pm 0.18$	$7.93 \pm 0.43$	$3.29 \pm 0.05$
Petal	WLR (m=2) (Ours)	<b><math>25.27 \pm 0.97</math></b>	$0.26 \pm 0.01$	<b><math>0.09 \pm 0.01</math></b>	<b><math>0.01 \pm 0.01</math></b>	<b><math>0.05 \pm 0.05</math></b>
	CFM	$233.26 \pm 1.10$	<b><math>0.23 \pm 0.01</math></b>	$0.18 \pm 0.02$	$0.30 \pm 0.27$	$0.14 \pm 0.08$
	MIOFlow	$45.15 \pm 0.19$	$0.26 \pm 0.02$	$0.26 \pm 0.09$	$0.32 \pm 0.27$	$0.24 \pm 0.16$
	TrajectoryNet	$1415.57 \pm 56.44$	$0.49 \pm 0.02$	$0.31 \pm 0.01$	$0.81 \pm 0.94$	$0.64 \pm 0.40$
	F & S (B Spline)	$36.59 \pm 0.42$	$0.27 \pm 0.01$	$0.11 \pm 0.01$	$0.36 \pm 0.33$	$0.11 \pm 0.10$
	WLF-SB	$53.44 \pm 6.01$	$1631.28 \pm 2818.99$	$38.75 \pm 46.76$	$1630.06 \pm 2816.59$	$38.74 \pm 46.74$
	WLF-UBOT+	$35.79 \pm 4.26$	$0.25 \pm 0.15$	$0.42 \pm 0.26$	$0.28 \pm 0.15$	$0.47 \pm 0.24$
Dyngen Tree	WLR (m=2) (Ours)	$13.18 \pm 0.09$	<b><math>0.82 \pm 0.02</math></b>	<b><math>1.76 \pm 0.01</math></b>	<b><math>1.55 \pm 1.13</math></b>	<b><math>1.32 \pm 0.94</math></b>
	CFM	$423.42 \pm 0.38$	$2.02 \pm 0.40$	$3.31 \pm 0.39$	$17.68 \pm 28.68$	$6.69 \pm 7.26$
	MIOFlow	$6.98 \pm 0.12$	$2.17 \pm 0.90$	$3.75 \pm 0.91$	$16.74 \pm 26.93$	$6.34 \pm 6.16$
	TrajectoryNet	$2692.71 \pm 139.31$	$1.38 \pm 0.20$	$2.97 \pm 0.32$	$1.91 \pm 1.33$	$3.14 \pm 1.09$
	F & S (B Spline)	<b><math>3.65 \pm 0.02</math></b>	<b><math>0.82 \pm 0.02</math></b>	<b><math>1.76 \pm 0.01</math></b>	<b><math>1.55 \pm 1.13</math></b>	<b><math>1.32 \pm 0.94</math></b>
	WLF-SB	$52.68 \pm 7.63$	$10.27 \pm 5.10$	$9.59 \pm 2.41$	$9.99 \pm 4.98$	$9.26 \pm 2.26$
	WLF-UBOT+	$39.32 \pm 7.79$	$2.53 \pm 0.37$	$4.34 \pm 0.44$	$3.13 \pm 0.44$	$4.52 \pm 0.49$
Dyngen Cycle	WLR (m=2) (Ours)	<u><math>1.62 \pm 0.01</math></u>	<b><math>0.06 \pm 0.01</math></b>	<b><math>0.58 \pm 0.01</math></b>	<b><math>0.09 \pm 0.04</math></b>	<b><math>0.73 \pm 0.32</math></b>
	CFM	$1055.50 \pm 2.92$	$0.11 \pm 0.05$	$0.77 \pm 0.21$	$0.76 \pm 1.29$	$0.85 \pm 0.54$
	MIOFlow	$26.86 \pm 0.12$	<b><math>0.06 \pm 0.02</math></b>	$0.65 \pm 0.12$	$0.11 \pm 0.07$	$0.75 \pm 0.37$
	TrajectoryNet	$5758.90 \pm 261.24$	$14.50 \pm 2.55$	$11.99 \pm 1.04$	$7.08 \pm 5.36$	$7.58 \pm 3.51$
	F & S (B Spline)	<b><math>1.36 \pm 0.01</math></b>	<b><math>0.06 \pm 0.01</math></b>	<b><math>0.58 \pm 0.01</math></b>	$0.10 \pm 0.04$	<b><math>0.73 \pm 0.32</math></b>
	WLF-SB	$93.09 \pm 10.48$	$222.18 \pm 306.67$	$37.46 \pm 31.30$	$218.33 \pm 304.44$	$36.60 \pm 31.75$
	WLF-UBOT+	$79.23 \pm 9.78$	$2.46 \pm 0.38$	$4.92 \pm 0.36$	$1.51 \pm 0.32$	$3.60 \pm 0.37$
CITE-seq	WLR (m=2) (Ours)	$120.91 \pm 0.43$	$35.54 \pm 0.48$	<b><math>31.72 \pm 0.15</math></b>	$31.70 \pm 20.70$	<b><math>12.14 \pm 13.23</math></b>
	CFM	$186.30 \pm 2.01$	$37.36 \pm 0.70$	$34.32 \pm 0.33$	$34.27 \pm 23.25$	$24.92 \pm 14.84$
	MIOFlow	$90.54 \pm 1.21$	$31.27 \pm 0.84$	$32.43 \pm 0.40$	$34.50 \pm 8.79$	$27.08 \pm 10.87$
	TrajectoryNet	$10349.98 \pm 192.70$	<b><math>11.21 \pm 0.14</math></b>	$31.75 \pm 0.23$	<b><math>20.06 \pm 7.96</math></b>	$39.84 \pm 7.03$
	F & S (B Spline)	<b><math>6.41 \pm 0.03</math></b>	$35.54 \pm 0.48$	<b><math>31.72 \pm 0.15</math></b>	$31.70 \pm 20.70$	<b><math>12.14 \pm 13.23</math></b>
	WLF-SB	$98.58 \pm 5.63$	$596.42 \pm 489.34$	$220.59 \pm 93.82$	$602.36 \pm 488.36$	$220.97 \pm 93.50$
	WLF-UBOT+	$105.10 \pm 8.29$	$30.15 \pm 0.28$	$32.58 \pm 0.20$	$34.13 \pm 0.26$	$29.65 \pm 0.12$

Table 4: Quantitative comparison of approximation methods (non-interpolatory) for trajectory inference. Best scores are in **bold** and second best are underlined. F & S (B Spline) is our modified version of F & S proposed by Chewi et al. (2021). The methods in this table include B-spline based approximation methods and neural network approaches, both of which provide smooth approximations rather than exact interpolations of the input point clouds. While our approach and F&S often appear numerically similar, this is expected: the original F&S method in Chewi et al. (2021) was limited to cubic spline interpolation in  $\mathbb{R}^d$ , so we implemented a B-spline approximation version of F&S (via LaneRiesenfeld), matching our own B-spline setting and thereby yielding closely aligned scores. Importantly, our method can naturally handle mass splitting and varying numbers of points across time, which F&S and many other interpolation methods cannot address. Hence, even though our numerical metrics coincide with F&S under comparable spline degrees, our approach offers greater flexibility in modeling realistic biological processes, such as cell differentiation with bifurcations or supercell formation, where the total mass and number of points vary over time.

Dataset	Method	Runtime (sec) ↓	Leave-one-out		Mean	
			MSE ↓	$W_1 \downarrow$	MSE ↓	$W_1 \downarrow$
Dyngen Cycle	W-4-point (Ours)	$2.99 \pm 0.14$	<b><math>0.04 \pm 0.01</math></b>	$0.40 \pm 0.01$	<b><math>0.07 \pm 0.04</math></b>	<b><math>0.42 \pm 0.31</math></b>
	F & S (cubic)	<u><math>0.03 \pm 0.01</math></u>	$0.04 \pm 0.01$	<b><math>0.39 \pm 0.01</math></b>	$0.08 \pm 0.04$	<b><math>0.42 \pm 0.31</math></b>
Uniform Gaussian Cycle	W-4-point (Ours)	$8.35 \pm 0.75$	<b><math>3.56 \pm 0.06</math></b>	<b><math>2.28 \pm 0.02</math></b>	<b><math>3.19 \pm 2.13</math></b>	<b><math>1.89 \pm 1.06</math></b>
	F & S (cubic)	<b><math>0.07 \pm 0.01</math></b>	$4.14 \pm 0.07$	$2.52 \pm 0.02$	$3.56 \pm 2.52$	$2.00 \pm 1.19$

Table 5: Quantitative comparison of interpolatory methods for trajectory inference. Best scores are in **bold**. We compare two interpolatory schemes: F & S (cubic) splines and our Wasserstein four-point scheme. Both ensure exact interpolation through the observed point clouds.