Wasserstein-Fisher-Rao Splines

Julien Clancy  
clancy.julien@gmail.com

Felipe Suárez  
felipesc@mit.edu

Massachusetts Institute of Technology*

April 6, 2022

Abstract

We study interpolating splines on the Wasserstein-Fisher-Rao (WFR) space of measures with differing total masses. To achieve this, we derive the covariant derivative and the curvature of an absolutely continuous curve in the WFR space. We prove that this geometric notion of curvature is equivalent to a Lagrangian notion of curvature in terms of particles on the cone. Finally, we propose a practical algorithm for computing splines extending the work of Chewi et al. (2020a).

1 Introduction

Let $\mu_1, \mu_2, \ldots, \mu_n$ be $n$ positive measures of differing total masses. How to interpolate them? This question is motivated by cellular trajectory reconstruction where $\mu_i$ is a population of cells at time $i$ (Schiebinger et al., 2019). Cells move in gene space as they evolve, but also divide and proliferate. While ad-hoc fixes for this issue have been proposed, e.g. via renormalization and using optimal transport (OT) (Chewi et al., 2020a), the conservation of mass property inherent to OT makes it a less suitable tool for this task.

Curve evolution in Wasserstein space is governed by the continuity equation

$$\partial_t \mu_t + \text{div}(v_t \mu_t) = 0,$$

which can be viewed as a simple restatement of conservation of mass, following the divergence theorem. This equation is underdetermined, and the geometry of $W^2$ is induced by selecting for each time $t$ the field $v_t$ with minimal kinetic energy:

$$v_t = \arg \min_{u_t} \int \|u_t\|^2 \, d\mu_t = \arg \min_{u_t} \|u_t\|_{\mu_t}^2.$$

It can be seen (Gigli, 2012) that the optimal $v_t$ lies in the closure of $\{\nabla \varphi \mid \varphi \in C_c^\infty\}$. The celebrated Benamou-Brenier theorem states that the $W^2$ distance, defined by optimal transport, is equal to the least total kinetic energy among all possible paths.

*J.C. was supported by NSF GRFP, and F.S. was supported by the MathWorks Fellowship and the NSF grant IIS-1838071.

1Formally, this equation is to be interpreted weakly in duality with functions $\phi \in C_c^\infty(\mathbb{R}^d \times \mathbb{R})$ via the divergence theorem i.e. $\int f \, d\mu_t = \int (v_t, \nabla \varphi) \, d\mu_t$.
Theorem 1. (Benamou and Brenier, 2000) Let $\mu_0, \mu_1$ be probability measures. Then

$$W_2^2(\mu_0, \mu_1) = \inf_{(\mu_t, v_t)} \int_0^1 \|v_t\|_{\mu_t}^2 \, dt,$$

where the infimum is taken over solutions of the continuity equations with prescribed boundary data $\mu_0$ and $\mu_1$.

Thus, in order to define a new metric between measures of arbitrary mass we have to alter the continuity equation, running the procedure in reverse. Turning back to intuition, where the term $\text{div}(v_t \mu_t)$ represents mass translation, we add another term representing growth or decay:

$$\partial_t \mu_t + \text{div}(v_t \mu_t) = 4\alpha_t \mu_t. \quad (2)$$

We call this the nonconservative continuity equation.\(^2\) The term in the right-hand side allows for a relative growth or decay in mass. While measures $(\mu_t)$ evolving according to (2) may have varying mass, they are granted to stay positive.

From here, we measure the magnitude of a pair $(v_t, \alpha_t)$ by

$$\| (v_t, \alpha_t) \|_{\mu_t}^2 = \int (|v_t|^2 + 4\alpha_t^2) \, d\mu_t,$$

where $\| \cdot \|$ stands for norm in WFR space and $| \cdot |$ for norm of vectors in Euclidean space. The Wasserstein-Fisher-Rao distance is then defined by

$$\text{WFR}(\mu_0, \mu_1)^2 = \inf_{(\mu_t, v_t, \alpha_t)} \int_0^1 \| (v_t, \alpha_t) \|_{\mu_t}^2 \, dt \quad (3)$$

where, again, the minimization is over solutions to (2) with $\mu_0$ and $\mu_1$ prescribed. For more detailed discussions of the genesis of the equation (2) and the development of the connection with optimal transport see Kondratyev et al., 2016; Liero et al., 2018; Chizat, 2017.

It was shown simultaneously in Liero et al. (2018); Chizat (2017) that (3) defines a metric on $\mathcal{M}_+(\mathbb{R}^d)$, the space of non-negative measures, and that it turns this into a geodesic space. Furthermore, Kondratyev et al. (2016) shows it has the structure of a pseudo-Riemannian manifold analogous to the Riemannian structure on $W_2$ (Otto, 2001), with inner product given by

$$\langle (v, \alpha), (w, \beta) \rangle_\mu = \int \langle v, w \rangle + 4\alpha\beta \rangle \, d\mu,$$

where the tangent space is

$$T_\mu(\mathcal{M}_+) = \text{clos}_{L^2(\mu)} \left\{ (\nabla \alpha, \alpha) \, | \, \alpha \in C_0^\infty(\mathbb{R}^d) \right\}.$$

Remark. Notice that this is the same tangent space as for $W_2$, with the difference being that the WFR Riemannian metric is the full $H^1$ Sobolev norm, while the $W_2$ Riemannian metric includes only the gradient.

There has been recent interest in understanding the curvature of these spaces (Chewi et al., 2020b), and specifically in understanding curvature-minimizing interpolating curves that generalize Euclidean splines (Benamou et al., 2019; Chen et al., 2018; Chewi et al., 2020a). We aim to replicate some of these advances in

\(^2\)The factor of 4 is for notational convenience in accordance with the literature.
the Wasserstein-Fisher-Rao case, which, as mentioned, plays an important role in applications. Specifically, we will 1) characterize the covariant derivative in WFR space and use this to define a notion of intrinsic splines; 2) define splines from a pseudo-Eulerian perspective, as measures on paths, and establish a relationship to intrinsic splines; and 3) define a more practically workable notion of WFR splines, analogous to the transport splines of Chewi et al. (2020a), and examine its relationship to intrinsic splines.

2 The Covariant Derivative

In this section, we derive an expression for the covariant derivative. We first present a new derivation of the covariant derivative in $W_2$ that is simpler than the definition in Gigli (2012). In turn, this new derivation is extended to WFR space.

The Riemannian metric on $W_2$ is given by $\langle v, w \rangle \mu = \int \langle v, w \rangle d\mu$, and the covariant derivative $\frac{D}{dt}$ and its associated Levi-Civita connection $\nabla$ must satisfy two properties:

1. **Leibniz rule.** If $\mu_t$ is a curve and $(v^1_t)$ and $(v^2_t)$ are two (tangent) vector fields along it, then
   \[ \frac{d}{dt} \langle v^1_t, v^2_t \rangle_{\mu_t} = \langle v^1_t, \frac{D}{dt} v^2_t \rangle_{\mu_t} + \langle \frac{D}{dt} v^1_t, v^2_t \rangle_{\mu_t}. \]  
   \hspace{1cm} (P1)

2. **Torsion-freeness.** If $X$ and $Y$ are vector fields, then
   \[ \nabla_X Y - \nabla_Y X = [X, Y]. \]  
   \hspace{1cm} (P2)

Let $(v_t)$ be the tangent field of the curve $(\mu_t)$ (as noted above, the minimal field is unique and is a gradient\textsuperscript{3}). The product rule yields

\[ \frac{d}{dt} \langle v^1_t, v^2_t \rangle_{\mu_t} = \frac{d}{dt} \int \langle v^2_t, v^2_t \rangle_{\mu_t} d\mu_t 
= \int \left( \langle \partial_t v^1_t, v^2_t \rangle + \langle \partial_t v^2_t, v^1_t \rangle \right) d\mu_t + \int \langle v^1_t, \partial_t v^2_t \rangle_{\mu_t} d(\partial_t \mu_t). \]

Because $(\mu_t, v_t)$ solves the continuity equation, the dual definition of $\text{div}(v_t \mu_t)$ and the divergence theorem give

\[ \int \langle v^1_t, v^2_t \rangle_{\mu_t} d(\partial_t \mu_t) = \int \langle \nabla v^1_t \cdot v_t, v^2_t \rangle + \langle \nabla v^2_t \cdot v_t, v^1_t \rangle_{\mu_t}. \]

Together with (P1), it yields

\[ \langle \frac{D}{dt} v^1_t, v^2_t \rangle_{\mu_t} + \langle v^1_t, \frac{D}{dt} v^2_t \rangle_{\mu_t} = \langle \partial_t v^1_t + \nabla v^1_t \cdot v_t, v^2_t \rangle_{\mu_t} + \langle v^1_t, \partial_t v^2_t + \nabla v^2_t \cdot v_t \rangle_{\mu_t}. \]

From here, it natural to postulate that

\[ \frac{D}{dt} v^1_t = P_{\mu_t} \left( \frac{D}{dt} v^1_t \right) = P_{\mu_t} \left( \partial_t v^1_t + \nabla v^1_t \cdot v_t \right), \]

(4)

where $P_{\mu_t}$ is the orthogonal projection onto $T_{\mu_t}(\mathcal{P}_2)$ in $L^2(\mu_t)$. We call $\frac{D}{dt}$ the total derivative. Note that if $v^1_t = v = \nabla \varphi_t$ then $\partial_t v_t + \nabla v_t \cdot v_t = \nabla \left( \partial_t \varphi_t + \frac{1}{2} |\nabla \varphi_t|^2 \right) \in L^2(\mu_t)$.

\textsuperscript{3}In general it is merely in the closure of the set of gradients, but if all measures involved are absolutely continuous then it is truly a gradient.
Therefore, we may define two new tangent fields along these curves by
\[ u_1^1 = v_0^2, \]
\[ u_1^2 = v_0^1. \]

With this definition, it is reasonable to interpret
\[ \nabla u_0^1 u_2^2 \bigg|_{t=0} = \frac{D}{dt} u_2^1 \bigg|_{t=0}, \]
with the derivative being taken along \( \mu_1^2 \), and similarly for \( \nabla u_0^1 u_1^1 \). Now, fix \( \varphi \) and consider the functional \( F : \mu \mapsto \int \varphi \, d\mu \). By the continuity equation, the derivative of \( F \) along \( u_0^2 \) at \( \mu \) is
\[ \frac{d}{dt} F[\mu_t] \bigg|_{t=0} = \int \varphi \, d \left( \partial_t \mu_t^1 \right) \bigg|_{t=0} = \int (\nabla \varphi, u_0^2) \, d\mu. \]
Then since the covariant derivative above respects the metric,
\[ u_0^1(u_0^2(F)) \big|_{\mu_t} = \frac{d}{dt} (\nabla \varphi, u_0^2) \bigg|_{t=0} = \left\langle \frac{D}{dt} \nabla \varphi, u_0^2 \right\rangle + \left\langle \nabla \varphi, \frac{D}{dt} u_0^2 \right\rangle \bigg|_{t=0}
= \left\langle \nabla^2 \varphi \cdot u_0^1, u_0^2 \right\rangle_{\mu} + \left\langle \nabla \varphi, \nabla u_0^2 u_1^1 \right\rangle_{\mu}, \]
where we have used the definition of \( \frac{D}{dt} \) to calculate \( \frac{D}{dt} \nabla \varphi \) on the third line. Performing the same calculation for \( u_0^2(u_0^1(F)) \big|_{\mu} \) and subtracting, since \( \nabla^2 \varphi \) is symmetric the first terms cancel and we get
\[ u_0^1(u_2^2(F)) \big|_{\mu} - u_0^2(u_1^1(F)) \big|_{\mu} = \left\langle \nabla \varphi, \nabla u_0^1 u_1^1 - \nabla u_0^2 u_1^1 \right\rangle_{\mu}. \]
Since gradients \( \nabla \varphi \) are dense in \( T_\mu(\mathcal{P}_2) \) this means that \( \frac{D}{dt} \) is indeed torsion-free.

Now we repeat the argument in WFR space.

**Theorem 2.** Let \( \mu_t \) be an absolutely continuous curve in Wasserstein-Fisher-Rao space satisfying the continuity equation with tangent fields \( (v_t, \alpha_t) \). The covariant derivative is given by
\[ \frac{D^2}{dt^2} \mu_t = \left( \partial_t v_t + \nabla v_t \cdot v_t + 4\alpha_t v_t \right) \bigg|_{\partial_t \alpha_t + \frac{1}{2} |\nabla \alpha_t|^2 + 2\alpha_t^2} \bigg|_{\partial_t^2 \mu_t}. \]

**Proof.** See appendix A. \( \square \)
Theorem 2 gives a geometrical proof of the dynamical characterization of geodesics in duality with (3) (Chizat, 2017, Theorem 1.1.15; Liero et al., 2018, Theorem 8.12). Indeed, an absolutely continuous geodesic \((\mu_t)\) in WFR space with tangent \(\dot{\mu}_t = (\nabla \alpha, \alpha)\) satisfies the Hamilton-Jacobi equation almost surely

\[
\partial_t \alpha_t + \frac{1}{2} |\nabla \alpha_t|^2 + 2 \alpha_t^2 = 0,
\]
or, equivalently, the curve \((\mu_t)\) is autoparallel:

\[
\nabla \dot{\mu}_t \dot{\mu}_t = \frac{D^2}{dt^2} \mu_t = 0.
\] (7)

3 E-Splines and P-splines

In this section, we define notions of \(E\)- and \(P\)-splines. As in the \(W^2\) case, we can define an intrinsic notion of curvature-minimizing interpolators, which we term \(E\)-splines, by

\[
\inf_{(\mu_t, v_t)} \int_0^1 \left\| \frac{D}{dt} v_t \right\|^2_{\mu_t} dt \text{ s.t. } \mu_{t_i} = \mu_i
\] (8)

Though the characterization (6) yields an explicit objective function, there is no practical way to optimize (8). Thus, we first define an analogy to the path splines (\(P\)-splines) introduced in Chen et al. (2018); Benamou et al. (2019). We give a brief introduction to them here (see Chewi et al., 2020a for more details).

3.1 Geodesics and \(P\)-splines in \(W^2\)

In the Wasserstein space \(W^2\) over a metric space \((X, d)\) it is known that geodesics can be represented as measures over paths (see Lisini, 2007). Specifically, let \(\Omega\) be the set all absolutely continuous paths in \(X\), let \(l\) be the length functional on \(\Omega\) i.e. for \(\omega \in \Omega\), \(l(\omega) := \int_0^1 |\dot{\omega}_t(t)| dt\), and let \(\epsilon_t\) be the time evaluation functional at time \(t\). If \(P^*\) is a measure over paths solution to

\[
\inf_{P \in P(\Omega)} \int l dP \text{ s.t. } (\epsilon_t)_# P = \mu_t \text{ for } t \in \{0, 1\},
\]

then the \(W^2\) geodesic between \(\mu_0\) and \(\mu_1\) is given by \(\mu_t = (\epsilon_t)_# P^*\). From this starting point and in the case of \(X = \mathbb{R}^d\), Chen et al. (2018) define \(P\)-splines as the minimizers of

\[
\inf_{P \in P(\Omega)} \int c dP \text{ s.t. } (\epsilon_t)_# P = \mu_t,
\] (9)

where this time \(\Omega\) is the set of paths with absolutely continuous derivative and \(c\) is the curvature cost \(c(\omega) = \frac{1}{2} \int_0^1 |\ddot{\omega}_t|^2 dt\). Letting \(W^2\) \(E\)-splines be defined by

\[
\inf_{(\mu_t, v_t)} \int_0^1 \left\| \frac{D}{dt} v_t \right\|^2_{\mu_t} dt \text{ s.t. } \mu_{t_i} = \mu_i
\] (10)

Chen et al. (2018) relates them:

**Theorem 3.** \((\text{Chen et al., 2018, Section 5.3})\) Let \((\mu_t)\) be some curve in \(W^2\) with derivative \((v_t)\). Then there is a measure \(P \in P(\Omega)\) such that \((\epsilon_t)_# P = \mu_t\), and the objective in (9) is equal to that of (10). Furthermore, \(P\) can be defined from the flow maps from the continuity equation.
This shows that problem (9) is a relaxation of (10). In Chewi et al. (2020a, Proposition 2), it is shown that this is not tight in the sense that there exists (Gaussian) measures \( \mu_1, \ldots, \mu_n \) that are a solution to problem (10), yet are suboptimal for (9).

This formulation cannot be directly extended to WFR since, in (9), the marginals of \( P \) must have the same mass at all times. The solution is to consider curves in a different base space, introduced in Liero et al. (2016); Chizat (2017), which we briefly describe.

### 3.2 The Cone Space

The cone space \( \mathcal{C} \) is the manifold \( \mathbb{R}^d \times \mathbb{R}_+ \), with \( \mathbb{R}^d \times \{0\} \) identified as a single point.\(^4\) Points \((x, r) \in \mathcal{C}\) are thought of as tuples of position and mass. The distance is given by

\[
d_{\mathcal{C}} ((x_0, r_0), (x_1, r_1))^2 = r_0^2 + r_1^2 - 2r_0r_1 \cos(|x_0 - x_1| \wedge \pi).
\]

Call the vertex of the cone (the point \( \mathbb{R}^d \times \{0\} \)) as \( \mathbf{0} \). Notice that if \(|x_0 - x_1| \geq \pi\) then \( d_{\mathcal{C}}((x_0, r_0), (x_1, r_1)) = r_0 + r_1 \). Since for any \((x, r)\) we have \( d_{\mathcal{C}}((x, r), \mathbf{0}) = r \), this reflects that fact that if \( x_0 \) and \( x_1 \) are far away then the shortest path is to the vertex and back out.

Writing \((v, p) = \frac{d}{dt}(x, r)\), the Riemannian metric on the cone space is given by

\[
\langle (v_1, p_1), (v_2, p_2) \rangle_{(x, r)} = \langle v_1, v_2 \rangle r^2 + p_1p_2. \tag{11}
\]

Given a measure \( \lambda \in \mathcal{M}(\mathcal{C}) \), we can project it to a measure \( \mathcal{P}\lambda \in \mathcal{M}(\mathbb{R}^d) \) via

\[
\int f(x) \, d\mathcal{P}\lambda(x) = \int r^2 f(x) \, d\lambda(x, r).
\]

The measure \( \lambda \) is then called a lift of \( \mathcal{P}\lambda \). The presence of the term \( r^2 \), as opposed to \( r \), is to simplify the parametrization of \( \mathcal{C} \). Observe that there are many possible liftings of a measure \( \mu \in \mathcal{M}(\mathbb{R}^d) \) to \( \lambda \in \mathcal{M}(\mathcal{C}) \), the most obvious of which is \( d\lambda(x, r) = \delta_1(r) \cdot d\mu(x) \), where \( \delta_p(\cdot) \) stands for the Dirac point measure at \( p \). As \( \mathcal{C} \) is (save for the point \( \mathbf{0} \)) a Riemannian manifold we can define its \( W_2 \) metric as usual

\[
W_{\mathcal{C}, 2}^2(\lambda, \eta) = \inf_{\gamma \in \Pi(\lambda, \eta)} \int d^2_{\mathcal{C}} \, d\gamma,
\]

where \( \Pi(\lambda, \eta) \) is the set of all couplings of measures \( \lambda \) and \( \eta \) on the cone. The following useful characterization is proved in Liero et al. (2016).

**Theorem 4.** (Liero et al., 2016, Theorem 7) For any measures \( \mu_0, \mu_1 \in \mathcal{M}_+(\mathbb{R}^d) \), we have

\[
\text{WFR}(\mu_0, \mu_1) = \inf_{\lambda_0, \lambda_1} W_{\mathcal{C}, 2}(\lambda_0, \lambda_1),
\]

where the \( \lambda_i \in \mathcal{P}(\mathcal{C}) \) project to \( \mu_i \), \( \mathcal{P}\lambda_i = \mu_i \) for \( i = 0, 1 \). Furthermore, there are optimal lifts \( \lambda_0, \lambda_1 \), and an optimal coupling \( \gamma \) between them.

This allows to characterize WFR geodesics in the same way as in Euclidean space.

\(^4\)Fraktur font characters will be reserved for objects relating to the cone space, consistent with Liero et al. (2018).
Proposition 1. (Liero et al., 2016, Section 3.4) Let $\mu_0, \mu_1 \in \mathcal{M}_+(\mathbb{R}^d)$, and let $\gamma \in \mathcal{M}(\mathcal{C}^2)$ be the optimal coupling of the optimal lifts in Theorem 4. For each pair of points $z_0, z_1 \in \mathcal{C}$, let $g_t^{z_0, z_1}$ be the geodesic between them. Then the curve of measures $(\mu_t)_t$ defined by

$$
\mu_t = \Psi[(g_t^{z_1, z_2})#\gamma]
$$

is a geodesic in WFR between $\mu_0$ and $\mu_1$.

We may view the above theorem as saying that WFR geodesics are identified with certain measures supported on geodesics in $\mathcal{C}$. In fact, Liero et al. (2016) prove that all curves have such a representation, and it is faithful to first order.

Theorem 5. (Liero et al., 2016, Theorem 15) Let $(\mu_t)$ be an absolutely continuous curve in WFR. Then there is a measure $P \in \mathcal{P}(\Omega_\mathcal{C})$ such that $\mu_t = \Psi[(e_t)_#P]$ and for a.e. $t \in [0, 1]$

$$
|\dot{\mu}_t|^2 = \int |\dot{z}_t|^2 dP(z)
$$

where $|\dot{\mu}_t|$ is the metric derivative, which is equal to the intrinsic Riemannian quantity $\|\dot{\mu}_t\|_{\mu_t}$.

As in the $W_2$ case, this analogy can be extended profitably to second order—that is, to $|\ddot{\mu}_t|$ in the WFR sense.

3.3 $P$-Splines on WFR

In light of the dual characterization of curvature in Wasserstein space of Theorem 3 and the careful reformulation of WFR as $W_2$ on the cone of Proposition 1 and Theorem 5, we define $P$-splines in WFR by

$$
\inf_{P \in \mathcal{P}(\Omega_\mathcal{C})} \int_0^1 |\dot{z}_t|^2 dt dP(z) \text{ s.t. } \Psi[(e_t)_#P] = \mu_t.
$$

We prove that this is indeed a relaxation of the $E$-spline problem (8).

Theorem 6. Let $\mu_t$ be a sufficiently smooth curve in WFR. Then there is a measure $P \in \mathcal{P}(\Omega_\mathcal{C})$ such that $\Psi[(e_t)_#P] = \mu_t$ for all $t$, and the $E$-cost of $\mu$ is equal to the $P$-cost of $P$. The measure $P$ is induced by the flow maps associated to the curve $\mu_t$.

Proof. See appendix B.

To understand the proof (and complete the description of the proposition) we must define the flow maps. In $W_2$, curves of measures are interpretable as particle flows via the flow maps defined by

$$
\dot{X}_t = v_t(X_t), \ X_0 = \text{Id}.
$$

It then holds that

$$
\mu_t = (X_t)_#\mu_0.
$$

There is a similar characterization for sufficiently smooth paths in WFR due to Maniglia (2007).
Proposition 2. (Maniglia, 2007, Proposition 3.6) Let $v \in L^1(W^{1,\infty}([R^d, R^d], [0, 1]))$ be a vector field and $\alpha \in C([R^d \times [0, 1])$ a bounded locally Lipschitz scalar function. For $\mu_0 \in \mathcal{M}_1([R^d])$, there is a unique weak solution to the nonconservative continuity equation (2) with initial measure $\mu_0$. Furthermore, this satisfies

$$\mu_t = (X_t) \# (R_t^2 \cdot \mu_0),$$

for the flow map $(X_t)$ and scalar field $(R_t)$ which solve the ODE system

$$\begin{cases} \dot{X}_t = v_t(X_t), & X_0 = \text{Id} \\ \dot{R}_t = 2\alpha_t(X_t) R_t, & R_0 = 1. \end{cases}$$

Analogous to what happens in the Wasserstein case, theorem 6 shows that the spline problem in Lagrangian terms can be relaxed in geometric terms via the flow map characterization of proposition 2. Thus, under absolute continuity of the measures and the curve, the Lagrangian and geometric formulations agree up to second order. The first order has three equivalent expressions, namely the metric derivative $|\mu_t|_2^2$, the WFR derivative $|\mu_t|_{2, \mu}^2$, and the Lagrangian form $\int |\dot{x}_t|^2 d\mu_t(z)$. A similar equivalence holds for the second order derivatives, i.e. the WFR covariant derivative $|\nabla_\mu \mu_t|_{2, \mu}^2$ and the Lagrangian form $\int |\dot{z}_t|^2 d\mu_t(z)$. We conjecture that this equivalence holds for higher order derivatives.

4 Transport Splines

In this section, we define a tractable and smooth interpolant of measures. Piecewise linear interpolation in $W_2$ has the virtue of being simple to construct, which allows for additional regularization, as it is often found in applications (Schiebinger et al., 2019; Lavenant et al., 2021). It yields however a curve that is not smooth. On the other hand, due to the inherent difficulty in solving the $P$-spline problem over $W_2$, Chewi et al. (2020a) define a different, and substantially more tractable, interpolant that they call transport splines. Very briefly, they proceed as follows:

1. Start with measures $\mu_i$ at times $t_i$.
2. Compute the $W_2$-optimal couplings $\gamma_{i\rightarrow i+1}$ from $\mu_i$ to $\mu_{i+1}$, which are induced by maps $T_{i\rightarrow i+1}$ provided the $\mu_i$ are absolutely continuous. Let $T_i = T_0 \circ \cdots \circ T_{i-1} \circ i$.
3. Define for each $x$ the flow map $X_t = S_t[x, T_1 x, \ldots, T_N x]$, where $S_t$ is the natural cubic spline interpolant in Euclidean space (with the base times $t_i$ implicit).
4. Output $\mu_t = (X_t) \# \mu_0$.

Not only is computing transport splines very fast, but if the measures $\mu_i$ are sampled from an underlying smooth curve in $W_2$, then the transport spline converges in supremum norm to the true curve at rate $O(\delta^2)$, where $\delta$ is the maximum distance between successive times $t_i$. Importantly for applications, the curve of measure is smooth, which is not true for a piecewise-geodesic interpolant. Other virtues, and their connection with $P$- and $E$-splines, are explored in Chewi et al. (2020a). In this section we aim to define an analogous interpolant in WFR.

The characterization in Theorem 4 is not directly useful here, even if we knew the optimal lifts $\lambda_0$, $\lambda_1$, it is not clear that the optimal coupling is induced by a map. Even if it were, that would not be enough for our purposes — we would want to associate a unique mass $r$ to each initial point $x$, and map the position-mass pair
\((x, r(x))\) to another position-mass pair \((x', r'(x'))\), with \(r'\) the unique mass at \(x'\). Instead we turn to a third characterization of WFR given by Liero et al. (2018).

**Theorem 7.** (Liero et al., 2016, Theorem 8; Liero et al., 2018, Theorem 6.6) Let \(\mu_0\) and \(\mu_1\) be measures, and define

\[
c(x, y) = -2\log \cos \left( |x - y| \wedge \frac{\pi}{2} \right).
\]

Then

\[
\text{WFR}(\mu_0, \mu_1)^2 = \inf_{\eta \in M_\mu} \text{KL}(\eta_0 \| \mu_0) + \text{KL}(\eta_1 \| \mu_1) + \int c(x, y) \, d\eta(x, y). \tag{15}
\]

Furthermore, the infimum is achieved, and if \(\mu_0\) and \(\mu_1\) are absolutely continuous with respect to the Lebesgue measure, the optimal \(\eta^*\) is unique and is induced by a map.

The cost \(c\) differs from the term in the cone metric \(d_\mathcal{C}\) by taking the minimum against \(\frac{\pi}{2}\), not \(\pi\). This is the difference between transport of points in \(\mathcal{C}\) and transport of measures on \(\mathcal{C}\). To transport \(z_0 = (x_0, r_0)\) to \(z_1 = (x_1, r_1)\) in \(\mathcal{C}\) we may reduce the mass at \(x_0\) to 0, then increase it again at \(x_1\) to \(r_1\), whereas to transport \(\delta_{z_0}\) to \(\delta_{z_1}\) these can be done simultaneously (having a superposition of two deltas), so that the WFR geodesic is at each time a combination of two deltas. The superposition results in a lower overall WFR cost when \(|x - y|\) is less than \(\frac{\pi}{2}\).

The optimal coupling \(\eta\) in Theorem 7 and the optimal coupling \(\gamma\) in Theorem 4 are intimately related, as another theorem of Liero et al. (2018) shows.

**Theorem 8.** (Liero et al., 2018, Theorem 6.2) Suppose \(\eta\) minimizes the objective in Theorem 7, and let \(\eta_i = \pi_i \eta\) be its marginals. Write

\[
\mu_i = \sigma_i \eta_i + \mu^+\,
\]

where \(\sigma_i = d\eta_i / d\mu_i\) and \(\mu_i\) is mutually singular with \(\eta^+_i\). Define the plan \(\gamma_0\) by

\[
d\gamma_0(z_0, z_1) = \delta_{\sqrt{\sigma_0(x_0)}(r_0)} \cdot \delta_{\sqrt{\sigma_1(x_1)}(r_1)} \cdot d\eta(x_0, x_1)
+ \delta_1(r_0) \cdot d\mu_0^+(x_0) \cdot \delta_0(z_1) + \delta_1(r_1) \cdot d\mu_1^+(x_1) \cdot \delta_0(z_0)
\]

Then \(\gamma_0\) is optimal for the objective in Theorem 4.

Now, suppose the \(\mu_0, \mu_1\) are such that there \(\mu_i \ll \eta_i\), so that \(\mu^+_i = 0\) in the theorem above. This happens, for instance, when the conditions of Proposition 2 are satisfied along the geodesic between them. The optimal \(\eta\) for Theorem 7 is supported on a map \(T\), and thus an optimal \(\gamma_\eta\) for Theorem 4 is supported on the assignment

\[(x_0, r_0(x_0)) \to (T(x_0), r_1(T(x_0)))\]

which associates to each \(x_0\) a unique mass \(r_0(x_0) = \sqrt{\sigma_0(x_0)}\) and maps it to another unique location and mass \(r_1(T(x_0)) = \sqrt{\sigma_1(T(x_0))}\). This is much stronger than \(\gamma_\eta\) being induced by a map on \(\mathcal{C}\).

Now, define the operator \(S^C_\mathbf{F} [z_0, \ldots, z_N]\) to be the Riemannian cubic interpolant in \(\mathcal{C}\) of the points \(z_0, \ldots, z_N\). We define transport splines over WFR by the following procedure

1. For each \(i\) solve (15) between \(\mu_i\) and \(\mu^+_{i+1}\) to obtain the optimal coupling \(\eta_{i \to i+1}\), and thus the map \(T_{i \to i+1}\). Let \(T_i = T_{0 \to 1} \circ \cdots \circ T_{i-1 \to i}\).
2. For each $x$, form a path $\tilde{X}_t(x)$ interpolating the $x_i = T_i(x)$ and a mass path $\tilde{R}_t(x)$ interpolating the masses $r_i(x_i)$.

3. Define the interpolating curve on $[t_{i-1}, t_i]$ by computing the Cone spline $(X_t, R_t)$ with endpoint velocity constraints $X_{t_j}(x) = \frac{d}{dt}\tilde{X}_t(x)|_{t=t_j}$ and $R_{t_j}(x) = \frac{d}{dt}\tilde{R}_t(x)|_{t=t_j}$ for $j = i-1, i$ and $i = 1, \ldots, n$.

In the first step, we obtain a family of transport maps and corresponding masses at each point $x$ from the coupling $\eta_{i\to i+1}$. Notice that two adjacent couplings $\eta_{i-1\to i}$ and $\eta_{i\to i+1}$ may not have the same $i$-th marginal, so the mass $\sqrt{\sigma_i}$ may be not uniquely defined. We remedy this by rescaling the coupling to have unit mass, so that the coupling is supported on $\sqrt{\mu_i(x_i)}$. Lifted plans on the cone are scale-invariant, thus the optimality of (15) is preserved under this scaling, as explained in Liero et al. (2016, Section §3.3). In the next step, we would ideally interpolate any sequence $(x_i, r_i)_i$ using Riemannian cubics on $C$, these are difficult to compute; unlike the Euclidean case, there appears to be no closed formula. We propose instead approximating the velocities at each knot from the (Euclidean) spline curves of $X$ and $R$ independently (Chewi et al., 2020a, Appendix D). We then solve in step 3 the cone spline problem by specifying the velocities computed in the previous step and using De Casteljau’s algorithm.

### 4.1 De Casteljau’s algorithm on the cone

In Euclidean space, an interpolating curve with minimal curvature is completely determined by the endpoint velocities. The De Casteljau’s algorithm computes points on this curve by iteratively finding points along the paths between the endpoints and two other control points. We describe the procedure for reference. The De Casteljau curve that interpolates $x_0$ and $x_3$ with speeds $v_0$ and $v_3$ at times $t = 0$ and $t = 1$ respectively is constructed as follows:

1. Compute control points $x_1 = x_0 + \frac{v_0}{2}$ and $x_2 = x_2 - \frac{v_0}{2}$.
2. Compute the first intermediate points $u_i(t) = (1-t)x_i + tx_{i+1}$, for $i = 0, 1, 2$.
3. Compute the second intermediate points $u_j(t) = (1-t)w_j + tw_{j+1}$, for $j = 0, 1$.
4. Compute the spline as $p(t) = (1-t)u_0 + tu_1$.

We extend this definition to the cone by replacing the linear interpolation by a geodesic interpolation (Absil et al., 2016; Gousenbourger et al., 2019). It is known that the geodesic has closed-form expression given by:

**Theorem 9.** (Liero et al., 2018, Section §8.1) The geodesic interpolator of the points $z_j = (x_j, r_j)$ is given by $z(t) = (x(t), r(t))$, where

$$ x(t) = (1 - \rho(t)) x_0 + \rho(t) x_1, $$

$$ r(t)^2 = (1 - t)^2 r_0^2 + \ell^2 r_1^2 + 2t(1-t) r_0 r_1 \cos |x_0 - x_1|, $$

$$ \rho(t) = \frac{1}{|x_1 - x_0|} \arccos \left( \frac{(1-t)r_0 + tr_1 \cos |x_1 - x_0|}{r(t)} \right). $$

Theorem 9 is only valid for distances $|x_1 - x_0| < \pi$, otherwise the geodesic simply goes from $x_0$ through the tip of the cone $\sigma$ and back to $x_1$. In terms of measures, this behavior corresponds to pure growth-decay and no transport. From the discussion above about geodesics on the cone and geodesics on WFR, and because we want to model scenarios where both transport and growth are present at every moment, we restrict the computation of cone splines to points that are contained within a.
ball of radius \( \pi/2 \). We will denote the geodesic interpolation of points \( z_0 \) and \( z_1 \) at time \( t \) as \( z_0 \#_t z_1 \).

In order to extend step 1 above, we must first understand how the velocities of a De Casteljau spline at \( t = 0, 1 \) relate to the control points \( x_1, x_2 \).

**Proposition 3.** Let \( (x_0, r_{x_0}) \) and \( (x_3, r_{x_3}) \) be given points on the cone, and \( (x_1, r_{x_1}) \) and \( (x_2, r_{x_2}) \) control points. Let

\[
(w_i(t), r_{w_i}(t)) := (x_i, r_{x_i}) \#_t (x_{i+1}, r_{x_{i+1}}), \quad i = 0, 1, 2 \\
(u_j(t), r_{u_j}(t)) := (w_j, r_{w_j}) \#_t (w_{j+1}, r_{w_{j+1}}), \quad j = 0, 1 \\
(p(t), r_p(r)) := (u_0, r_{u_0}) \#_t (u_1, r_{u_1}),
\]

be the De Casteljau spline on the cone, then

\[
\dot{p}(0) = 3 \frac{r_{x_0}}{r_{x_0}} \frac{|x_0 - x_1|}{|x_0 - x_1|} (x_1 - x_0) \quad (16) \\
\dot{r}_p(0) = 3 (r_{x_1} \cos |x_0 - x_1| - r_{x_0}) \quad (17) \\
\dot{p}(1) = 3 \frac{r_{x_2}}{r_{x_3}} \frac{|x_3 - x_2|}{|x_3 - x_2|} (x_3 - x_2), \quad (18) \\
\dot{r}_p(1) = 3 (r_{x_3} - r_{x_2} \cos |x_3 - x_2|). \quad (19)
\]

**Proof.** See Appendix C.

In particular, at an interval \([t_i, t_{i+1}]\), in order to achieve the position and mass velocities given by \( v_i = \frac{d}{dt} X_i(x) \mid_{t=t_i} \) and \( s_i = \frac{d}{dt} \tilde{R}_i(x) \mid_{t=t_i} \), we have to choose the control point \( (x_1, r_1) \) as \( x_1 = x_{t_i} + c_1 \frac{v_i}{\|v_i\|} \) and \( r_1 = c_2 r_{t_i} \), where \( c_1, c_2 \) are solved from (16-17):

\[
c_1 = \text{arctan} \left( \frac{\|v_i\|}{\frac{s_i}{r_{t_i}} + 3\delta^{-1}} \right), \\
c_2 = \frac{\delta}{3} \left( \frac{\|v_i\|}{\frac{s_i}{r_{t_i}} + 3\delta^{-1}} \right)^2,
\]

where \( \delta = t_{i+1} - t_i \). Similarly for \( t_{i+1} \), we choose \( (x_2, r_2) \) with \( x_2 = x_{t_{i+1}} - c_3 \frac{v_{i+1}}{\|v_{i+1}\|} \), \( r_2 = c_4 r_{t_{i+1}} \), and solve (18-19):

\[
c_3 = \text{arctan} \left( \frac{\|v_{i+1}\|}{3\delta^{-1} - \frac{s_{i+1}}{r_{t_{i+1}}}} \right), \\
c_4 = \frac{\delta}{3} \left( \frac{\|v_{i+1}\|}{3\delta^{-1} - \frac{s_{i+1}}{r_{t_{i+1}}}} \right)^2.
\]

The derivation of the above equations requires that \( c_1, c_3 \geq 0 \); it reflects the fact that \( x_1 - x_0 \) and \( x_3 - x_2 \) point in the right direction. This translates to \( s_i \geq -3\delta^{-1} r_{t_i} \) and \( s_{i+1} \leq 3\delta^{-1} r_{t_{i+1}} \), which constrains the possible velocities that are achievable by the De Casteljau’s algorithm. To comprehend why this happens, notice that unlike the Euclidean case, the second variable \( r \) is constrained to be non-negative. This limits the possible masses of the control points \( x_1, x_2 \), which in turn limits the endpoint derivatives \( \dot{r}_p(0) \) and \( \dot{r}_p(1) \) (equations (17),(19)).

There are two limitations to the cone transport spline problem: the maximum \( \frac{\pi}{3} \) diameter in the position space and the bounds in knot velocities in mass space, as described above. Furthermore, the velocities are estimated from two independent spline problems. A simple solution to these limitations is to 1) scale down all distances to a space with a smaller diameter and 2) scale up the knot times to allow for smaller endpoint derivatives.
Figure 1: De Casteljau cone interpolation of measures $\mu_1, \ldots, \mu_4$. Measure values are shown as a heat map (top) and in a three-dimensional plot (bottom) as a function of time. Knot times $(t_1, \ldots, t_4)$ are $(0, 1, 2, 10)$ (left), $(0, \frac{10}{3}, \frac{20}{3}, 10)$ (center), and $(0, 8, 9, 10)$ (right). $\sigma = 0.06$.

5 Numerical Experiments

In this section, we illustrate the behavior of the cone transport spline via the De Casteljau’s algorithm. We showcase different aspects of the paths qualitatively by considering spline interpolation of measures in one and two dimensions. In section 5.1 we consider interpolation with varying times. We show in section 5.2 two alternative ways of solving a spline problem given by discrete measures by considering a problem in two dimensions.$^5$

5.1 Time effects and Linear interpolation

In this subsection, we look at the effects of time distribution of knots for a one dimensional cone spline problem. The map obtained from solving (7) is induced by a map only when the measures are absolutely continuous with respect to the Lebesgue measure. Indeed, the WFR plan from a single particle to two particles cannot be induced by a map. The same effect is present when solving problem (7) computationally. Here we choose to solve problem (7) using entropic regularization (Chizat et al., 2018; Peyré et al., 2019) and derive an approximation of the map from its solution by taking expectation$^6$ of the marginals: $T(x) = \mathbb{E}_\eta[y|x]$, inspired by Pooladian and Niles-Weed (2021).

Let $\gamma_\sigma(x) = \exp\left(-\frac{1}{2\sigma^2}x^2\right)1_{[-2, 2]}(x/\sigma)$. For our first experiment, we interpolate the measures $d\mu_1(x) = \gamma_\sigma\left(x - \frac{1}{2}\right)dx$.

$^5$The MATLAB code used for generating the figures can be found here https://github.com/felipesua/WFR_splines.

$^6$We can assume $\eta$ is a probability measure by the scale invariance property of the projection from the cone.
\[ d\mu_2(x) = \frac{1}{2} (\gamma_\sigma(x - 0.3) + \gamma_\sigma(x - 0.7)) \, dx, \]
\[ d\mu_3(x) = (\gamma_\sigma(x - 0.3) + \gamma_\sigma(x - 0.7)) \, dx, \]
\[ d\mu_4(x) = \frac{1}{2} 1_{[0,1]}(x) \, dx, \]
which we interpolate at three set of knot times \((t_1, \ldots, t_4)\): \((0, 1, 2, 10), (0, \frac{10}{3}, \frac{20}{3}, 10), \) and \((0, 8, 9, 10)\) in figure 1.

### 5.2 Space discretization

We give details for the interpolation of measures in two dimensions in figure 2. For this example suppose we want to interpolate a set of particles. Since we need to represent them as absolutely continuous measures to be able to obtain a map, we convolve with a kernel and discretize the domain like in the previous section. As noted by Lavenant et al. (2021), gridding grows exponentially with the dimension which makes it infeasible for high-dimensional applications. We show here the result of fine gridding and compare it to sampling uniformly from the support.

Suppose the resulting interpolating measures are as follows. Let
\[ \gamma_\sigma(x) = \exp \left( -\frac{1}{2\sigma^2} \|x\|^2 \right) 1_{B_2(0)}(x). \]

We interpolate
\[ d\mu_1(x) = \frac{3}{4} \gamma_{2\sigma}(x) \, dx, \]
\[ d\mu_2(x) = 0.65 \left( \gamma_\sigma(x - (\frac{\sqrt{3}}{20}, \frac{\sqrt{3}}{20})) + \gamma_\sigma(x - (0, -\frac{\sqrt{3}}{20})) + \gamma_\sigma(x - (\frac{\sqrt{3}}{20}, -\frac{\sqrt{3}}{20})) \right) \, dx, \]
\[ d\mu_4(x) = \gamma_{2\sigma}(x - (\frac{1}{5}, 0)) \, dx. \]

### Acknowledgements

We thank Philippe Rigollet and Sinho Chewi for helpful comments and suggestions on the manuscript.

### A Derivation of the covariant derivative

**Theorem 2.** Let \( \mu_t \) be an absolutely continuous curve in Wasserstein-Fisher-Rao space satisfying the continuity equation with tangent fields \((v_t, \alpha_t)\). The covariant derivative is given by

\[ \frac{D^2}{dt^2} \mu_t = \left( \frac{\partial_t v_t + \nabla v_t \cdot v_t + 4 \alpha_t v_t}{\partial_t \alpha_t + \frac{1}{2} \nabla \alpha_t^2 + 2 \alpha_t^2} \right). \]

**Proof.** Let \( u_t^i = (u_t^i, \beta_t^i) \) be two tangent fields along a curve \( \mu_t \), which has derivative \((v_t, \alpha_t)\). Metric compatibility then reads

\[
\frac{d}{dt} \langle u_t^1, u_t^2 \rangle_{\mu_t} = \int \langle \partial_t u_t^1, u_t^2 \rangle + \langle u_t^1, \partial_t u_t^2 \rangle + 4 \partial_t \beta_t^1 \beta_t^2 + 4 \beta_t^1 \partial \beta_t^2 d\mu_t \\
+ \int \langle u_t^1, u_t^2 \rangle + 4 \beta_t^1 \beta_t^2 d(\partial_t \mu_t) \\
= T_1 + T_2
\]
Figure 2: Subsampled (left) and gridded (right) interpolating measures. Measures $\mu_1, \ldots, \mu_4$ are shown in blue, orange, yellow, and purple respectively (top). Interpolated paths are computed for the points shown in red (middle) as a function of time varying by color. The width of select curves is changed according to the value of the measure (bottom). $\sigma = 0.01$
By the continuity equation and the dual definition of \( \text{div} v_t \mu_t \), the second integral becomes

\[
T_2 = \int \langle \nabla u_t^1 \cdot v_t, u_t^2 \rangle + \langle u_t^2, \nabla u_t^1 \cdot v_t \rangle + 4\alpha_t \langle u_t^1, u_t^2 \rangle \\
+ \int 4\beta_t^2 \langle \nabla \beta_t^1, v_t \rangle + 4\beta_t^1 \langle \nabla \beta_t^2, v_t \rangle + 16\alpha_t \beta_t^1 \beta_t^2 \, d\mu_t
\]

Now, we need this to be equal to \( \langle u_t^1, \frac{D}{dt} u_t^2 \rangle + \langle \frac{D}{dt} u_t^1, u_t^2 \rangle \). Some terms are uniquely attributable to one or the other, such as \( \int \langle \partial_t u_t^1, u_t^2 \rangle \), but some are not, such as \( 4\beta_t^2 \langle \nabla \beta_t^1, v_t \rangle \). This can arise from either of the two terms: as \( \nabla \beta_t^1 = u_t^1 \),

\[
\langle (0, \langle \nabla \beta_t^1, v_t \rangle), u_t^2 \rangle_{\mu_t} = \langle u_t^1, (4\beta_t^2 v, 0) \rangle_{\mu_t}
\]

Indeed, by splitting these terms and gathering the others, the possible covariant derivatives that satisfy metric compatibility are of the form

\[
\frac{D}{dt} u_t = \mathcal{P}_{\mu_t} \left( \partial_t u_t + \nabla u_t \cdot v_t + 2\alpha_t \beta_t + 4\alpha_t \beta_t v_t \right)
\]

for real \( p \). Let us check the torsion-free identity with \( p = 1/2 \). In this case, if \( u_t = (\nabla \varphi, \varphi) \) is constant in time, then

\[
\frac{D}{dt} (\nabla \varphi) = \mathcal{P}_{\mu_t} \left( \nabla^2 \varphi \cdot v_t + 2\alpha_t \nabla \varphi + 2\varphi v_t \right)
\]

Now, with the setup as in the \( W_2 \) case, defining \( F: \varphi \mapsto \int \varphi \, d\mu \) and writing \( \varphi = (\nabla \varphi, \varphi) \), we have from the continuity equation

\[
\partial_t F[\mu_t^i] = \int \langle \varphi, v_t^i \rangle \, d\mu_t^i
\]

where \( v_t^i \) is the derivative of \( \mu_t^i \). As above, we have

\[
u_0^1(u^2(F))[\mu] = \left. \frac{d}{dt} \langle \varphi, \mu_t^2 \rangle \right|_{t=0}
\]

\[
= \left. \left( \frac{D}{dt} \varphi, u_t^2 \right)_{\mu_t^2} + \left( \varphi, \nabla u_t u_t^2 \right)_{\mu_t^2} \right|_{t=0}
\]

Recalling that at \( t = 0 \) we have \( u_0^2 = v_0^1 = (v_0^1, \alpha_0^1) \), the first term becomes (we may ignore the projection, since \( u_0^2 \) is already tangent)

\[
Q_1 = \left. \left( \nabla^2 \varphi \cdot v_0^1 + 2\alpha_0^1 \nabla \varphi + 2\varphi v_0^1 \right), (v_0^1, \alpha_0^1) \right]_{\mu_0^2}
\]

while the corresponding term from \( u_0^2(u^2(F))[\mu] \) is

\[
Q_2 = \left. \left( \nabla^2 \varphi \cdot v_0^1 + 2\alpha_0^1 \nabla \varphi + 2\varphi v_0^1 \right), (v_0^2, \alpha_0^2) \right]_{\mu_0^2}
\]

and we must check that these agree. The “top-left”, “top-right”, and “bottom-right” terms are identical for both. The top-middle of the first is equal to the bottom-left for the second, and vice-versa. Thus we have shown that the covariant derivative is given by

\[
\frac{D}{dt} u_t = \mathcal{P}_{\mu_t} \left( \partial_t u_t + \nabla u_t \cdot v_t + 2\alpha_t \beta_t + 2\alpha_t \beta_t v_t \right)
\]

and in specific,

\[
\frac{D^2}{dt^2} u_t = \left( \partial_t v_t + \nabla v_t \cdot v_t + 4\alpha_t v_t \right)
\]

This quantity is tangent, so no projection is necessary.
B Proof of equivalence in curvature

**Theorem 6.** Let $\mu_t$ be a sufficiently smooth curve in WFR. Then there is a measure $P \in \mathcal{P}(\Omega_\varepsilon)$ such that $\mathbb{P}[(e_t)\# P] = \mu_t$ for all $t$, and the $E$-cost of $\mu$ is equal to the $P$-cost of $P$. The measure $P$ is induced by the flow maps associated to the curve $\mu_t$.

**Proof.** Making the definition explicit, we wish of our measure $P$ that the cost of (12) is equal to

$$\int_0^1 \int |\partial_t v_t + \nabla v_t \cdot v_t + 4\alpha_t v_t|^2 + 4(\partial_t \alpha_t + \frac{1}{2} |\nabla \alpha_t|^2 + 2\alpha_t^2)^2 \, d\mu_t \, dt \tag{21}$$

Let $\lambda_0 \in \mathcal{P}(\mathcal{C})$ be a lift of $\mu_0$. By Proposition 2, the measure $\lambda_t = (X_t, R_t)_{\#} \lambda_0$ is a lift of $\mu_t$.

In order to compute the covariant derivative of a curve on the cone, we compute the Christoffel symbols. These are given by the formulas

$$\nabla_{\partial x_i} \partial x_j = \Gamma^k_{ij} \partial_k, \quad \Gamma^k_{ij} = \frac{1}{2} g^{kl} \left( \frac{\partial}{\partial j} g_{il} + \frac{\partial}{\partial i} g_{lj} - \frac{\partial}{\partial l} g_{ij} \right).$$

Let $\partial x_1, \ldots, \partial x_n, \partial r$ be the coordinate basis of the tangent space on the cone, hence

$$(g_{ij}) = \begin{pmatrix} r^2 I_n & 0 \\ 0 & 1 \end{pmatrix}, \quad \frac{\partial}{\partial r} g_{ij} = \begin{cases} 2r & i = j \neq r, \ k = r, \\
0 & \text{else.} \end{cases}$$

Since $(g^{ij})$ and $\partial_r g_{ij}$ are only defined along the diagonal and $\partial_l g_{ij} = 0$ for $l \neq r$, the only terms that do not vanish are

$$\Gamma^r_{ij} = r^{-1} \delta_{jk}, \quad \Gamma^r_{ij} = -r \delta_{ij}$$

thus the Levi-Civita connection on the cone is given by

$$\nabla_{\partial\phi} \partial_r = 0, \quad \nabla_{\partial\phi} X = r^{-1} X, \quad \nabla X_1 X_2 = -r \langle X_1, X_2 \rangle \partial_r.$$  

From this, the covariant derivative of a curve $z_t = (x_t, r_t)$ on the cone is given by

$$\dot{z}_t = \nabla_{\dot{x}_t} \dot{z}_t = \left( \ddot{x}_t + 2\frac{\dot{r}_t}{r_t} \dot{x}_t, \ \ddot{r}_t - r_t \dot{x}_t \right)$$

thus from the Riemannian metric (11)

$$|\dot{z}_t|^2 = |r_t \ddot{x}_t + 2\frac{\dot{r}_t}{r_t} \dot{x}_t|^2 + |\dot{r}_t - r_t |\dot{x}_t|^2|^2$$

Let $\lambda_0$ be any lift of $\mu_0$ and define $P$ to place mass $\lambda_0(z_0)$ on the path $z_t = (X_t(x_0), R_t(r_0))$, so that $P$ is supported on the flow map curves in $\Omega_\varepsilon$. By applying the total derivative to the defining equations of the flow maps, these curves satisfy

$$\dot{x}_t = v_t(x_t)$$
$$\ddot{x}_t = \partial_v v_t + \nabla v_t \cdot v_t$$
$$\dot{r}_t = 2r_t \alpha_t$$
$$\ddot{r}_t = 2r_t \left( \partial_r \alpha_t + \nabla \alpha_t \cdot v_t + 2\alpha_t^2 \right)$$

Now, we have

$$\int_{\Omega_\varepsilon} \int_0^1 |\dot{z}_t|^2 \, dt \, dP(z) = \int_\varepsilon \int_0^1 |\dot{z}_t|^2 \, dt \, d\lambda_0(z_0)$$
Expanding out, this is
\[
\int_0^1 \int_0^1 \left| r_t \ddot{x}_t + 2 \dot{r}_t \dot{x}_t \right|^2 + \left| \dot{r}_t - r_t |\dot{x}_t|^2 \right|^2 \, dt \, d\lambda_0(z_0)
\]

Let us deal with each term separately so the expressions do not become unwieldy. For the first

\[
\int_0^1 \int_0^1 \left| r_t \ddot{x}_t + 2 \dot{r}_t \dot{x}_t \right|^2 \, dt \, d\lambda_0(z_0)
\]

\[
= \int_0^1 \int_0^1 r_t^2 |\partial_t v_t + \nabla v_t \cdot v_t + 4 \alpha_t v_t|^2 (x_t) \, dt \, d\lambda_0(z_0)
\]

\[
= \int_0^1 \int_0^1 |\partial_t v_t + \nabla v_t \cdot v_t + 4 \alpha_t v_t|^2 (x_t) \, d(r_t^2 \cdot \lambda_0)(z_0) \, dt
\]

\[
= \int_0^1 \int_0^1 |\partial_t v_t + \nabla v_t \cdot v_t + 4 \alpha_t v_t|^2 (x_t) \, d\mu_t(x) \, dt
\]

We have used that \((X_t, R_t) \# \lambda_0 = \lambda_t\) and \(\mathfrak{P} \lambda_t = \mu_t\). The second term is dealt with in exactly the same way,

\[
\int_0^1 \int_0^1 |\dot{r}_t - r_t |\dot{x}_t|^2 \, dt \, d\lambda_0(z_0)
\]

\[
= \int_0^1 \int_0^1 4r_t^2 |\partial_t \alpha_t + \frac{1}{2} \nabla \alpha_t \cdot v_t + 2 \alpha_t|^2 (x_t) \, dt \, d\lambda_0(z_0)
\]

\[
= \int_0^1 \int_0^1 4 |\partial_t \alpha_t + \frac{1}{2} \nabla \alpha_t \cdot v_t + 2 \alpha_t|^2 (x_t) \, d(r_t^2 \cdot \lambda_0)(z_0) \, dt
\]

\[
= \int_0^1 \int_0^1 4 |\partial_t \alpha_t + \frac{1}{2} \nabla \alpha_t \cdot v_t + 2 \alpha_t|^2 (x_t) \, d\mu_t(x) \, dt.
\]

\[\square\]

### C Cone De Casteljau’s algorithm

**Proposition 3.** Let \((x_0, r_{x_0})\) and \((x_3, r_{x_3})\) be given points on the cone, and \((x_1, r_{x_1})\) and \((x_2, r_{x_2})\) be control points. Let

\[
(w_i(t), r_{w_i}(t)) := (x_i, r_{x_i}) \# t(x_{i+1}, r_{x_{i+1}}), \quad i = 0, 1, 2
\]

\[
(u_j(t), r_{u_j}(t)) := (w_j, r_{w_j}) \# t(w_{j+1}, r_{w_{j+1}}), \quad j = 0, 1
\]

\[
(p(t), r_{p}(t)) := (u_0, r_{u_0}) \# t(u_1, r_{u_1}),
\]

be the De Casteljau spline on the cone, then

\[
\dot{p}(0) = 3 \frac{r_{x_1}}{r_{x_0}} \sin \frac{|x_0 - x_1|}{r_{x_0}} (x_1 - x_0)
\]

(16)

\[
\dot{r}_p(0) = 3 (r_{x_1} \cos \frac{|x_0 - x_1|}{r_{x_0}} - r_{x_0})
\]

(17)

\[
\dot{p}(1) = 3 \frac{r_{x_2}}{r_{x_3}} \sin \frac{|x_3 - x_2|}{r_{x_3}} (x_3 - x_2)
\]

(18)

\[
\dot{r}_p(1) = 3 (r_{x_3} - r_{x_2} \cos \frac{|x_3 - x_2|}{r_{x_3}})
\]

(19)

**Proof.** Let \(\theta_i := |x_{i+1} - x_i|\). Since the expressions for \(w, u\) and \(P\) are all the same with different interpolating points, we compute the derivatives of \(w \rho, r\) and \(\theta\)
with the subscripts removed. To compute each derivative, we just replace with the corresponding subscript.

For \( \theta = |x_1 - x_0| \),

\[
\theta \dot{\theta} = (x_1 - x_0, \dot{x}_1 - \dot{x}_0).
\]

For a point \( w_0 = (1 - \rho)x_0 + \rho x_1 \),

\[
u \dot{w}_0 = (x_1 - x_0) \dot{\rho} + (1 - \rho) \dot{x}_0 + \rho \dot{x}_1.
\tag{22}
\]

For the mass \( \rho' = \rho_0^2 (1 - t)^2 + \rho_1^2 t^2 + 2 \rho_0 \rho_1 t (1 - t) \cos(\theta) \),

\[
r \dot{r} = 2 r_0 \dot{r}_0 (1 - t)^2 - r_0^2 (1 - t) + r_1 \dot{r}_1 t^2 + r_1^2 t 
+ (\dot{r}_1 r_0 + \dot{r}_0 r_1) t (1 - t) \cos(\theta) 
+ r_0 r_1 (1 - 2 t) \cos(\theta) - r_0 r_1 t (1 - t) \sin(\theta) \dot{\theta}.
\tag{23}
\]

For the local time \( \rho = \frac{1}{r} \arccos \left( \frac{r_0 (1 - t) + r_1 t \cos(\theta)}{r} \right) \),

\[
\dot{\rho} + \rho \dot{\theta} = \frac{1}{r} \frac{r_0^2 (1 - t) t \sin(\theta)}{r_0 r_1 (1 - t) \cos(\theta)} (\dot{r}_1 r_0 - \dot{r}_0 r_1) 
+ (1 - t) t \cos(\theta) r_0 r_1 \dot{\theta} 
+ r_0 r_1 \sin(\theta) + r_1^2 \dot{\theta} t^2).
\tag{24}
\]

For the first interpolation points \( w_0, w_1, w_2 \) we have in particular that \( \dot{x}_i, \dot{r}_x_i = 0 \), hence

\[
\dot{\theta}_{w_i} = 0,
\]

\[
\dot{w}_i = (x_{i+1} - x_i) \dot{\rho}_{w_i},
\]

\[
\dot{r}_{w_i} = \frac{1}{r_{w_i}} \left( -r_{w_i}^2 (1 - t) + r_{x_{i+1}}^2 t + r_{w_i} r_{x_{i+1}} (1 - 2 t) \cos(\theta_{w_i}) \right),
\]

\[
\dot{\rho}_{w_i} = \frac{r_{x_i} r_{x_{i+1}}}{\theta_{w_i}} \sin(\theta_{w_i}) - \rho_{w_i} \frac{\dot{\theta}_{w_i}}{\theta_{w_i}}.
\]

Thus at \( t = 0 \),

\[
\dot{\theta}_{w_i}(0) = 0
\]

\[
\dot{\rho}_{w_i}(0) = \frac{r_{x_{i+1}} \sin(\theta_i)}{r_{x_i} \theta_i}
\tag{25}
\]

\[
\dot{r}_{w_i}(0) = r_{x_{i+1}} \cos(\theta_i) - r_{x_i}
\tag{26}
\]

\[
\dot{w}_i(0) = (x_{i+1} - x_i) \frac{r_{x_{i+1}} \sin(\theta_i)}{r_{x_i} \theta_i}
\tag{27}
\]

and \( t = 1 \),

\[
\dot{\theta}_{w_i}(1) = 0
\]

\[
\dot{\rho}_{w_i}(1) = \frac{r_{x_i} \sin(\theta_i)}{r_{x_{i+1}} \theta_i}
\tag{28}
\]

\[
\dot{r}_{w_i}(1) = r_{x_{i+1}} - r_{x_i} \cos(\theta_i)
\tag{29}
\]

\[
\dot{w}_i(1) = (x_{i+1} - x_i) \frac{r_{x_i} \sin(\theta_i)}{r_{x_{i+1}} \theta_i}.
\tag{30}
\]
We substitute these expression back into the formulas for the derivatives of position (22), mass (23) and local time (24) for \( u_j \) and \( p \), for \( j = 0, 1 \). Notice that for both \( t = 0 \) and \( t = 1 \) the depdence on \( \dot{\theta} \) vanishes.

\[
\dot{\rho}_{u_j}(0) = \frac{r_{x_{j+1}} \sin(\theta_j)}{r_{x_j}} \tag{33} \\
\dot{r}_{u_j}(0) = 2 \left( r_{x_{j+1}} \cos(\theta_j) - r_j \right) \tag{34} \\
\dot{u}_j(0) = 2 (x_{j+1} - x_j) \frac{r_{x_{j+1}} \sin(\theta_j)}{r_{x_j}} \frac{\theta_j}{\theta_{j+1}} \tag{35}
\]

\[
\dot{\rho}_{u_j}(1) = \frac{r_{x_j} \sin(\theta_{j+1})}{r_{x_{j+1}}} \tag{36} \\
\dot{r}_{u_j}(1) = 2 \left( r_{x_{j+1}} - r_j \cos(\theta_{j+1}) \right) \tag{37} \\
\dot{u}_j(1) = 2 (x_{j+1} - x_j) \frac{r_{x_{j+1}} \sin(\theta_{j+1})}{r_{x_j}} \frac{\theta_j}{\theta_{j+1}} \tag{38}
\]

Finally for \( p \),

\[
\dot{\rho}_{p}(0) = \frac{r_{x_1} \sin(\theta_0)}{r_{x_0}} \frac{\theta_0}{\theta_{j+1}} \tag{39} \\
\dot{r}_{p}(0) = 3 \left( r_{x_1} \cos(\theta_0) - r_{x_0} \right) \tag{40} \\
\dot{p}(0) = 3 (x_1 - x_0) \frac{r_{x_1} \sin(\theta_0)}{r_{x_0}} \tag{41}
\]

\[
\dot{\rho}_{p}(1) = \frac{r_{x_2} \sin(\theta_2)}{r_{x_3}} \frac{\theta_2}{\theta_{j+1}} \tag{42} \\
\dot{r}_{p}(1) = 3 \left( r_{x_3} - r_{x_2} \cos(\theta_2) \right) \tag{43} \\
\dot{p}(1) = 3 (x_3 - x_2) \frac{r_{x_2} \sin(\theta_2)}{r_{x_3}} \frac{\theta_2}{\theta_{j+1}} \tag{44}
\]

\[\square\]

**References**

Absil, P.-A., Gousenbourger, P.-Y., Striewski, P., and Wirth, B. (2016). Differentiable piecewise-bézier surfaces on riemannian manifolds. *SIAM Journal on Imaging Sciences*, 9(4):1788–1828.

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

Benamou, J.-D., Gallouët, T. O., and Vialard, F.-X. (2019). Second-order models for optimal transport and cubic splines on the wasserstein space. *Foundations of Computational Mathematics*, 19(5):1113–1143.

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., Gouic, T. L., Rigollet, P., Stepaniants, G., and Stromme, A. J. (2020a). Fast and smooth interpolation on wasserstein space. arXiv preprint arXiv:2010.12101.

Chewi, S., Maunu, T., Rigollet, P., and Stromme, A. J. (2020b). Gradient descent algorithms for bures-wasserstein barycenters. In Conference on Learning Theory, pages 1276–1304. PMLR.

Chizat, L. (2017). Unbalanced optimal transport: Models, numerical methods, applications. PhD thesis, PSL Research University.

Chizat, L., Peyré, G., Schmitzer, B., and Vialard, F.-X. (2018). Scaling algorithms for unbalanced optimal transport problems. Mathematics of Computation, 87(314):2563–2609.

Gigli, N. (2012). Second Order Analysis on $(\mathcal{P}_2(M), W_2)$. American Mathematical Soc.

Gousenbourger, P.-Y., Massart, E., and Absil, P.-A. (2019). Data fitting on manifolds with composite bézier-like curves and blended cubic splines. Journal of Mathematical Imaging and Vision, 61(5):645–671.

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

Lavenant, H., Zhang, S., Kim, Y.-H., and Schiebinger, G. (2021). Towards a mathematical theory of trajectory inference. arXiv preprint arXiv:2102.09204.

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

Liero, M., Mielke, A., and Savaré, G. (2018). Optimal entropy-transport problems and a new hellinger–kantorovich distance between positive measures. Inventiones mathematicae, 211(3):969–1117.

Lisini, S. (2007). Characterization of absolutely continuous curves in wasserstein spaces. Calculus of variations and partial differential equations, 28(1):85–120.

Maniglia, S. (2007). Probabilistic representation and uniqueness results for measure-valued solutions of transport equations. Journal de mathématiques pures et appliquées, 87(6):601–626.

Otto, F. (2001). The geometry of dissipative evolution equations: the porous medium equation.

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.

Pooladian, A.-A. and Niles-Weed, J. (2021). Entropic estimation of optimal transport maps. arXiv preprint arXiv:2109.12004.
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.